Defeasible Knowledge

Emil Frederik Lundbjerg Moeller
Merton College, Oxford
DPhil Thesis

Defeasible Knowledge (DPhil thesis in Philosophy)
Emil F L Moeller, Merton College
September 2014

ABSTRACT

This dissertation collects five papers that discuss potential consequences of the defeasibility of knowledge. Chapters 1, 2, 3, and 5 provide arguments for a number of important epistemological consequences of the defeasibility of knowledge. Chapter 4 discusses consequences that others have claimed the defeasibility of knowledge to have. Chapter 1 argues that closure principles for knowledge are in tension with the defeasibility of knowledge. Chapter 2 argues that one of Williamson's famous arguments against the KK principle relies in a problematic way on a closure principle that is incompatible with defeat. Chapter 3 argues that a view on which knowledge just is belief safe from error is in tension the defeasibility of knowledge. Chapter 4 discusses two arguments to the effect that defeat by higher-order evidence sometimes involves the violation of rational ideals or rules of rational belief formation. As part of a response to an objection to contextualism about 'know', Chapter 5 investigates a number of social epistemological consequences of the defeasibility of knowledge.

Contents

Acknowledgements

Most of what I know about epistemology I have learned from John Hawthorne. During the last three years he has patiently and gently helped me get clearer in my thinking about the issues in the thesis and about epistemology more generally. He has given insightful and incisive comments on all five chapters of the thesis from which they have benefitted tremendously. The thesis is indebted first and foremost to him. I am also grateful to Harvey Lederman for helpful comments on chapters 2, 4, and 5, and to fellow graduate students, postdocs, faculty members and visitors at the Oxford Philosophy Faculty for discussions of arguments and ideas in the thesis. A special thanks goes to my friend Emma Loftus. Had it not been for her enthusiasm and care, I would have withered in my room, and both the thesis and I would have suffered. She made the rainy days in Oxford tolerable.

Finally, I would like to express my gratitude to Kat for her love and moral support, and to Tøger for being Tøger. Without them I would not have pulled through.

Introduction

This dissertation collects five papers that discuss potential consequences of the idea that one can lose knowledge by receiving misleading evidence. Insofar as belief is a necessary condition for knowledge, strong misleading evidence often destroys knowledge because it destroys belief. The focus of this dissertation, however, is the more substantial idea that if the acquired evidence is of the right strength or kind, one will lose knowledge even if one retains belief. The idea that knowledge is defeasible in this way is taken for granted in much epistemological theorizing and can be motivated by considering cases like the following:

> *Red Lighting*[1] *– Part 1*
>
> At time t_1 Suzy is looking at a red object in front of her and based on her experience she forms the belief that the object in front of her is red. There is nothing abnormal about Suzy's eyesight or the lighting in the room.

> *Premier League Reports*[2] *– Part 1*
>
> *The Times* and *The Guardian* are both extremely reliable in reporting Premier League results and Jack tends to trust them equally. Man U just won an important match against Chelsea and on the front page of *The Guardian* Jack finds the Man U victory reported. He therefore forms the belief that Man U won.

When presented with these cases the inclination is to say that both Suzy's belief and Jack's belief amount to knowledge. Consider now the following developments:

---

[1] This case is a version of Lasonen-Aarnio's (2010:1) *Trick on Suzy*.
[2] This case is a modified version of a case from Hawthorne (2004:71).

Red Lighting – Part 2

At time t_2 a person, Sally, whom Suzy trusts and considers highly reliable tells Suzy that the object she was looking at at t_1 was illuminated by red lighting that would make objects of any colour look red.

Premier League Reports – Part 2

Shortly after reading *The Guardian*, Jack reads on the front page of *The Times* that Man U lost the match.

In light of Suzy's misleading conversation with Sally and Jack's misleading encounter with *The Times* we are inclined to say that even if Suzy still believes that the object was red and even if Jack still believes that Man U won the match, these beliefs do not amount to knowledge.

The standard way for philosophers to think about cases like *Red Lighting* and *Premier League Reports* is to see them as cases exemplifying the phenomenon of knowledge defeat. That is, philosophers tend to think of Suzy as having lost knowledge as a consequence of the misleading testimony from Sally and to think of Jack as having lost knowledge as a result of his misleading encounter with *The Times*[3,4].

The five papers or chapters of the dissertation discuss various potential consequences of the idea that knowledge is defeasible in this way[5]. Whereas chapters 1, 2, 3 and 5 provide

[3] See Chapter 3 for a more thorough discussion of these cases.

[4] Philosophers sometimes distinguish between *rebutting* and *undercutting* defeaters (see Pollock 1986:38-39). A rebutting defeater for a belief that p is typically specified a reason for believing the negation of p. An undercutting defeater for a belief that p is typically specified as reason for no longer believing p without hereby being a reason for believing the negation. Philosophers attracted to this distinction typically think of the evidence Suzy receives in *Part 2* as an undercutting defeater and the evidence Jack receives in *Part 2* as a rebutting defeater.

[5] The above characterization of knowledge defeat is diachronic: a subject knows a proposition at time t_1, but loses that knowledge at t_2 as a result of acquiring misleading evidence. In the philosophical literature it is

arguments for a number of important epistemological consequences of the defeasibility of knowledge, Chapter 4 discusses consequences that other authors have claimed the defeasibility of knowledge to have.

Chapter 1 provides a treatment of how closure principles for knowledge interact with the defeasibility of knowledge. I argue that the traditional formulations of knowledge closure are in tension with the defeasibility of knowledge, and I provide a number of reasons for thinking that the tension is not easily resolved and may have important ramifications for epistemological theorizing.

Chapter 2 provides a discussion of one of Williamson's (2000) central arguments against the KK principle (i.e. the principle stating that if one knows that p, one knows that one knows p). I present a number of reasons for thinking that Williamson's (2000) style of argument reveals problems with his preferred closure principle for knowledge rather than problems with KK. Roughly speaking, I argue that Williamson's (2000) preferred closure principle for knowledge is in tension with the idea that knowledge is defeasible. The kind of case Williamson (2000) uses in his argument against KK is more appropriately seen as a case exemplifying the tension between knowledge closure and knowledge defeat than as a case exemplifying the falsity of KK.

Chapter 3 provides a detailed study of how a simple safety account of knowledge interacts with the defeasibility of knowledge. Although the argument turns out to be far from straightforward, I ultimately conclude that our judgments in certain defeat cases

---

standardly assumed that there is a related non-diachronic notion of defeat on which defeat need not involve any *losing* of knowledge. Imagine, for example, that Jack in *Premier League Reports* had read the false match report in *The Times* prior to reading the correct match report in *The Guardian*. In this alternative scenario it is standard to describe the misleading evidence Jack acquires by reading *The Times* as a *defeater* that prevents Jack from coming to know that Man U won by reading the correct match report in *The Guardian*. In the chapters of the dissertation the defeasibility of knowledge refers to both the diachronic and the non-diachronic notion of knowledge defeat.

cannot be satisfactorily explained by reference to failures of safety. A view according to which knowledge just is belief safe from error is therefore in tension with the defeasibility of knowledge.

Chapter 4 discusses arguments by Christensen (2010a) and Lasonen-Aarnio (2014a) to the effect that we violate rational ideals or rules of rational belief formation by revising our beliefs in the face of certain kinds of misleading higher-order evidence. I point out that Christensen's (2010a) and Lasonen-Aarnio's (2014a) arguments do not provide us with a reason for thinking that rational ideals or rules of rational belief formation are violated in the kinds of cases they discuss.

Chapter 5 responds to an objection to contextualism about 'know'. According to the objection, contextualism is unable to account for our 'consumption' of knowledge claims made in other contexts. I provide two interpretations of the objection and respond to both. As part of my response to one of these interpretations I point out that irrespective of what semantics is correct for 'know', the defeasibility of knowledge creates a number of difficulties for our cross-contextual 'consumption' of knowledge claims.

While most epistemologists agree that we can lose knowledge by receiving misleading evidence, some authors have recently suggested that the phenomenon of knowledge defeat may be much more limited than epistemologists have tended to think. These defeat skeptical authors have presented a number of considerations in defense of their more *indefeasibilist* conceptions of knowledge.

One argument for the limits of defeat starts from Williamson's (2011, *forthcoming*) case for the possibility of improbable knowing. According to Williamson (2011, *forthcoming*), it is possible for a subject S to know p while the probability on S's evidence

that S knows p is arbitrarily close to 0. At least initially, this result is rather surprising, since one might have thought that knowing p is incompatible with having strong evidence than one does not know p. Williamson (2011, *forthcoming*), however, argues that the phenomenon of improbable knowing is an unavoidable consequence of the very intuitive idea that we require a margin for error in at least some (and plausibly much) of our knowledge[6].

As Williamson (*forthcoming*:37-41) has noted, another argument for the possibility of improbable knowing comes from combining (a version of[7]) multi-premise closure for knowledge with the claim that knowing is not a luminous condition[8] (i.e. that we are not always in a position to know that we know). By multi-premise closure, if one knows each of a number of premises and deduces the conjunction of those premises, then one knows the conjunction. Imagine that there is a large number of propositions that one knows, but does not know that one knows. For each premise there is some small probability on one's evidence that one does not know the truth of that premise. Multi-premise closure nonetheless allows one to come to know the conjunction of those propositions. On certain ways of spelling out the case, competent deduction will hereby allow one to know the conjunction although it is incredibly unlikely on one's evidence that one knows.

---

[6] To take an example from Williamson (2000:114-115), imagine that you are looking at a tree from some distance and attempt to make a judgment about the height of the tree on the basis of your visual experience. Given the limitations of your eyesight and ability to judge heights, you cannot determine the height of the tree to the nearest inch just by looking. If, for example, the tree is 666 inches tall, then you cannot come to know that the tree is not 665 inches tall just by looking. You need a greater margin for error. According to Williamson (2011, *forthcoming*), improbable knowing is a consequence of our need for margins for error in at least some (and plausibly much) of our knowledge.

[7] The statement of multi-premise closure Williamson (*forthcoming*) assumes in his discussion of this alternative route to improbable knowing is rather strong and far from uncontroversial. See Chapter 2 for a discussion.

[8] See also Williamson (2009b) and Hawthorne & Srinivasan (2013) for such arguments for improbable knowing.

If either of the mentioned arguments for improbable knowing are sound, then the question naturally arises of how widespread the phenomenon of improbable knowing actually is. Are there, for example, putative cases of defeat that in fact are cases of improbable knowing?[9]

Consider, for instance, the following kinds of cases:

> A: After having (competently) diagnosed a patient and prescribed medication you realize that you have been awake for 36 hours. As a result, you start to worry about possible diagnostic errors given your sleep-deprived state.
>
> B: You realize that someone you consider an epistemic peer has reached a different conclusion than you by reasoning from the same body of evidence (in fact your perceived peer has made a mistake).
>
> C: Someone you trust (falsely) informs you that you have been given a drug that makes you prone to make errors in reasoning.
>
> D: After having (competently) gone through a sequence of reasoning to check that everything is set for your climb down the mountain, you acquire strong evidence that you are in serious danger of suffering from hypoxia[10] caused by the high altitude (in fact you are not hypoxic).

These are all cases in which you receive (misleading) evidence that some of your beliefs are the result of a flawed process (evidence of this kind is sometimes called *higher-order*

[9] This question is raised by Williamson (*forthcoming*:42): ”Knowledge claims are often thought to be defeated by various sorts of misleading evidence. In how many cases is the correct account that the subject knows, even though it is almost certain on the subject’s evidence at the time that they do no not know? That is left an open question.”

[10] Hypoxia is a condition caused by oxygen deprivation. The condition can cause hard-to-detect cognitive impairments.

evidence). If the evidence in question is sufficiently strong, many will be inclined to see A, B, C, and D as cases in which the evidence you receive about the possible epistemic shortcomings of some of your beliefs can prevent those beliefs from being known[11]. But if Williamson's (2011, *forthcoming*) arguments for improbable knowing are sound, then perhaps we should consider the possibility of interpreting A, B, C, and D as cases of improbable knowing rather than as ordinary cases of knowledge defeat?[12] And perhaps we should similarly consider interpreting *Part 2* of *Red Lighting* and *Part 2* of *Premier League Reports* as circumstances in which Suzy continues to know that the object in front of her is red and Jack continues to know that Man U won the match[13], although it is incredibly unlikely on their evidence that they know?

In connection with this defeat skeptical idea, it is worth noting that cases like A, B, C, and D are exactly the kinds of cases that Christensen (2010a) and Lasonen-Aarnio (2014a) seem to think of as cases in which subjects violate rational ideals or rules of rational belief formation by revising their beliefs in the face of (misleading) evidence that those beliefs are flawed. At least on certain ways of developing these claims, Christensen's (2010) and Lasonen-Aarnio's (2014) consideration might be seen as providing an independent motivation for thinking that strong (misleading) evidence that one's beliefs are flawed does not always prevent those beliefs from being known.

---

[11] Much will depend on exactly how the cases are spelled out, and as the growing literature on the epistemology of disagreement reveals, there are, of course, difficult questions about when and how higher-order evidence affects the epistemic status of our beliefs (see, e.g., Christensen & Lackey (2013) for some recent contributions to the disagreement debate). Even so, I think that many are inclined to say that evidence that some of your beliefs are flawed at least sometimes has defeating force with respect to those beliefs.

[12] Much more would have to be said to motivate such an interpretation of cases like A, B, C, and D. My aim here and in the following is not to spell out detailed arguments for thinking that defeat is more limited than many of us tend to think, but merely to hint at some of the considerations that have motivated certain authors to suggest that it might be.

[13] Assuming, of course, that Suzy continues to believe that the object is red and Jack continues to believe that Man U won the match.

Another motivation for a more defeat skeptical conception of knowledge comes from externalist theories of knowledge. Simple externalist theories of knowledge are often claimed to have difficulties accommodating our judgments about knowledge in a range of putative defeat cases, but perhaps those attracted to a simple externalist theory of knowledge could just reject the relevant cases as actual cases of defeat? Cases like A, B, C, D, *Red Lighting* and *Premier League Reports* might all be cases in which the subject of the case satisfies some popular externalist condition on knowledge, and some externalists about knowledge might therefore be attracted simply to see the relevant cases as cases of knowledge (see, e.g., Hawthorne & Srinivasan 2013 and Lasonen-Aarnio 2010).

Of course, any argument for rejecting certain putative defeat cases as actual cases of defeat will have to be wedded to an error theory about our initial judgments in the relevant cases. There are, however, various proposals for how this might be done (see Hawthorne & Srinivasan (2013), Lasonen-Aarnio (2010, 2014a), Williamson (2011, *forthcoming*)).

While the chapters of the dissertation provide little or no explicit discussion of the limits of defeat, several chapters have some relevance for the above kinds of motivations for accepting a more indefeasibilist conception of knowledge[14].

As previously mentioned, Chapter 3 provides a detailed study of how a simple safety account of knowledge interacts with the defeasibility of knowledge and concludes

[14] There are a number of motivations for accepting a more indefeasibilist conception of knowledge that are not discussed in any of the chapters of the dissertation. For example, Hawthorne & Srinivasan (2013) and Lasonen-Aarnio (2010) present a number of considerations to the effect that our judgments in defeat cases may be the result of our confusing unreasonable or blameworthy belief with an absence of knowledge. Lasonen-Aarnio (2014b) has also sketched a possible argument from the dogmatism puzzle to an indefeasibilist conception of knowledge. Neither of these motivations for a more indefeasibilist conception of knowledge are discussed in the chapters of the dissertation.

that the simple safety account is unable to accommodate our judgments in a number of putative defeat cases. As such, Chapter 3 supports the thought that those attracted to a simple externalist theory of knowledge will have to take on board certain indefeasibilist commitments and will need an error theory of some kind to explain our judgments in a range of putative defeat cases.

Chapter 4 argues that certain considerations provided by Christensen (2010a) and Lasonen-Aarnio (2014a) do not ultimately give us a reason for thinking that we violate rational ideals or rules of rational belief formation in cases like A, B, C, and D. If the argument of that chapter is correct, then it is unclear in which (if any) cases we violate rational ideals or rules of rational belief formation by revising our beliefs in the face of strong misleading higher-order evidence. It is therefore at best unclear what support a more indefeasibilist conception of knowledge could draw from the claimed existence of such violations if we have no clear grasp of the cases in which such violations occur.

For those who take closure principles to be an indispensable part of a proper theory of knowledge, Chapter 1 might be turned into an argument for a more indefeasibilist conception of knowledge. I suspect, however, that many will see the argument of Chapter 1 as raising worries for knowledge closure rather than as putting pressure on our judgments about defeat. If that is right, then the argument of Chapter 1 might also raise worries for the previously mentioned argument from multi-premise closure and the assumed failure of KK to improbable knowing. If we have independent reason to think that defeat raises problems for knowledge closure, then perhaps the suggested case of improbable knowing is more appropriately seen as yet another case in which defeat considerations reveal problems for knowledge closure.

Finally, Williamson's (2011, *forthcoming*) argument for why our need for margins for error leads to improbable knowing is an extension of his argument against KK discussed in Chapter 2 of the dissertation. If the argument of Chapter 2 is correct, then Williamson's argument against KK relies in a problematic way on closure principles that are incompatible with defeat, in which case Williamson's (2011, *forthcoming*) case for improbable knowing is under significant pressure. As such, Chapter 2 questions whether an important motivation for a more indefeasibilist conception of knowledge is ultimately defendable.

Chapter 5 is not directly relevant for the question of whether knowledge is defeasible in the range of cases we typically take it to be[15]. While the discussion in Chapter 5 is limited to dealing with a particular objection raised for contextualism, the defeat related considerations presented in that chapter point to some broader social epistemological issues pertaining to defeat. These include how misleading evidence affects the testimonial transmission of knowledge and how defeat might affect the value of knowing that someone knows. As such, the defeat related considerations in Chapter 5

[15] Since Chapter 5 provides a discussion of an objection to contextualism about 'knows', it is worth noting that a contextualist semantics has certain features that might look attractive to proponents of a more indefeasibilist conception of knowledge. In particular, contextualists sometimes claim that the epistemic position that a subject S must be in with respect to a proposition p for "S knows that p" to be true as uttered in a conversational context C can change when new non-p possibilities are made salient in C. Plausibly, attending to (misleading) evidence that p is false makes certain non-p possibilities salient in one's conversational context. Attending to misleading evidence might therefore change the epistemic position that S must be in for "S knows that p" to be true as uttered in one's current conversational context. This opens up the possibility of claiming that the misleading evidence acquired by S does not change S's epistemic position with respect to p in any important respects, but merely changes the epistemic position that S must be in for "S knows that p" to be true as uttered by those who attend to the relevant evidence. As such, someone attracted to a more indefeasibilist line of thought might hold that what explains our reaction in *Part 2* of *Premier League Reports* is not really that reading *The Times* has destroyed Jack's knowledge of the outcome of the Man U match. Rather, our attending to the misleading evidence Jack acquires by reading *The Times* has changed the standards that Jack must satisfy for "Jack knows that Man U won" to be true as uttered by us (and by himself). On this kind of view, misleading evidence changes the standards operative in a conversational context, but need not seriously affect the epistemic position of the subject acquiring that evidence. I think such a view faces a number of difficulties, but will leave discussion of these difficulties for another occasion.

might provide the starting point for further investigations into the potential social epistemological consequences of the defeasibility of knowledge.

Since the dissertation is a collection of papers and not a traditional research monograph, each of the chapters or papers of the dissertation is written to stand alone. The reader of the whole dissertation will thus find a number of cases, definitions and considerations repeated throughout, which will at least make reading the dissertation faster. The bibliography is at the end of the whole dissertation.

CHAPTER 1

Knowledge Closure and Knowledge Defeat

Knowledge closure has long been a topic of controversy in epistemology. Despite its initial appeal, various theoretical considerations have led a number of philosophers to reject some or all of the available formulations of knowledge closure[16]. Other philosophers remain unconvinced by the objections standardly leveled against knowledge closure and continue to defend it in both its single- and multi-premise formulations (see, e.g., Hawthorne (2004:31-50, 2005), Williamson (2009a,b)).

In the first part of this paper, I provide a generalization of a potential objection to knowledge closure that so far has received little attention in the literature. The objection originates from Hawthorne (2004:183-5), who notes that his preferred formulation of multi-premise closure (MPC) is in tension with the widely accepted[17] idea that one can lose knowledge by receiving sufficiently strong misleading evidence. Below I argue that the tension noted by Hawthorne (2004) arises for knowledge closure more generally; single-premise closure and more traditional formulations of knowledge closure are in tension with the defeasibility of knowledge as well.

---

[16] Famously, Nozick (1981) and Dretske (2000) both endorse anti-skeptical epistemologies that involve the denial of knowledge closure in both its single- and multi-premise formulations. Other philosophers sympathetic to the idea that knowledge that p is compatible with a small epistemic chance, evidential probability or objective chance that non-p have worried specifically about multi-premise closure, because the chance or probability that a deduced conjunction is false might be big even though the chance or probability that each conjunct is false is small. See, e.g., Hawthorne & Lasonen-Aarnio (2009) for an objection to multi-premise closure along those lines, and see Williamson (2009a,b) for a defense of multi-premise closure against this kind of worry.

[17] While most philosophers think that one can lose knowledge by receiving sufficiently strong misleading evidence, some authors have recently suggested that misleading evidence might not destroy knowledge in as broad a range of cases as philosophers have tended to think. See, e.g., Lasonen-Aarnio (2010) and Hawthorne & Srinivasan (2013).

In the second part of the paper, I argue that the tension between knowledge closure and knowledge defeat is not easily resolved and should not be readily ignored. If knowledge is defeasible, the ramifications for knowledge closure and its place in epistemological theorizing may be rather significant.

Part 1: A Tension Between Closure and Defeat

Let me start by providing the formulation of multi-premise closure (MPC) proposed by Williamson (2000:117) and refined by Hawthorne (2004:33, 2005):

MPC: Necessarily, if S knows p_1, …, p_n, competently deduces q, and thereby comes to believe q, while retaining knowledge of p_1, …, p_n throughout, then S knows q.

The tension between MPC and knowledge defeat noted by Hawthorne (2004:183-5) can now be brought out by the following line of reasoning[18]. Assume that a subject, S, knows p_1, p_2, …, p_n. Assume that S is in roughly the same epistemic position with respect to each of the propositions and roughly equally confident in all of them. Imagine, for example, that p_1, p_2, …, p_n are a number of propositions that S has learned by reading a book that is widely accepted as trustworthy on the relevant issues. Imagine now that S engages in a competent multi-premise deduction. S considers each of the things she believes on the basis of having read the book and deduces the conjunction:

[18] Although my presentation of the tension between MPC and knowledge defeat differs somewhat from Hawthorne's (2004:183-5), the general idea is the same.

$p_1, p_2, \ldots, p_n$

$(p_1 \wedge p_2 \wedge \ldots \wedge p_n)$

Assume that S thereby comes to believe the conjunction. We have no particular reason to think that S by the process of deduction loses knowledge of any of the premises[19], and according to MPC, S therefore comes to know the conjunction.

The tension with knowledge defeat now arises in the following way. Imagine that S receives misleading evidence E with defeating force with respect to S's knowledge of the conjunctive conclusion. Imagine, for example, that someone S considers highly trustworthy convincingly informs S that one of the conjuncts is false. The informant, however, provides no indication of which conjunct. Upon receiving this piece of misleading evidence, we would expect S to give up belief in the conjunctive conclusion and even if S were to retain belief in the conclusion, we are inclined to think that S no longer knows the conclusion.

What, then, about S's knowledge of the premises? If n is very low, we might expect S to give up belief in at least some (if not all) of the conjuncts upon receiving misleading evidence E. Even if S were to retain belief in all of the conjuncts it might seem compelling to think that receiving E will defeat S's knowledge of at least some (if not all) of the conjuncts. Whether or not S continues to believe all of the conjuncts, it is therefore plausible to think that if n is very low, the misleading evidence that defeats knowledge of

[19] If the number of premises is sufficiently high, then someone might suggest that S can lose knowledge of the premises simply by thinking of the possibility that one of the premises is false. To my mind, that result looks overly skeptical. It naturally leads to the idea that a subject can lose all of her knowledge simply by thinking of the possibility that one of the propositions she believes is false.

the conjunctive conclusion also results in a loss of knowledge of at least some of the premises.

What if n is very high? If n is very high, we would not expect S to give up belief in all of the conjuncts upon being told that one of the conjuncts is false. In fact, given that S is roughly as confident with respect to each of the conjuncts and given that E is not targeted at any particular conjunct, we would not expect S to give up belief in any conjunct. As such, we would expect S to continue believing all of the conjuncts if n is very high. Does S also retain knowledge of all of the conjuncts upon receiving E? It does not seem particularly plausible to think that receiving E defeats S's knowledge of some of the conjuncts while allowing S to retain knowledge of some of the other conjuncts. After all, E is not targeted at any particular conjunct and S was in roughly the same epistemic position with respect to each of the conjuncts before receiving E. At the same time, the claim that S loses knowledge of all of the conjuncts looks overly skeptical. If n is very high and if S knew all of the conjuncts before being told that one of the conjuncts is false, then we are not inclined to think that S stops knowing all of the conjuncts upon receiving E. As such, there is some attractiveness to the thought that S does in fact retain knowledge of all of the conjuncts upon receiving misleading evidence E.

So far, this does not pose a problem for MPC as nothing in the provided formulation of MPC prohibits that S can lose knowledge of the conclusion of a competent deduction without losing knowledge of the premises. A worry arises, however, if we consider the possibility that S once again starts thinking about the premises and once again infers the conjunctive conclusion. If S were able to competently deduce the conjunctive conclusion before receiving E, why should S not be able to do so after having received E? We are, after all, considering a fairly simple inference from a set of premises to the

conjunction of those premises. Assuming that S still knows each of the premises, it therefore seems that if MPC is true, then S could come to know the conjunctive conclusion even after having received E. This, however, looks problematic, because S now knows the conjunctive conclusion while at same time having misleading evidence E with assumed defeating force with respect to S's knowledge of the conjunctive conclusion. But presumably that isn't possible. It therefore looks as if MPC isn't compatible with the idea that knowledge is defeasible[20].

It is worth noting that the incompatibility isn't restricted to formulations of multi-premise closure along the lines of Hawthorne's (2004). Consider, for example, the following formulation of multi-premise closure inspired by traditional formulations of closure discussed in, e.g., Cohen (1988), Dretske (2000), Nozick (1981), and Stine (1976):

MPC*: If S knows p_1, …, p_n, knows p_1, …, p_n entails q, and believes q on that basis, then S knows q.

---

[20] Hawthorne (2004:184) briefly mentions the possibility of a strategy for handling the tension between MPC and knowledge defeat inspired by discussions of vagueness. The idea is that once S receives E, E renders it determinate that S does not know all of the conjuncts, but for each conjunct it is indeterminate whether E destroys S's knowledge of that conjunct ('determinate' and 'indeterminate' should here be understood in accordance with one's favorite theory of vagueness). To my mind, it seems like a significant skeptical concession that S no longer determinately knows any of the conjuncts after having received E.
Furthermore, Hawthorne (2004:184, footnote 50) seems to find it a natural assumption that indeterminate propositions cannot be known. If that is true, then S cannot know for each of p_1,…,p_n that she knows it. But there are several reasons to think that S can in fact know of each of p_1,…,p_n that she knows it even after she has received E. First of all, it is rather implausible to think that we cannot obtain higher-order knowledge from books when we have strong evidence that one of the claims in the book is mistaken. Generalized to our other sources of knowledge, the natural consequence is that we rarely have higher-order knowledge of anything. But surely that can't be right. Combined with the knowledge norm of assertion, this conclusion also has the consequence that we are rarely in a position to assert "I know that p" for some p that we have come to know by reading a book or have come to know by relying on other sources of information of which we have strong evidence that they are occasionally mistaken. Finally, it is worth noting that if S knew for each of the conjuncts that S knew it before receiving E (which seems plausible), then we can run an argument for why S retains higher-order knowledge that is parallel to the argument for why S retains knowledge of each of the conjuncts.
Finally, it is worth noting that even if one finds Hawthorne's (2004) suggested vagueness response attractive, it isn't clear how it can be invoked to handle the below argument for why single-premise closure and knowledge defeat are in tension (an argument that easily generalizes to Hawthorne's formulation of multi-premise closure as well).

As before, we are assuming that if n is sufficiently large, then S continues to know each of the conjuncts after having received E. It does not seem plausible that receiving E destroys S's knowledge that $p_1, \ldots, p_n$ entails $p_1 \wedge \ldots \wedge p_n$ (after all, S merely has to know that the truth of the conjunction follows from the truth of the conjunct). Furthermore, we can simply assume that S believes $p_1 \wedge \ldots \wedge p_n$ on the basis of her knowledge of the conjuncts and her knowledge of the entailment. Although S has evidence E with defeating force against knowledge of the conjunctive conclusion, S therefore knows $p_1 \wedge \ldots \wedge p_n$ according to MPC*. It therefore looks as if MPC* also isn't compatible with the idea that knowledge is defeasible.

A somewhat similar argument can now be given for the incompatibility between single-premise closure and the defeasibility of knowledge. Here is Hawthorne's (2004, 2005) formulation of single-premise closure (SPC):

SPC: Necessarily, if S knows p, competently deduces q, and thereby comes to believe q, while retaining knowledge of p throughout, then S knows q.

Assume that a subject, S, knows p. Imagine now that S engages in a competent single-premise deduction and comes to the conclusion that q:

p
---
q

Assume that S thereby comes to believe that q. We have no particular reason to think that S by the process of deduction loses knowledge of p, and according to SPC, S therefore comes to know that q. Imagine as before that S receives misleading evidence E* with defeating force with respect to S's knowledge of the conclusion q. Imagine, for example, that someone S considers an expert on matters concerning q convincingly informs S that q is false. Upon receiving this piece of misleading evidence, we would expect S to give up belief in the conclusion q and even if S were to retain belief in q, we are inclined to think that S no longer knows that q.

What, then, about S's knowledge of the premise p? If the deduction is extremely simple and transparent, we might expect S to give up her belief in p. Even if S were to retain her belief in p, it is compelling to think that receiving E* will defeat S's knowledge of p.

What if the deduction from p to q is extremely complicated and involves a large number of difficult inferential steps?[21] If the deduction is extremely complicated, then we can easily imagine that S will not give up belief in p upon receiving misleading evidence E*. If, for example, S was antecedently highly confident that p, and perhaps much more confident that p than that the complicated deduction is correct, then S might doubt the correctness of the inference rather than give up belief in p when being told that E*. What then of S's knowledge that p? Is that defeated even if S continues to believe that p? Not necessarily. It seems plausible that at least in certain circumstances where the deductive inference from p to q is highly complicated and where S is in a very strong epistemic

[21] Schecter (2013) argues that the mere risk of error in reasoning is sufficient to make us unjustified in believing the conclusion of a sufficiently long and complicated deduction (even if the deduction is competent and the premise remains known throughout). I make no such assumptions about the effect of risk accumulation here.

position with respect to p before receiving E*, S can retain knowledge that p even after receiving E*. Although E* defeats S's knowledge of the conclusion q, it need not defeat S's knowledge of the premise p.

So far, this does not pose a problem for SPC as nothing in the provided formulation of SPC prohibits that S can lose knowledge of the conclusion of a competent deduction without losing knowledge of the premise. A worry arises, however, if we consider the possibility that S once again starts thinking about what follows from p and once again comes to the conclusion that q. If S were able to competently deduce q from p before receiving E*, why should S not be able to do so after having received E*? Presumably S's ability to make competent deductions is not destroyed by S's receiving E*[22]. Additionally, there is no reason to think that S cannot retain her knowledge of p while carrying out the deduction from p to q. When S initially made the competent deduction from p to q, S was able to retain her knowledge of p even as her knowledge of q was destroyed by her receiving E*. Why then think that S's knowledge of p cannot be retained when S competently deduces q from p after having received E*[23]? It therefore seems that S could come to know q even after having received E* if SPC is true. This, however, looks

---

[22] Some might try to gerrymander the notion of competence so that S can no longer competently deduce q from p once S has received E*. Doing so will, however, result in a notion of competence that is highly unnatural and deviates significantly from our ordinary understanding of what competence involves. Deductive competence is not normally thought of as something that misleading evidence can destroy. Additionally, the resulting notion of competence seems to have the slightly odd consequence that while S cannot competently deduce q from p after having received E*, S can still engage in competent deductions from p to other propositions irrespective of whether these deductions are significantly more complicated than the deduction from p to q.

[23] Plausibly, the situation changes if S competently deduces q from p a hundred times in a row. In that case, it might be compelling to think that S's knowledge of the premise p is eventually destroyed or that E* will eventually lose its defeating force with respect to the conclusion q. What is needed to block the above argument, however, is that the first competent deduction from p to q that S makes after having received E* either destroys S's knowledge that p or undermines E*'s defeating force with respect to q. If the deduction from p to q is extremely complicated and if S's epistemic position with respect p is very strong, then it is very difficult to see why S's first competent deduction from p to q after S has received E* must have this effect.

problematic, because S now knows q while at same time having misleading evidence E* with assumed defeating force with respect to S's knowledge of q. But presumably that isn't possible. It therefore looks as if SPC isn't compatible with the idea that knowledge is defeasible.

As we have seen above, a number of formulations of knowledge closure are in tension with the idea that knowledge is defeasible. In the case of multi-premise closure, knowledge of large conjunctions gives rise to a tension between MPC and MPC* on the one hand and the defeasibility of knowledge on the other. In the case of single-premise closure, the possibility of highly complicated competent deductions gives rise to a tension between SPC and the defeasibility of knowledge.

It is interesting to note, however, that more traditional formulations of knowledge closure do not seem committed to the possibility of acquiring knowledge of large conjunctions on the basis of one's knowledge of the conjuncts. Additionally, traditional formulations of knowledge closure make no mention of competent deduction. Consider, for example, the following formulation of what we might call *Traditional Closure* (see, e.g., Cohen (1988), Dretske (2000), Nozick (1981), and Stine (1976)):

TC: If S knows *p*, knows *p* entails *q*, and believes *q* on that basis, then S knows *q*.

Given how TC deviates from MPC, MPC*, and SPC, the proponents of knowledge closure might hope that endorsing TC[24] will dissolve the tension between knowledge closure and

[24] Even if TC was not in tension with the idea that knowledge is defeasible, there are a number of other reasons to be dissatisfied with closure principles formulated along the lines of TC. For a discussion, see Hawthorne (2004, 2005).

the idea that knowledge is defeasible. As the following kind of scenario brings out, this hope is in vain.

Assume that S knows p. Assume also that S knows that p entails r_1, knows that r_1 entails r_2, knows that r_2 entails r_3, ..., knows that r_{n-1} entails r_n, and knows that r_n entails q (we need not assume that S knows the conjunction of these). Assume that S has come to know p and all these other propositions by reading a book that is considered trustworthy on the relevant issues. Assume further that S has come to believe r_1 on the basis of her knowledge that p and her knowledge that p entails r_1. According to TC, S knows r_1. Similarly, S has come to believe r_2 on the basis of her knowledge that r_1 and her knowledge that r_1 entails r_2. According to TC, S therefore knows r_2 as well. In the same way, S has come to know r_3, know r_4, ..., know r_{n-1}, know r_n, and know q. In other words, S has come to know q by a process that accords with TC.

Assume now that S receives misleading evidence E** that defeats S's knowledge that q. Assume further that E** is not targeted specifically at the premise p or any of the other propositions that S has come to know by reading the book (i.e. the propositions r_1 entails r_2, r_2 entails r_3, ..., r_{n-1} entails r_n, r_n entails q). Imagine, for example, that someone S trusts and considers an authority on q tells S that q is false without giving any indications as to the truth-value of p or any of the other propositions S has learned from the book. We can for the sake of argument assume that prior to receiving E**, S was roughly as confident in p as in each of the other propositions S has come to know by reading the book.

Does receiving evidence E** defeat S's knowledge of p or S's knowledge of any of the other propositions S has learned from the book? The considerations here are exactly analogous to those presented in the discussion of MPC and MPC*. If n is very low, we

might expect S to give up belief in p and/or at least some (if not all) of the other propositions from the book upon receiving misleading evidence E**. Even if S were to retain belief in all of the propositions from the book it might seem compelling to think that receiving E** will defeat S's knowledge of p and/or some (if not all) of the other propositions from the book.

What if n is very high? Given that S was roughly as confident in p as in each of the other propositions from prior to receiving E**, there is no reason to think that S would give up belief in p or in one of other propositions while retaining her belief in the rest of them. At the same time, we would not expect S to stop believing p and all the other propositions if n is very high. Rather, we would expect S to continue believing p and all the other propositions S has learned from the book even after receiving evidence that q is false.

Does S then retain knowledge of p and of all the other propositions from the book upon receiving E**? If n is sufficiently high, then it seems overly skeptical to insist that S loses her knowledge of p and her knowledge of all of the other propositions. At the same time, it seems arbitrary to single out p or one of the other propositions and insist that S loses her knowledge of that particular proposition rather than any of others. As a consequence, it is natural to think that S retains her knowledge of p and her knowledge of all the other propositions S has learned from the book.

Using her knowledge that p and her knowledge of the other propositions she has learned by reading the book, S can now come to know r_1, r_2, ..., r_{n-1}, r_n, and q by going through a series of belief forming processes in accordance with TC (again, there is no assumption that S knows the conjunction of the beliefs she hereby acquires). In other words, TC allows that S can come to know q while having misleading evidence E** with

defeating force against q. But presumably that isn't possible. It therefore looks as if TC isn't compatible with the idea that knowledge is defeasible.

If the above line of reasoning is correct, then the tension between knowledge closure and the defeasibility of knowledge is quite general. At least on standard formulations of both single- and multi-premise closure, knowledge closure and knowledge defeat are incompatible.

Part 2: Defeat-Friendly Closure Principles

In this part of the paper I argue that the tension between knowledge closure and knowledge defeat is not easily resolved and should not be readily ignored. If knowledge is defeasible, the ramifications for knowledge closure and its place in epistemological theorizing may be rather significant.

In section **2.1**, I focus on the attempt to resolve the tension between knowledge closure and knowledge defeat by adding a defeat clause[25] to the standard formulations of closure. I argue that any defeat clause that deals satisfactorily with the full range of defeat cases threatens to deplete the resulting closure principles of informative content.

In section **2.2**, I argue that if knowledge is defeasible, then some of the standard uses of closure considerations in various epistemological puzzles and arguments may not be defensible.

In section **2.3**, I argue that closure principles with defeat clauses do not give us a very satisfying articulation of how and when we can extend our knowledge by deduction.

---

[25] See Kvanvig (2006:261-262) and Schecter (2013:436-437) for the idea that closure principles should incorporate defeat clauses of some kind.

2.1. Closure principles with defeat clauses. Using SPC as an example, the following provides an initial suggestion for how the tension between knowledge closure and knowledge defeat might be resolved by invoking defeat clauses:

SPC_{Defeat}: Necessarily, if S knows p, competently deduces q, and thereby comes to believe q, while retaining knowledge of p throughout *and having no defeaters*[26] *for q*, then S knows q.

In the epistemological literature, defeaters have been invoked to deal with a variety of different ways in which a subject can fail to know a proposition.

In the early literature on Gettier-cases, defeasibility accounts of knowledge (e.g., Klein 1976, 1981) proposed the addition of a no-defeat condition to the standard JTB analysis of knowledge as a strategy for dealing with Gettier-cases and fake barn cases. On such accounts, a defeater is not a proposition believed by the subject or a piece of evidence or a mental state had by the subject. Rather, it is some fact or true proposition that in one way or another prevents the subject's (justified true) belief in a given proposition from being known. Defeaters of this kind are sometimes called *factual* or *propositional* defeaters (see, e.g. Bergmann (2006) and Lackey (2008)).

In the literature on externalist accounts of knowledge and justification, no-defeat conditions have typically been proposed to enable externalist views to deal with evidence, beliefs or other mental states had by the subject that can prevent the subject from knowing

[26] A defeater is here the same as what some philosophers refer to as an *undefeated* defeater (see, e.g., Bergmann (2006) and Lackey (2008)).

a given proposition. The kind of defeaters dealt with by such no-defeat conditions are sometimes called *evidential* defeaters, *mental state* defeaters and *psychological* defeaters.

Some have also suggested that beliefs and other mental states that a subject need not in fact have, but *should* or *ought to* have can serve as defeaters that in certain circumstances prevent the subject from knowing a given proposition (see, e.g., Lackey (2008, chap. 2)). Defeaters of this kind are sometimes called *normative* defeaters.

These various categories of defeaters are often not very carefully defined and invoking them to do substantial epistemological work raises a number of difficult questions. To what extent do the various categories of defeaters overlap and how well do they translate into different epistemological frameworks? If one thinks that knowing is a mental state (Williamson 2000:21-48), does any failure to know a given proposition amount to having a *mental state* defeater for that proposition? If one thinks that a subject's evidence is all and only the propositions known by the subject (Williamson 2000:184-208, Hawthorne 2004:139), does a lack of knowledge automatically translate into a case of *evidential* defeat? Can we make sense of the category of *normative* defeaters and the idea that mental states we *should* have, but do not in fact have, can have defeating force?

While I suspect that the answer to these questions may raise a number of worries for the kind of defeat clauses I will discuss below, I will not try to settle these questions here. At this stage, I merely want to note that the notion of a defeater has been used in so many ways that *generically speaking* we can understand a defeater as some aspect of a subject's situation that prevents the subject from knowing a given proposition. On this very generic understanding of the notion of a defeater, the following epistemic principle is presumably true:

(NO-DEFEAT) If S has no defeaters for p, then S knows that p.

For those who think that the falsity of p or the failure to believe p are not properly speaking *defeaters*, a slightly more restricted notion of a defeater is compatible with the truth of the following epistemic principle:

(NO-DEFEAT*) If S believes p, p is true and S has no defeaters for p, then S knows that p.

If the initially proposed statement of SPC_{Defeat} relies on a notion of a defeater that is inclusive enough to make principles like NO-DEFEAT* true, then SPC_{Defeat} will also be true. The problem, however, is that on this understanding of the notion of a defeater, SPC_{Defeat} will also be as uninformative and uninteresting as NO-DEFEAT*.

On this very general understanding of defeat, SPC_{Defeat} will no more articulate the sense in which deduction is a way of extending one's knowledge than the following principle articulates the sense in which guessing is a way of extending one's knowledge:

(GUESSING) If S guesses that p, thereby acquiring the true belief that p while having no defeaters for p, then S knows p.

If the tension between knowledge closure and knowledge defeat is to be resolved by adding a defeat clause to the closure principles discussed in **Part 1**, the defeat clause will

therefore have to be restricted in one way or another[27]. As I will argue in the following, however, it is very difficult to specify a defeat clause that deals satisfactorily with the full range of defeat cases without making the resulting closure principles as uninteresting and uninformative as NO-DEFEAT*.

What would be a natural way of restricting the defeat clause added to SPC? Since the problem cases for MPC, MPC*, SPC and TC in **Part 1** all rely on the idea that misleading evidence can defeat knowledge, some sort of *evidential* defeat clause might seem like a natural suggestion. Using SPC as an example, the resulting closure principle could look roughly as follows:

$SPC_{Evidence}$: Necessarily, if S knows p, competently deduces q, thereby coming to believe q, while retaining knowledge of p throughout and having no *evidential* defeaters for q, then S knows q.

In order to make sense of $SPC_{Evidence}$, we need a gloss on the notion of evidential defeat. One rather straightforward proposal is to understand evidential defeat in terms of a subject's evidential support for a proposition. That is, one might suggest that a piece of evidence E is an *evidential* defeater with respect to S's knowledge that p if adding E to S's total evidence has the effect that p is no longer supported sufficiently by S's total evidence for S to know that p.

---

[27] In his discussion of how risk accumulation might cause problems for single-premise closure for justification, Schecter (2013:439, footnote 32) also notes that adding a generic defeat clause to the standard formulation of single-premise closure for justification "…would trivialize the principle", although he does not spell out why.

A significant problem with this proposal is that we are normally inclined to say that the known premise p from which S deduces q is part of S's total evidence. In order for the conclusion q to be deducible from p, q must follow from or be entailed by p. Since p is still part of S's total evidence after S has received the misleading evidence E, q is still entailed by S's total evidence after S has received E. In terms of evidential probabilities (Williamson 2000:209:237), the probability of q on S's evidence is 1 even after S has received E. Presumably evidential support does not get any stronger than that, and insofar as E works as a defeater that prevents S from coming to know q, the relevant notion of defeat is therefore plausibly not *evidential* defeat[28] if evidential defeat is understood as suggested here.

On many conceptions of evidence, having a piece of evidence E involves having certain mental states (e.g. beliefs with a certain propositional content). Instead of restricting the notion of defeat relevant to knowledge closure to *evidential* defeaters, one could therefore try to focus on what is sometimes called *mental state* defeaters or *psychological* defeaters. Using SPC as an example, the resulting closure principles could look roughly as follows:

SPC_{Mental}: Necessarily, if S knows p, competently deduces q, thereby coming to believe q, while retaining knowledge of p throughout and having no *mental state* defeaters for q, then S knows q.

[28] One could try to defend $SPC_{Evidence}$ by suggesting that on a proper understanding of S's total evidence, the premise p from which S deduces q is not part of S's evidence. To my mind, that proposal does not look particularly attractive. Furthermore, the proposal is incompatible with the idea that a subject's total body of evidence consists of all and only the propositions known by the subject (E=K). Whatever one thinks of this proposal, some of the most prominent defenders of knowledge closure are sympathetic to the identification of a subject's evidence with her knowledge (Williamson (2000:184-208), 2009a,b and Hawthorne 2004:139). Consequently, they will not be attracted to defending $SPC_{Evidence}$ by giving up the rather intuitive idea that the known premise p is part of S's evidence.

Insofar as we can provide a suitable distinction between *mental state* defeaters and other kinds of defeaters, it might seem that SPC_{Mental} will be able to handle the problem case for SPC from **Part 1** without becoming as uninteresting as NO-DEFEAT*.

A problem, however, is that SPC_{Mental} (as well as $SPC_{Evidence}$) has difficulties with defeat cases involving forgetting or memory loss[29]. Imagine, for example, that everything is as in the SPC case from **Part 1**. S has received misleading evidence E* with defeating force against the conclusion q, and after having competently deduced q from the known premise p, S still doesn't know q. Frustrated about the uncertainty that E* has given her with respect to the question of whether q, S decides to take a pill that will make her forget E* and likewise forget that she has taken the pill. After taking the pill, S once again considers what follows from p and engages in an extremely complicated but nonetheless competent deduction to the conclusion that q. S no longer has E* or any other current mental states that can defeat her knowledge that q, and according to $SPC_{Evidence}$ and SPC_{Mental} it therefore seems that S knows q. But presumably, taking pills to get rid of disturbing or misleading evidence does not put S in a position to know q if S were not in a position to know q before taking the pill.

One straightforward way of trying to avoid this problem for SPC_{Mental} is to complicate the provided defeat clause by adding a qualification about mental states previously had by the subject:

[29] I am here assuming that a *mental state* defeater preventing S from knowing q at time t is a mental state had by S at time t.

SPC_{Time}: Necessarily, if S knows p at time t_1, competently deduces q, thereby coming to believe q at time t_2, while retaining knowledge of p throughout and having at no time earlier than or identical to t_2 had mental states that prevent S from knowing q at t_2, then S knows q at t_2.

The resulting principle is not particularly elegant, and the idea of past mental states that defeat current knowledge might look slightly odd. Still, SPC_{Time} might seem to avoid the above problem case for SPC_{Mental}.

An alternative to SPC_{Time} is to deal with the problem case for SPC_{Mental} by invoking the category of *normative* defeaters. That is, although S has gotten rid of E* by taking a pill, one might try to argue that there is *some* sense in which S *should*[30] or *ought to* have E*, and SPC_{Mental} can therefore be developed in the following way in an attempt to avoid the above problem case:

SPC_{M+N}: Necessarily, if S knows p, competently deduces q, thereby coming to believe q, while retaining knowledge of p throughout and having no *mental state* defeaters or *normative* defeaters for q, then S knows q.

While SPC_{Time} and SPC_{M+N} might seem to avoid problems with cases of memory loss, one might worry that there are other problem cases that they will not be able handle. Consider, for instance, the following analogue of fake barn county (see, e.g., Goldman 1976):

[30] I should note that I am not at all convinced that any sense can be made of the suggestion that S *should* or *ought to* have E after S has taken the pill. In fact, I find the category of normative defeaters somewhat suspicious.

PILL. A pharmaceutical company has invented a new pill. The pill affects people's ability to form beliefs on the basis of competent deduction without affecting their ability to make competent deductions. That is, subjects under the influence of the pill will be just as deductively competent as they usually are, but while the subjects think that they form a belief in the conclusion of their competent deduction, they in fact form a belief in a completely unrelated proposition. If S competently deduces q from her knowledge that p, the pill will cause S to form a belief in a completely unrelated proposition r, although it will seem to S that she has formed a belief in the proposition she has deduced from p. The effect of the pill wears off after 5 hours, and during those 5 hours there are typically a few seconds where the subject's belief formation is unaffected.

Assume that Sarah works at the pharmaceutical company, but is unfamiliar with the pill and its effects. Unbeknownst to Sarah, some of the other employees have slipped a pill into Sarah's morning coffee in order to test the effect. After she has completed her coffee, they present her with a proposition p and informs her that p is true. She thereby comes to know p. On the request of the other employees, Sarah starts thinking about what follows from p and competently deduces that q. Just as she has completed her deduction, Sarah coincidently has a few seconds where her belief formation is unaffected by the pill and she therefore forms the belief that q on the basis of her deduction from p.

According to SPC_{Time} and SPC_{M+N} (as well as $SPC_{Evidence}$ and SPC_{Mental}), Sarah knows that q. But Sarah's situation looks exactly analogous to a standard fake barn case. The pill has made competent deduction an extremely unreliable belief forming mechanism, and Sarah's competent deduction from her knowledge that p could therefore easily have resulted in a false belief. As such, there is some attraction to the idea that Sarah does not know q.

One way to avoid the worry that PILL gives rise to for SPC_{Time} and SPC_{M+N} is to invoke the category of *factual* or *propositional* defeaters in a further development of these principles. In the case of SPC_{M+N}, the natural development would be SPC_{M+N+F}:

SPC_{M+N+F}: Necessarily, if S knows p, competently deduces q, thereby coming to believe q, while retaining knowledge of p throughout and having no *mental state* defeaters, *normative* defeaters or *factual* defeaters for q, then S knows q.

The worry for the resulting closure principle, however, is that it avoids the problem cases at the cost of informative content. The defeat clause in SPC_{M+N+F} invokes so many different kinds of defeaters that it is unclear whether it can be distinguished from the generic SPC_{Defeat}. Consequently, it is unclear whether principles like SPC_{M+N+F} are any more informative than NO-DEFEAT*.

This raises the worry that any coherent defeat clause inclusive enough to handle all the ways in which SPC is in tension with knowledge defeat will make the resulting closure principle as uninteresting as NO-DEFEAT*. The worry, of course, generalizes to MPC, MPC* and TC.

The above considerations are not meant to be decisive. Gendler & Hawthorne (2005) have given us reason to be wary of fake-barn-style case like the above, and some may therefore reject the conclusion that Sarah doesn't know in PILL. Others may think that less inclusive categories of defeaters can be invoked to deal with the problem cases without making the resulting closure principles completely uninteresting.

Nonetheless, I think it is fair to say that attempting to resolve the tension between knowledge closure and knowledge defeat by adding a defeat clause to existing closure principles is far from straightforward.

2.2. The role of closure in epistemological theorizing. A number of important epistemological problems and arguments rely in some way on closure considerations. In this section I argue that if we accept the idea that knowledge is defeasible, then some of the standard uses of closure considerations in epistemology are under pressure.

Consider first the lottery problem. Here is rough statement[31] of the problem.

LOTTERY. If someone of modest means announces that he knows he will not have enough money to go on an African safari this year, we are inclined to treat his utterance as true. If he announces that he knows he will not win a major prize in a lottery this year, we are far less inclined to accept his judgment as true (assuming that he occasionally buys a lottery ticket). But the proposition that the person will not have enough money to go on an African safari this year entails that he will not win a major prize in a lottery. If the person knows the former, isn't he at least in a position to know the latter by performing a simple deduction?

In motivating the lottery problem, knowledge closure plays a role at two different stages. In the above statement of LOTTERY, the idea that I can come to know that I will not win

[31] My statement of the lottery problem relies heavily on Hawthorne (2004:1-2).

the lottery by competent deduction from my knowledge that I will not have enough money to go to Africa relies on some notion of closure. Furthermore, the idea that I cannot know in advance of a lottery drawing that my losing ticket is a loser is also sometimes supported by closure considerations.

Let us start by considering how closure considerations are sometimes[32] used to support the idea that I cannot know before the lottery drawing has been announced that my losing ticket is a loser.

Suppose for *reductio* that before the result of the lottery drawing has been announced I can know of my losing ticket that it is a losing ticket. If I can know such things about my own ticket, then presumably I can also know of each of my friends' losing tickets that they are losing tickets. If I can know such things about each of my friends' lottery tickets, then by MPC I can come to know that all of my friends' losing tickets are losing tickets. Suppose I have 998 friends that just as me have losing tickets. Given MPC and the assumption that I can know of the losers that they are losers, competent deduction allows me to know that we all have losing tickets. Suppose I also know that there are exactly 1,000 tickets[33] in the lottery. Assuming MPC, competent deduction allows me to know which ticket is the winning ticket. But surely that is impossible!

The original formulation of MPC therefore plays an important role in motivating the idea that I cannot know before the result of a lottery drawing has been announced that my lottery ticket is a loser. There is reason to suspect, however, that a defeat-friendly

[32] See, e.g., Hawthorne (2004:6-7)

[33] Some might think that even if I am sometimes in a position to know of my ticket that it is a losing ticket prior to the lottery drawing, the lottery will need to have more than a 1,000 tickets for this to be the case. Those attracted to this line of thought should think of the example as involving a lottery with sufficiently many tickets to allow me to know of my losing ticket that it is a loser.

statement of MPC will not be able to play a similar role in motivating the idea that I cannot know that I won't win the lottery. To see why, consider the following modified version of the case initially used to bring out the tension between MPC and the idea that knowledge is defeasible:

Assume that a subject, S, has come to believe a number of propositions p_1, p_2, …, p_n about Chinese history by reading a book that is widely accepted as trustworthy on the relevant issues. Assume that S is equally confident in all of p_1, p_2, …,p_n. Assume (contrary to the first MPC case) that one of the propositions in the book is false. If n is sufficiently large, then we are normally inclined to say that S can know some or many of the true things that S has come to believe by reading the book. After all, denying that we can acquire knowledge from books that contain a few insignificant falsehoods looks overly skeptical and could lead to widespread skepticism if generalized to our other testimonial and non-testimonial sources of beliefs.

If the false proposition has no significance to how the other propositions in the book are motivated, then it even seems possible that S can know all of the true things that S has come to believe by reading the book. If S reads the entire book, then it is not as if the truth of some of the beliefs S has acquired by reading the book are more lucky or accidental than the truth of other beliefs S has acquired by reading the book[34]. For instance, it does not seem that S's epistemic position is any worse with respect to propositions that are written on pages close to the page containing the false proposition. As such, it seems

[34] Even if someone insisted that S cannot come to know the truth-value of claims written on pages close to the page containing the false proposition, S's knowledge of all the other propositions in the book would be sufficient to generate paradoxical consequences if MPC holds. I leave it to the reader to spell out the details.

plausible that there can be circumstances in which S's epistemic position with respect to all of the true propositions is roughly the same.

Imagine now that someone S (rightly) considers highly trustworthy convincingly informs S that exactly one of the propositions S has come to believe by reading the book is false. The informant, however, provides no indication of which proposition. We can assume that the informant does in fact know that exactly one of the propositions in the book is false and that S therefore comes to know by relying on the informant's say-so.

What happens to the beliefs that S has acquired by reading the book? If n is sufficiently large, then we would expect S to behave just as in the first MPC case and continue to believe all of the propositions p_1, …, p_n upon being told that one of the propositions is false. Does S's knowledge of the falsity of one the propositions he believes destroy his knowledge of the true propositions he believes on the basis of having read the book? It looks overly skeptical to claim that S loses knowledge of all of the true propositions in the book. Even without the testimony of a trustworthy informant, we often know that big books will contain or are likely to contain a few mistakes, but normally this is not thought of as something that prevents us from coming to know by reading big books. Similarly, most of us realize that we have at least a few false beliefs, but this is rarely seen as something that prevents whatever true beliefs we might have from being known.

And although we can imagine scenarios in which S's knowledge that one of the propositions in the book is false destroys S's knowledge of *some* of the true propositions in the book, there is no reason to think that it necessarily must. If S is in roughly the same epistemic position with respect to each of the propositions before learning that there is a false proposition in the book, and if S does not lose knowledge of all the true propositions by learning of the existence of a falsehood, then it is unclear why S must necessarily lose

knowledge of *any* of the true propositions in the book. As such, there is some inclination to think that S can retain knowledge of all the true propositions in the book upon learning that one of the propositions in the book is false.

What can S learn by deduction in a situation like this? If n is sufficiently large, then there is some inclination to think that S can come to know small conjunctions of propositions from the book by deducing them from his knowledge of individual propositions. If S knows that Shanghai is the largest city in China and if S knows that Shanghai is not the capital of China, then it seems plausible that competent deduction will allow S to know that Shanghai is the largest city in China but not the capital.

Nonetheless, it seems that S's knowledge that one of the propositions in the book is false has consequences for what S can come to know by competent deduction. Consider the possibility that S thinks about all the things he has come to believe by reading the book. S knows that one of the propositions is false (although he does not know which one), and by competent deduction S comes to believe the conjunction of all but one of the propositions he has come to believe by reading the book. As it turns out, all of the conjuncts are true. We are assuming that S knows all of the true propositions he has come to believe by reading the book, and so S has come to believe the conjunction on the basis of competent deduction from known premises. According to MPC, S therefore knows the conjunction, but just as in first MPC case, that result looks implausible.

The implausibility of the idea that S can come to know the conjunction by deduction is further supported by the observation that if S can know the conjunction of the true propositions in the book, then by competent deduction S can come to know which of the propositions in the book is false. Assume that p_1, …, p_n are all the propositions in the book and p_1,…, p_{n-1} are all the true propositions in the book. Given S's knowledge that

exactly one of the propositions in the book is false, we can assume that S knows that $p_1 \wedge \ldots \wedge p_n$ is false. If S by competent deduction can come to know that $p_1 \wedge \ldots \wedge p_{n-1}$ is true, then from this knowledge and his knowledge that $p_1 \wedge \ldots \wedge p_n$ is false, S can competently deduce that p_n is false. Assuming MPC, S can therefore come to know that p_n is false, but presumably that is impossible.

It is rather natural to think that the above scenario provides us with yet another case in which MPC is incompatible with the idea that knowledge is defeasible. If that is the correct diagnosis, then a defeat-friendly statement of multi-premise closure should give us the result that S in the above case can in fact know $p_1, \ldots, p_{n-1}$, but cannot come to know $p_1 \wedge \ldots \wedge p_{n-1}$ and the falsity of p_n by competent deduction.

The case, however, is almost exactly analogous to a lottery case, and this suggests that a defeat-friendly statement of multi-premise closure is compatible with the idea that if the number of lottery tickets is sufficiently high, then we can know of our losing tickets that they are losers, although this does not put us in a position to know of the winning ticket that it is a winner.

Consequently, a defeat-friendly statement of multi-premise closure cannot be used to support the idea that we cannot know that we won't win the lottery, and therefore cannot play the role that the standard formulations of multi-premise closure play in motivating this particular part of the lottery problem.

As I noted earlier in this section, knowledge closure does not merely play a role in motivating the thought that we cannot know of a losing lottery ticket that it is a loser. In stating the lottery problem, knowledge closure is perhaps even more important in

motivating the thought that if I know that I will not have enough money to go on an African safari this year, then by competent deduction I can come to know that I will not win a major prize in the lottery this year.

In this respect, the statement of the lottery problem relies on knowledge closure in much the same way as typical statements of the skeptical problem (e.g. Hawthorne 2004:36-46), the problem of easy knowledge (e.g. Cohen 2002) and the dogmatism puzzle (e.g. Harman 1973:148).

In standard formulations of the skeptical problem, knowledge closure motivates the thought that if I knows I have hands, then by competent deduction I can come to know that I am not a handless brain in a vat. In the problem of easy knowledge, knowledge closure motivates the thought that I can come to know that my perceptual faculties are reliable by competent deduction from my perceptual knowledge that an object in front of me is red. In the dogmatism puzzle, knowledge closure motives the thought that if I know some proposition p, then competent deduction will allow me to know that all future evidence against p is misleading.

In all these cases, the epistemological puzzle or problem is motivated by the tension between knowledge closure and the supposed inclination to think that we can know the premise but not the conclusion of the relevant deductions. Consequently, one of the possible reactions to these puzzles or problems is to question whether knowledge closure holds in the relevant cases[35].

---

[35] Famously, Dretske (2000, 2005) and Nozick (1981) have defended views that attempt to resolve the skeptical problem by denying that I can come to know that I am not a handless brain in a vat by competent deduction from my knowledge that I have hands. See Sharon & Spectre (2010) for philosophers who takes the dogmatism puzzle to reveal problems with closure.

As I will argue in the following, the defeasibility of knowledge may have some relevance for at least some of the motivations for thinking that knowledge closure holds in these cases. To see how defeat relates to the assumption of closure in these cases, let us start by taking a look at one of the prominent arguments for holding on to closure in our treatment of the skeptical problem.

In his discussion of Dretske's closure denying response to the skeptical problem, Hawthorne (2004:39, 2005:32) has provided what appears to be one of the central arguments for why we should be wary of denying the closure of our knowledge in cases like the above. Specifically, he has pointed out how embarrassing consequences seem to arise if the denial of closure is combined with the idea that knowledge is the norm of assertion. Call the proposition that I have hands 'P' and the proposition that I am not a handless brain in a vat 'Q':

> "I ask S whether she agrees that P. She assert that she does: "Yes," she says. I then ask S whether she realizes that Q follows from P. "Yes," she says. I then ask her whether she agrees that Q. "I am not agreeing to that," she says. I ask her whether she now wishes to retract her earlier claims. "Oh no," she says. "I'm sticking by my claim that P and my claim that P entails Q. I'm just not willing to claim that Q." Our interlocutor now resembles perfectly Lewis Carroll's Tortoise, that familiar object of ridicule who was perfectly willing to accept the premises of a *modus ponens* argument but was unwilling to accept the conclusion (see Carroll, 1895). If we embrace the thesis that knowledge is the norm of assertion, then a Dretske style position on closure will turn us into a facsimile of Carroll's Tortoise." (2005:32)

The closure of our knowledge in the case of the skeptical problem, the lottery problem, the problem of easy knowledge, and the dogmatism puzzle therefore allows us to embrace the

thesis that knowledge is the norm of assertion without being turned into Tortoise-like objects of ridicule.

This line of argument for why we should accept the closure of our knowledge in the case of these epistemological puzzles may, however, look slightly less pressing if we accept the idea that knowledge is defeasible. To see why, consider again the problem cases for closure from **Part 1**.

In both the MPC case and the MPC* case, the subject of the case, S, knows p_1, …, p_n, knows that p_1, …,p_n entails $p_1 \wedge \ldots \wedge p_n$, and knows that $p_1 \wedge \ldots \wedge p_n$ follows from these. Yet, if knowledge is defeasible, S is not in a position to know $p_1 \wedge \ldots \wedge p_n$. In her attempts to adhere to the knowledge norm of assertion, S is therefore turned into a facsimile of Carroll's Tortoise.

In the TC case, we seem to get similar results. Here a subject, S, knows p, knows that p entails r_1, knows that r_1 entails r_2, knows that r_2 entails r_3, ..., knows that r_{n-1} entails r_n, and knows that r_n entails q. S, however, is also assumed to have evidence E** that prevents S from coming to know q. Consequently, it looks like we can get a situation where S knows the two premises of the form [r_i] and [r_i entails r_{i+1}] without being in a position to know r_{i+1}. In her attempts to adhere to the knowledge norm of assertion, S is therefore turned into a facsimile of Carroll's Tortoise.

The SPC case from **Part 1** is slightly different. In the relevant case, S does not know that the premise p entails the conclusion q after she has received misleading evidence E*, and the case therefore does not give rise to similar undesirable consequences.

This, however, does not mean that cases in which SPC and knowledge defeat are in tension never give rise to such consequences when combined with the knowledge norm of assertion.

Consider, for example, PILL from the last section. In the original statement of the case, Sarah has been given a pill that for 5 hours affects her ability to form beliefs on the basis of her competent deduction. During those 5 hours, there are a few seconds where her belief formation is unaffected by the drug. Let us for present purposes strengthen the case by assuming that Sarah has been informed that she has been drugged and also (misleadingly) been told that the effect of the pill is constant.

Those attracted to the idea that knowledge is defeasible will surely think that Sarah cannot come to know by competent single-premise deduction from a known premise. Assume that Sarah by testimony comes to know the proposition p∧(p entails q). Sarah can easily deduce q from p∧(p entails q) and for a few seconds she is also able to form the belief that q on the basis of her deduction.

But since she cannot come to know q, we have a situation where Sarah in her attempt to adhere to the knowledge norm of assertion will accept the premises of what looks a lot like a simple and transparent *modus ponens* argument but will be unwilling to accept the conclusion.

If knowledge is defeasible, we are therefore forced to accept a range of cases of closure failure where adhering to the knowledge norm of assertion will turn us into facsimiles of Carroll's Tortoise.

The fact that denying closure as a response to various epistemological puzzles leads to similar consequences when combined with the knowledge norm of assertion may

therefore seem slightly more tolerable for those who accept that knowledge is defeasible. In other words, the defeasibility of knowledge seems to slightly diminish the force of one of the central arguments for holding on to knowledge closure in our diagnosis of the skeptical problem, the lottery problem, the problem of easy knowledge and the dogmatism puzzle.

On a somewhat more speculative note, it is worth pointing out that those who endorse defeat-friendly statements of knowledge closure could try to argue that they are not committed to the idea that we can come to know that we are not handless brains in vats on the basis of competent deduction from our knowledge that we have hands.

If there is any reason to think that we cannot come to know that we are not handless brains in vats because we have defeaters of some kind, then failing to acquire knowledge of this fact by competent deduction from our knowledge that we have hands need not conflict with a defeat-friendly statement of knowledge closure. And similarly in the case of the lottery problem, the problem of easy knowledge and dogmatism puzzle.

Those who accept the idea that knowledge is defeasible could therefore try to develop a response to these epistemological puzzles or problems by denying the closure of our knowledge in the relevant cases without hereby rejecting their preferred defeat-friendly formulation of knowledge closure (more would, of course, have to be said to determine whether there is any plausibility to such views).

In addition to the epistemological puzzles considered above, there are, of course, a number of other important puzzles and arguments that rely on closure consideration in one way or another. For those attracted to the idea that knowledge is defeasible, the question arises of

whether these puzzles and arguments retain their force if we endorse defeat-friendly statements of knowledge closure (assuming, of course, that plausible defeat-friendly statements of knowledge closure can be provided).

Take as an example Williamson's (2000, chap. 5) famous Mr Magoo argument for the incompatibility of inexact knowledge and the KK principle. Williamson's argument proceeds as a *reductio* against KK with knowledge closure as one of the premises. Does Williamson's argument still go through if his preferred closure principle is replaced with a defeat-friendly statement of knowledge closure?[36].

Although my considerations in this section have merely scratched the surface, enough has been said to indicate that accepting the idea that knowledge is defeasible may have ramifications for some of the standard uses of closure considerations in epistemological theorizing.

2.3. An axiom about knowledge? Knowledge closure is sometimes motivated by the thought that closure principles like SPC and MPC articulate "what is an extremely intuitive idea" (Hawthorne 2004:33), namely that "…deduction is a way of extending one's knowledge..." (Williamson 2000:117) or "…that one can add to what one knows by deductive inference from what one already knows." (Hawthorne 2005:30).

In this section I argue that the principles that result from adding defeat clauses to standard formulations of closure do not seem to give us a very satisfying articulation of the idea that deduction is a way of extending one's knowledge.

---

[36] See Chapter 2 for a detailed discussion.

To see why, consider again the case initially used to bring out the tension between SPC and knowledge defeat in **Part 1**. In the relevant case, a subject, S, competently deduces q from her knowledge that p, but because she has misleading evidence E* and because the deduction is extremely complicated, the deduction does not give S knowledge that q (although she retains her knowledge of the premise p throughout).

Assume for present purposes that S is certain of the premise p. Although the deduction from p to q is extremely complicated, we can assume that S would be able to establish that q follows from p if S repeated the deduction from p to q ten times. E* would then lose its defeating force with respect to the conclusion q and S would therefore come to know q.

How many times S has to go through the deduction from p to q for E* to lose its defeating force and for S to come to know q plausibly depends on S's deductive skill. A subject that is just short of being logically omniscient would plausibly be able to establish that q follows from p in her first attempt at deducing q from p. For such a subject, E* would not prevent knowledge that q even as the subject deduces q from p for the first time.

These considerations suggest that the following case is possible:

MARY&MARIE. Two subjects, Mary and Marie, have exactly the same evidence relevant to the question of whether q (including E*). They both know p with certainty. The only difference between Mary and Marie is that Mary has significantly better deductive skills that Marie. They both engage in competent deduction from their knowledge that p and come to the conclusion that q. They both thereby come to believe q and they both retain knowledge of p throughout. Yet, only Mary knows that q.

Let $SPC_{Defeat\text{-}clause}$ be the principle that we get by adding a defeat clause to SPC. According to $SPC_{Defeat\text{-}clause}$, the difference between Mary and Marie is that only the latter of the two has misleading evidence with defeating force, which is why Mary knows but Marie doesn't.

It seems, however, that the epistemologically significant difference between Mary and Marie is not very well captured by pointing to a difference in misleading evidence. They both started out with the same evidence relevant to the question of whether q, they both knew p with certainty, they both competently deduces q from p, but afterwards one of them knows q and the other doesn't. The fundamental epistemological difference concerns their deductive skills. The difference in evidence is a consequence rather than an explanation. Even if $SPC_{Defeat\text{-}clause}$ provides the right verdict in the case, it therefore isn't clear that $SPC_{Defeat\text{-}clause}$ provides a deep and fundamental articulation of how the concept of knowledge operates in a case like MARY&MARIE.

This does not give us anything like a conclusive reason for rejecting $SPC_{Defeat\text{-}clause}$, but at least it suggests that $SPC_{Defeat\text{-}clause}$ does not provide a fully satisfactory articulation of how and when we can extend our knowledge by deduction. As an attempt to capture the circumstances in which we can gain knowledge by deduction, $SPC_{Defeat\text{-}clause}$ does not seem to carve the concept of knowledge by its joints or to "…articulate something very important about how the concept of knowledge operates" (Hawthorne 2004:36). As such, it is questionable whether $SPC_{Defeat\text{-}clause}$ "…expresses something deep about the nature of knowledge" (Cohen 1999:64) or is appropriate characterized as "…something like an axiom about knowledge" (Cohen 2002:312).

Insofar as closure principles were meant to capture the fundamental workings of the concept of knowledge, principles like $SPC_{Defeat\text{-}clause}$ do not seem very satisfactory.

Conclusion. The standard formulations of both single- and multi-premise closure are in tension with the idea that knowledge is defeasible. The tension is not easily resolved and should not be readily ignored. If knowledge is defeasible, the ramifications for knowledge closure and its place in epistemological theorizing are potentially significant.

CHAPTER 2

Inexact Knowledge, Positive Introspection and Knowledge Closure

1. Introduction. According to the KK principle, also sometimes called *positive introspection*, if one knows that p, one knows that one knows p. If belief is a necessary condition on knowledge, KK fails for the boring reason that one can know p without having formed the belief that one knows p. While a slightly more guarded version of the KK principle would plausibly avoid this and other trivial objections, Williamson (2000) has famously argued that KK fails for a much more interesting reason. Specifically, Williamson has argued that it is possible to know p, believe that one knows p, without knowing that one knows p, because one is not in a strong enough epistemic position to know that one knows. Williamson's (2000:114-134) Mr Magoo argument for why the KK principle is incompatible with inexact knowledge is one of his central arguments for this conclusion[37].

Below I provide a number of reasons for thinking that Williamson's style of argument reveals problems with his preferred closure principle for knowledge rather than problems with KK[38].

---

[37] A version of the argument can also be found in Williamson (1992).

[38] The purpose of this paper is not to defend KK in full generality. There are arguments against KK that I cannot deal with here. For all I say here, Williamson's (2000, chap. 4) anti-luminosity argument may, for example, succeed as an argument against KK. My aim in the following is merely to provide some reasons for thinking that Williamson's (2000) Mr Magoo argument does not in fact reveal problems with KK. Since Williamson's (2011, *forthcoming*) argument for improbable knowing and his treatment of the paradox of the *Surprise Examination* (Williamson 2000, chap. 6) extends the reasoning of his Mr Magoo argument, the potential failure of that argument may be rather important even if there are other reasons to think that KK fails. See also section (**8**) below for more on improbable knowing and the paradox of the *Surprise Examination*.

2. An argument against KK. While Williamson's famous Mr Magoo argument against the KK principle focuses on a case of perceptual knowledge, it will for present purposes be more convenient to restate the argument using a case of knowledge by testimony. As Williamson himself notes, his style of argument "…generalizes to sources beyond present perception, such as memory and testimony" (2000:119), and the case below is parallel to Williamson's Mr Magoo case in all respects central to the argument against KK[39].

WALDO. Imagine that Cholly gives Waldo a book entitled *"50 Extraordinary Facts"*. The book consists of 50 pages with one extraordinary claim on each page. Cholly tells Waldo that thorough fact-checking has established that all but one of the claims in the book are true. Had Cholly provided Waldo with no further information about the book, then Waldo would not be in a position to know of any claim in the book that it is true[40]. Cholly, however, also tells Waldo that the false claim is just around the middle of the book, although Cholly cannot recall the exact page number. Waldo reads the book and starts thinking about exactly what claim might be false. On the basis of Cholly's testimony, there is no way for Waldo to determine exactly which page contains a falsehood. In fact, there is no way for Waldo to know that the claims on the pages immediately preceding and following the page with the false claim are true. Nevertheless, Cholly's testimony has given Waldo knowledge of the truth-value of some of the claims in the book. Waldo knows (among other things) that the 4th and the 44th claim are true. In fact, the claim on page 27 is false, but Waldo does not know that.

---

[39] Notice also that Williamson's (2000:135-146) attempt to apply the lessons of his Mr Magoo case to the paradox of the *Surprise Examination* relies on the possibility of restating his argument against KK in terms of testimonial knowledge.

[40] As will become clear in section (**6**), I am inclined to think that things might have been different if the number of claims in the book had been higher.

Waldo reflects on the limitations of what he can say about the truth-value of claims in the book and comes to realize that he needs some margin for error in order to know that a particular claim is true. This provides us with the first premise of the argument:

(LIMIT) Waldo knows that if the claim on page i+1 is not true, then he cannot know that the claim on page i is true[41,42].

Williamson's style of argument proceeds as a *reductio ad absurdum* and the second premise of the argument is therefore a restricted version of the KK principle:

(KK) For any pertinent proposition p, if Waldo knows p then he knows that he knows p.

The third premise is the closure of Waldo's knowledge:

(C) If p and all members of the set X are pertinent propositions, p is a logical consequence of X, and Waldo knows each member of X, then by competent deduction Waldo can come to know p[43].

---

[41] Following Williamson (2000:115), we can make the conditional "if the claim on page i+1 is not true, then he cannot know that the claim on page i is true" as uncontentious as possible by reading 'if' as the truth-functional conditional.

[42] My statement of (LIMIT) it formulated in terms of what Waldo is in a position to know, whereas Williamson's (2000) original statement of Mr Magoo's margin for error is formulated simply in terms of what Mr Magoo knows. This is an innocuous consequence of the fact that I restate Williamson's case against KK using slightly weaker assumptions about the closure of Waldo's and Mr Magoo's knowledge. (LIMIT) is entirely in the spirit of Williamson's original argument. See the next footnote for a more detailed discussion.

[43] In Williamson's (2000) original Mr Magoo argument, (C) is stated as follows: " If p and all members of the set X are pertinent propositions, p is a logical consequence of X, and Mr Magoo knows each member of X, *then he knows p*" (2000:116, *my italics*). In other words, Mr Magoo is not merely in a position to know the consequences of what he knows, but automatically knows them without having to carry out any deductions. My dynamic restatement of (C) is entirely in the spirit of Williamson's (2000) original argument, and is chosen merely because it makes it easier to discuss the role of closure in Williamson's argument against KK. Williamson (2000) should find my restatement of (C) completely acceptable, since my version

We can see how these premises interact by assuming that Waldo knows of some *i* that the claim on page *i* is true:

(B_i) Waldo knows that the claim on page *i* is true.

From (KK) it follows that:

(KB_i) Waldo knows that he knows that the claim on page *i* is true.

From (KB_i), (LIMIT) and (C) we now get that:

> Waldo can come to know that the claim on page *i*+1 is true by competent deduction from (LIMIT) and his knowledge that he knows the claim on page *i* is true.

On the assumption that Waldo carries out the deduction we get that:

(B_{i+1}) Waldo knows that the claim on page *i*+1 is true.

The *reductio* now proceeds by observing that Waldo knows the truth-value of some of the claims in the book:

(B_1) Waldo knows that the claim on page 1 is true.

---

of (C) fits naturally with how (C) is supposed to be motivated by its connection to what he calls *intuitive closure*: "Assumption (C) is true if deduction is a way of extending one's knowledge: that is, if knowing p_1, …, p_n, competently deducing q, and thereby coming to believe q is in general a way of coming to know q. Call that principle *intuitive closure*" (2000:117).

By (KK) Waldo knows that he knows the claim on page 1 is true, and by (LIMIT) and (C) Waldo can come to know that the claim on page 2 is true. By assuming that Waldo carries out the relevant deductions and by repeating the argument for values of *i* from 1 to 26 starting from (B_1) we reach the conclusion (B_{27}):

(B_{27}) Waldo knows that the claim on page 27 is true.

(B_{27}) is false, for the claim on page 27 is false and knowledge is factive. We can assume that Waldo is able to carry out the deductions competently, and as a consequence, either (LIMIT), (KK) or (C) has to be rejected. Williamson's own conclusion is that this style of argument reveals the falsity of (KK).

Plausibly, Waldo can know that he requires some margin for error for his beliefs about the truth-value of particular claims in the book to amount to knowledge. I will therefore assume that (LIMIT) is correct[44]. Instead, I will focus on the feasibility of blocking Williamson's style of argument against (KK) by denying (C).

Before considering the plausibility of (C), I should say a bit more about the details of Waldo's situation. Parallel to Williamson's original Mr Magoo argument, the primary focus of the above argument is the observation that (KK), (C) and (LIMIT) together lead to the impossible result that Waldo knows of the false claim on page 27 that it is true.

[44] Sharon & Spectre (2008) and Dokic & Égre (2009) have argued that Williamson's (2000) margin for error principle is in fact the problematic assumption of his argument against KK. I do not find the arguments against his use of the margin for error principle at all convincing, but will leave the discussion of these argument for another occasion.

It is worth noticing, however, that given (KK), (C) and Waldo's knowledge of his margin for error, Waldo (just as Mr Magoo in Williamson's original case) is in an impossible situation even before he comes to know that the false claim on page 27 is true.

In order to reach the conclusion that the claim on page 27 is true, Waldo needs to know that he knows that the claim on page 26 is true. But given Waldo's margin for error, he cannot know that the claim on page 26 is true if the claim on page 27 is false, and therefore cannot know that he knows the claim on page 26 is true. We therefore have the impossible situation that there is a true claim in the book that Waldo is not in a position to know (i.e. the claim on page 26), but which the assumption of (KK), (C) and (LIMIT) nonetheless allows Waldo to know.

Notice furthermore that whereas (KK), (C) and Waldo's knowledge of his margin for error by one route leads to the impossible result that Waldo can come to know that the false claim on page 27 is true, the same three assumptions can likewise lead to the opposite (and similarly impossible) result that Waldo can come to know that the claim on page 27 is false.

To see how, consider again Waldo's knowledge of his margin for error:

(LIMIT) Waldo knows that if the claim on page i+1 is not true, then he cannot know that the claim on page i is true.

Quite plausibly, Waldo's knowledge of his margin for error amounts to something slightly stronger:

(LIMIT*) Waldo knows that if the claim on page i+1 is not true, then he cannot know that the claims on pages (i+1)±1 are true.

Assume as before that Waldo knows that the claim on page 1 is true and given (KK) also knows that he knows that the claim on page 1 is true. From this knowledge and Waldo's knowledge of his margin for error, he competently deduces that the claim on page 2 is true, and given (C) he comes to know that the claim on page 2 is true. Waldo continues in this way until he comes to know that the claim on page 26 is true along with all the previous claims in the book. Assume that Waldo now engages in parallel reasoning from his knowledge that the claim on page 50 is true. Given (KK), (LIMIT*) and (C) he can hereby come to know that the claim on page 28 is true along with all the later claims in the book. From the knowledge Waldo has hereby acquired and from his knowledge that one claim in the book is false, Waldo competently deduces that the claim on page 27 is false. Given (C), Waldo knows that the claim on page 27 is false. But presumably that isn't possible.

We can now sum up these further paradoxical aspects of Waldo's situation:

(α) The assumption of (KK), the closure of Waldo's knowledge and (LIMIT*) allows Waldo to come to know the actual truth-value of claims that (given his margin for error) he is not in a position to know.

That is, Waldo can come to know that the claim on page 26 is true and that the claim on page 27 is false, which (given his margin for error) are truths that Waldo is not in a position to know. Furthermore:

(β) The assumption of (KK), the closure of Waldo's knowledge and (LIMIT*) allows Waldo to come to know the actual truth-value of claims that (given his knowledge of his margin for error) he knows he is not in a position to know.

That is, Waldo can come to know that the claim on page 27 is the only false claim in the book, but given Waldo's knowledge of his margin for error, he knows that he is not a position to know exactly which claim in the book is false.

The paradoxical consequences (α) and (β) are just as problematic as the consequence that Waldo can come to know that the false claim on page 27 is true. Furthermore, the result that Waldo can come to know that the false claim on page 27 is true arises from a simple continuation of the reasoning that led to the paradoxical consequence (α) (i.e. of the reasoning that led to the result that Waldo can come to know that the claim on page 26 is true).

Consequently, a diagnosis of how to avoid (α) will also provide us with a diagnosis of how to avoid the result that Waldo can come to know that the false claim on page 27 is true. In my discussion of the plausibility (C), the central focus will therefore be on whether (C) is the assumption that should be given up in order to avoid the paradoxical consequences (α) and (β) in WALDO.

3. An initial refinement of intuitive closure. Before I provide an outline of my reasons for suggesting that (C) rather than (KK) might be the problematic assumption in the above argument, the connection between (C) and a more general closure principle for knowledge should be noted. In Williamson's (2000:117) discussion of his Mr Magoo case, (C)[45] is motivated in the following way:

[45] My version of (C) is slightly different from Williamson's (2000). See my initial statement of (C) for a discussion.

> ”Could we reject the assumption (C) that Mr Magoo’s knowledge of the pertinent propositions is deductively closed? Assumption (C) is true if deduction is a way of extending one’s knowledge: that is, if knowing p_1, …, p_n, competently deducing q, and thereby coming to believe q is in general a way of coming to know q. Call that principle *intuitive closure*. Since by hypothesis Mr Magoo satisfies the conditions for the intuitive closure principle to apply, rejecting (C) is tantamount to rejecting intuitive closure.” (2000:117).

The stated version of *intuitive closure* amounts to some version of multi-premise closure for knowledge, and according to Williamson (2000), this is a principle that we should be hesitant to give up:

> ”We should in any case be very reluctant to reject intuitive closure, for it *is* intuitive. If we reject it, in what circumstances can we gain knowledge by deduction?” (2000:118)

As is made clear above, *intuitive closure* in its multi-premise formulation can be stated as follows:

Intuitive Closure$_{\text{MPC}}$: Knowing p_1, …, p_n, competently deducing q, and thereby coming to believe q is in general a way of coming to know q.

For those skeptical about multi-premise closure for knowledge, Williamson (2000:117-118) argues that his case against (KK) can be reworked so that the deductive closure of Waldo’s knowledge follows from a more restricted, single-premise version of *intuitive closure*:

Intuitive Closure$_{\text{SPC}}$: Knowing p, competently deducing q, and thereby coming to believe q is in general a way of coming to know q.

Although Williamson (2000) thinks we should be hesitant to give up *intuitive closure*, most philosophers will reject it in both its single- and multi-premise formulation. As Hawthorne (2004:33) notes, *intuitive closure* does not allow for the possibility that knowledge of the premises is destroyed at some point during the deductive inference, in which case the subject might not come to know after all:

> "…imagine a case where, at t_1, S knows $p_1,\ldots,p_n$ and starts upon a long deduction on the basis of which he comes to believe q at t_3. At t_2, sometime between t_1 and t_3, S's knowledge of some or all of $p_1,\ldots,p_n$ gets destroyed by misleading counterevidence. We do not intuitively suppose that in such a case the person knows that q at t_3." (2004:33)

Dropping the somewhat unclear qualification that competent deduction from known premises 'in general' is a way of coming to know, Hawthorne (2004:33) therefore proposes the following refined version of multi-premise closure (MPC):

MPC: Necessarily, if S knows $p_1,\ldots,p_n$, competently deduces q, and thereby comes to believe q, while retaining knowledge of $p_1,\ldots,p_n$ throughout, then S knows q.

Hawthorne's (2004:34) statement of single-premise closure (SPC) incorporates the same qualification:

SPC: Necessarily, if S knows p, competently deduces q, and thereby comes to believe q, while retaining knowledge of p throughout, then S knows q.

Insofar as Williamson wants to retain the idea that knowledge is defeasible[46], he should therefore give up *Intuitive Closure*$_{\text{MPC}}$ and *Intuitive Closure*$_{\text{SPC}}$ in favour of MPC and SPC.

4. Outline of an argument. My argument for why (C) rather than (KK) is the problem in WALDO now proceeds as follows:

In section (**5**), I provide an initial discussion of whether paradoxical consequences such as (α) and (β) still arise in WALDO if we assume MPC/SPC instead of Williamson's *intuitive closure*. If (α) and (β) do not arise if we replace *intuitive closure* with MPC/SPC, then (C) rather than (KK) looks like the problematic assumption in WALDO. Although I identify a suggestion for how (α) and (β) might be avoided if we assume MPC/SPC, I ultimately conclude that the considerations in favour of the relevant suggestion are far from decisive. As such, it isn't clear that replacing *intuitive closure* with MPC undermines the plausibility of (C).

In section (**6**), I present a number of cases in which it is clear that Hawthorne's MPC and SPC (as well as Williamson's *intuitive closure*) give rise to paradoxical consequences analogous to (α) and (β) without the involvement of KK or any similar principles. In other words, I show that there are a number of circumstances in which it is clearly the assumption of knowledge closure rather than KK that is responsible for the generation of paradoxical consequences like (α) and (β).

---

[46] It is worth noting that Williamson in some places expresses attraction to a view on which misleading evidence cannot defeat knowledge in the range of cases epistemologists have traditionally been inclined to think. See, e.g., Williamson (*forthcoming*:42). Dialectically the interest of his argument against KK will, however, be significantly reduced if it relies on assumptions that are incompatible with standard views about defeat. At least for that reason it is interesting to see what consequences replacing *intuitive closure* with MPC/SPC has for his argument against KK.

This gives us some reason to suspect that *if* the assumption of MPC/SPC, (KK) and (LIMIT*) does in fact result in the paradoxical consequences (α) and (β) in WALDO, it is MPC/SPC rather than (KK) that is to blame. In other words, since MPC and SPC are responsible for the generation of paradoxical consequences such as (α) and (β) in other cases, there is some attraction to thinking that *if* (α) and (β) still arise in WALDO when we replace *intuitive closure* with MPC/SPC, then MPC/SPC rather than (KK) is responsible. Consequently, there is some attraction to thinking that (C) rather than (KK) should be given up in WALDO.

In section (**7**) I provide further support for this conclusion by presenting a case that has a number of similarities with WALDO, although it does not involve any assumptions about KK.

In the relevant case, Williamson's *intuitive closure* leads to consequences parallel (α) and (β). Just as in WALDO, it is unclear whether consequences like (α) and (β) arise if we assume MPC instead of *intuitive closure*. What is clear, however, is that *if* consequences like (α) and (β) do in fact arise in the relevant case, then MPC is to blame. Given the similarities between the proposed case and WALDO, I suggest that we have reason to opt for an analogous conclusion in WALDO. Consequently, (C) rather than (KK) should be given up in WALDO.

In section (**8**), I note what implications giving up (C) in WALDO might have for the paradox of the *Surprise Examination* and for Williamson's (2011, *forthcoming*) argument for the possibility of improbable knowing.

In section (**9**), I consider a worry for my argument against (C) and argue that it can be overcome.

In the following section, I start by considering whether the paradoxical consequences (α) and (β) arise in WALDO if we replace *intuitive closure* with MPC/SPC.

5. WALDO and MPC/SPC. Replacing *intuitive closure* with MPC/SPC opens the possibility that some of Waldo's competent deductions from known premises result in a loss of knowledge of the premises rather than his gaining knowledge of the conclusion. That is, MPC/SPC is compatible with the claim that if Waldo knows his margin for error and knows that he knows that the claim on page *i* is true, then competently deducing that the claim on page *i*+1 is true could result in a loss of knowledge of one or both of the premises rather than Waldo's gaining knowledge of the conclusion. Could this possibility undermine the plausibility of (C) and point to a way of avoiding the paradoxical consequences (α) and (β) without having to give up MPC/SPC or (KK)?

It does not seem particularly plausible that Waldo's competent deductions could result in a loss of knowledge of his margin for error. If Waldo's competent deductions is to result in a loss of knowledge of one of the premises, the only plausible candidate premise is Waldo's knowledge that he knows the claim on page *i* is true.

Could Waldo's competent deduction to the conclusion that the claim on page *i*+1 is true result in Waldo's losing knowledge that he knows the claim on page *i* is true rather than his gaining knowledge of the conclusion?

Here is one line of thought to the effect that it could. By stipulation of the case, Waldo knows that one of the claims just around the middle of the book is false, but also

knows that he is not in a position to know exactly which claim. In fact, Waldo knows that he cannot know the truth-value of claims close to the middle of the book.

Assuming that page i+1 is close to the middle of the book, Waldo knows that he cannot know that the claim on page i+1 is true. Realizing that the truth of the claim on page i+1 follows from his knowledge of his margin for error and his (known) belief that he knows the claim on page i is true, there is some attraction to the idea that Waldo could hereby lose his knowledge that he knows the claim on page i is true.

It is worth noting, however, that even if Waldo loses his knowledge that he knows the claim on page i is true, this might not be sufficient to prevent the paradoxical consequences (α) and (β) from arising in WALDO. To see why, consider the two different ways in which Waldo could lose his knowledge that he knows the claim on page i is true.

On the one hand, Waldo could lose both his knowledge that he knows the claim on page i is true and his knowledge that the claim on page i is true. On the other hand, Waldo could lose his knowledge that he knows the claim on page i is true while retaining his knowledge that the claim on page i is true.

Let us start by looking at the first option. Even if Waldo loses both his knowledge that the claim on page i is true and his knowledge that he knows, this presumably does not result in Waldo's losing all of his knowledge about the truth-value of claims in the book. That is, Waldo presumably still knows that the claim on page 1 is true. Given MPC, (KK) and Waldo's knowledge of his margin for error, it therefore seems that Waldo can regain his knowledge that the claim on page i is true and by (KK) his knowledge that he knows.

Even if one of Waldo's competent deductions results in his losing knowledge that he knows the claim on page i is true instead of his gaining knowledge that the claim on

page *i*+1 is true, it therefore seems that Waldo is always able to regain knowledge that he knows the claim on page *i* is true.

In order to avoid the paradoxical consequences (α) and (β) while retaining MPC and (KK) we would therefore have to claim that no matter how many times Waldo attempts to redeploy his knowledge of that he knows the claim on page *i* is true in an attempt to come to know that the claim on page *i*+1 is true, his deduction will always result in a loss of knowledge of the premise rather than his gaining knowledge of the conclusion. Assuming that Waldo is extremely persistent in his deductive efforts, the consequence is an *unstable* situation in which Waldo constantly loses and regains his knowledge that he knows the claim on page *i* is true.

Let us now consider the possibility that Waldo's competent deduction to the conclusion that the claim on page *i*+1 is true results in a loss of knowledge that he knows the claim on page *i* is true *without* resulting in a loss of knowledge that the claim on page *i* is true.

Strictly speaking, this suggestion is incompatible with (KK). There is, however, at least one way of understanding the suggestion that is compatible with the spirit of (KK). Specifically, one could understand this suggestion as the idea that while Waldo loses his knowledge that he knows the claim on page *i* is true (e.g. by losing his belief that he knows the claim on page *i* is true), he can regain this higher-order knowledge by reflecting on what he knows. In order for this suggestion to avoid the consequences (α) and (β), one would have to postulate an *unstable* situation in which a persistent Waldo constantly loses and regains his knowledge that he knows the claim on page *i* is true.

It is beyond the scope of the present paper to determine whether any of these *unstable* suggestions give us a plausible way of avoiding (α) and (β) while retaining MPC, Waldo's knowledge of his margin for error and some principle in the spirit of (KK). At this point, it is therefore unclear whether (α) and (β) arise in WALDO if we replace *intuitive closure* with MPC/SPC. Consequently, it is at this stage unclear whether (C) or (KK) is the problematic assumption in WALDO.

6. Paradoxical consequences without KK. In this section I present a number of cases in which MPC and SPC (as well as Williamson's *intuitive closure*[47]) give rise to paradoxical consequences analogous to (α) and (β) without the involvement of KK or any similar principle. This gives us some reason to suspect that *even if* (α) and (β) arise in WALDO when we replace *intuitive closure* with MPC/SPC, it is MPC/SPC rather than (KK) that should be blamed. In other words, it gives us reason to suspect that the problematic assumption in WALDO is the closure requirement build into (C) rather than (KK).

I will start by providing two cases involving MPC[48]:

FALSE TESTIMONY. Assume that a subject, S, knows a number of propositions p_1, p_2, …, p_n. Assume that S is in roughly the same epistemic position with respect to each of the propositions and roughly equally confident in all of them. Imagine, for example, that p_1, p_2, …, p_n are a number of propositions that S has learned by reading a book that is widely

[47] Below I focus on how MPC and SPC leads to consequences like (α) and (β) in the various cases, but it should be clear that the same consequences arise if we assume *intuitive closure* instead.
[48] See Chapter 1 for two almost identical cases.

accepted as trustworthy on the relevant issues. Imagine now that S engages in a competent multi-premise deduction. S considers each of the things he believes on the basis of having read the book and deduces the conjunction:

$p_1, p_2, \ldots, p_n$

$(p_1 \wedge p_2 \wedge \ldots \wedge p_n)$

Assume that S thereby comes to believe the conjunction. We have no particular reason to think that S by the process of deduction loses knowledge of any of the premises, and according to MPC, S therefore comes to know the conjunction.

Imagine now that S receives misleading evidence E with defeating force with respect to S's knowledge of the conjunctive conclusion. Imagine, for example, that someone S considers highly trustworthy convincingly (but falsely) informs S that one of the conjuncts is false. The informant, however, provides no indication of which conjunct. Upon receiving this piece of misleading evidence, we would expect S to give up belief in the conjunctive conclusion and even if S were to retain belief in the conclusion, we are inclined to think that S no longer knows the conclusion.

What, then, about S's knowledge of the premises? If n is very low, we might expect S to give up belief in at least some (if not all) of the conjuncts upon receiving misleading evidence E. Even if S were to retain belief in all of the conjuncts it might seem compelling to think that receiving E will defeat S's knowledge of at least some (if not all) of the conjuncts. Whether or not S continues to believe all of the conjuncts, it is therefore plausible to think that if n is very low, the misleading evidence that defeats knowledge of

the conjunctive conclusion also results in a loss of knowledge of at least some of the premises.

What if n is very high? If so, we would not expect S to give up belief in all of the conjuncts upon being told that one of the conjuncts is false. In fact, given that S is roughly as confident with respect to each of the conjuncts and given that E is not targeted at any particular conjunct, we would not expect S to give up belief in any conjunct. As such, we would expect S to continue believing all of the conjuncts if n is very high.

Does S also retain knowledge of all of the conjuncts upon receiving E? It does not seem particularly plausible to think that receiving E defeats S's knowledge of some of the conjuncts while allowing S to retain knowledge of some of the other conjuncts. After all, E is not targeted at any particular conjunct and S was in roughly the same epistemic position with respect to each of the conjuncts before receiving E. At the same time, the claim that S loses knowledge of all of the conjuncts looks overly skeptical. If n is very high and if S knew all of the conjuncts before being told that one of the conjuncts is false, then we are not inclined to think that S stops knowing all of the conjuncts upon receiving E. As such, there is some attractiveness to the thought that S does in fact retain knowledge of all of the conjuncts upon receiving misleading evidence E[49].

[49] Someone might suggest that if S does in fact retain knowledge of each of the conjuncts upon receiving E, then receiving E must at least have had the effect that S does not know for each of the conjuncts that she knows it. FALSE TESTIMONY is therefore an example of a case in which KK fails. There are a number of reasons why I think this diagnosis is not correct. First of all, it is rather implausible to think that we cannot obtain higher-order knowledge by reading books when we have strong evidence that one of the claims in the book is mistaken. Generalized to our other sources of knowledge, the natural consequence is that we rarely have higher-order knowledge of anything. But surely that can't be right. Combined with the knowledge norm of assertion, this conclusion also has the consequence that we are rarely in a position to assert "I know that p" for some p that we have come to know by reading a book or have come to know by relying on other sources of information of which we have strong evidence that they are occasionally mistaken. Finally, it is worth noting that if S knew for each of the conjuncts that he knew it before receiving E (which seems plausible), then we can run an argument for why S retains higher-order knowledge that is parallel to the argument for

Paradoxical consequences now arise if we consider the possibility that S once again starts thinking about what follows from p_1, p_2, …, p_n and infers the conjunctive conclusion. If S were able to competently deduce the conjunctive conclusion before receiving E, why should S not be able to do so after having received E? We are, after all, considering a fairly simple inference from a set of premises to the conjunction of those premises. Assuming that S still knows each of the premises, it therefore seems that if MPC is true, then S could come to know the conjunctive conclusion even after having received E. This, however, looks problematic, because S now knows the conjunctive conclusion while at the same time having misleading evidence E with assumed defeating force with respect to S's knowledge of the conjunctive conclusion. But presumably that isn't possible[50].

As opposed to WALDO, my FALSE TESTIMONY case involves a subject who is misled or deceived by false testimony from a trustworthy informant. The TRUE TESTIMONY case below differs from FALSE TESTIMONY in exactly this respect:

TRUE TESTIMONY. Assume that a subject, S, has come to believe a number of propositions p_1, p_2, …, p_n about Chinese history by reading a book that is widely accepted as trustworthy on the relevant issues. Assume that S is equally confident in all of p_1, p_2, …,p_n. Assume (contrary to above) that one of the propositions in the book is false. If n is sufficiently large, then we are normally inclined to say that S can know some or many of the true things that S has come to believe by reading the book. After all, denying that we

---

why S retains first-order knowledge of each of the conjuncts. FALSE TESTIMONY should therefore not be interpreted as a case of KK failure.

[50] Hawthorne (2004:184) briefly mentions the possibility of a strategy for handling the tension between MPC and knowledge defeat inspired by discussions of vagueness. For a critical discussion, see Chapter 1.

can acquire knowledge from books that contain a few insignificant falsehoods looks overly skeptical and could lead to widespread skepticism if generalized to our other testimonial and non-testimonial sources of beliefs.

If the false proposition has no significance for how the other propositions in the book are motivated, then it even seems possible that S can know all of the true things that S has come to believe by reading the book[51]. If S reads the entire book, then it is not as if the truth of some of the beliefs S has acquired by reading the book are more lucky or accidental than the truth of other beliefs S has acquired by reading the book. For instance, it does not seem that S's epistemic position is any worse with respect to propositions that are written on pages close to the page containing the false proposition. Plausibly, there are possible circumstances in which S's epistemic position with respect to all of the true propositions is roughly the same.

Imagine now that someone S (rightly) considers highly trustworthy convincingly informs S that exactly one of the propositions S has come to believe by reading the book is false. The informant, however, provides no indication of which proposition. We can assume that the informant does in fact know that one of the propositions in the book is false and that S therefore comes to know by relying on the informant's say-so.

What happens to the beliefs that S has acquired by reading the book? If n is sufficiently large, then we would expect S to behave just as in FALSE TESTIMONY and continue to believe all of the propositions p_1, …, p_n upon being told that one of the propositions is false. Does S's knowledge of the falsity of one the propositions he believes destroy his knowledge of the true propositions he believes on the basis of having read the

[51] It is worth noting that even if one insists that S can only know some, but not all of the truths in the book, MPC still gives rise to paradoxical consequences similar to (α) and (β) in a case like TRUE TESTIMONY. I leave it as an exercise to the reader to spell out the details.

book? It looks overly skeptical to claim that S loses knowledge of all of the true propositions in the book. Even without the testimony of a trustworthy informant, we often know that big books will contain or are likely to contain a few mistakes, but normally this is not thought of as something that prevents us from coming to know by reading big books. Similarly, most of us realize that we have at least a few false beliefs, but this is rarely seen as something that prevents whatever true beliefs we might have from being known.

And although we can imagine scenarios in which S's knowledge that one of the propositions in the book is false destroys S's knowledge of *some* of the true propositions in the book, there is no reason to think that it necessarily must. If S is in roughly the same epistemic position with respect to each of the propositions before learning that there is a false proposition in the book, and if S does not lose knowledge of all the true propositions by learning of the existence of a falsehood, then it is unclear why S must necessarily lose knowledge of *any* of the true propositions in the book. As such, there is some inclination to think that S can retain knowledge of all the true propositions in the book upon learning that one of the propositions in the book is false [52].

What can S learn by deduction in a situation like this? If n is sufficiently large, then there is some inclination to think that S can come to know small conjunctions of propositions from the book by deducing them from her knowledge of individual propositions. If (by reading the book) S has come to know that Shanghai is the largest city in China and if S has come to know that Shanghai is not the capital of China, then it seems plausible that competent deduction will allow S to know that Shanghai is the largest city in

[52] Notice the parallel between TRUE TESTIMONY and a lottery with one winning ticket. Some might think that the parallel puts pressure on the plausibility of the claim that S can know all the true propositions in the book. Others might see the parallel as some support for the idea that if the number of lottery tickets is sufficiently large, then we can in fact know of a losing ticket that it is a losing ticket (although this need not put us in a position to know that the winning ticket is in fact a winner). For a discussion, see Chapter 1.

China but not the capital. Nonetheless, it seems that S's knowledge of the falsity of one of the propositions in the book has consequences for what S can come to know by competent deduction. Consider the possibility that S thinks about all the things he has come to believe by reading the book. S knows that one of the propositions is false (although he does not know which one), and by competent deduction S comes to believe the conjunction of all but one of the propositions he has come to believe by reading the book. As it turns out, all of the conjuncts are true. We are assuming that S knows all of the true propositions he has come to believe by reading the book, and so S has come to believe the conjunction on the basis of competent deduction from known premises. According to MPC, S therefore knows the conclusion, but just as in FALSE TESTIMONY, that result looks implausible.

The implausibility of the idea that S can come to know the conjunction by deduction is further supported by the observation that if S can know the conjunction of the true propositions in the book, then by competent deduction S can come to know which of the propositions in the book is false. Assume that p_1, …, p_n are all the propositions in the book and p_1,…, p_{n-1} are all the true propositions in the book. Given S's knowledge that one of the propositions in the book is false, we can assume that S knows that $p_1 \wedge \ldots \wedge p_n$ is false. If S by competent deduction can come to know that $p_1 \wedge \ldots \wedge p_{n-1}$, then from this knowledge and his knowledge that $p_1 \wedge \ldots \wedge p_n$ is false, S can competently deduce that p_n is false. Assuming MPC, S can therefore come to know that p_n is false, but presumably that is impossible[53].

[53] One could try to defend the conclusion that S can come to know $p_1 \wedge \ldots \wedge p_{n-1}$ as well as the falsity of p_n by interpreting TRUE TESTIMONY as a case of improbable knowing. That is, one could try to defend the conclusion that S can come to know $p_1 \wedge \ldots \wedge p_{n-1}$ as well as the falsity of p_n by claiming that these would be cases in which S knows $p_1 \wedge \ldots \wedge p_{n-1}$ and knows not-p_n although it is incredibly unlikely on S's evidence that S knows. Even with the idea of improbable knowing on the table, I think many will find the conclusion that S can come to know that p_n is false extremely unpalatable. Furthermore, taking TRUE TESTIMONY to be a

We can now see that FALSE TESTIMONY and TRUE TESTIMONY have paradoxical consequences analogous to (α) and (β) in WALDO. A kind of analogue of (α) follows straightforwardly:

(α*) The assumption of knowledge closure (MPC) allows S to know certain truths that S is not in a position to know.

In FALSE TESTIMONY, MPC allows S to come to know the conjunction $p_1 \wedge \ldots \wedge p_n$, which his possession of misleading evidence E with defeating force against knowledge of the conjunction makes impossible. In TRUE TESTIMONY, MPC allows S to come to know the conjunction $p_1 \wedge \ldots \wedge p_{n-1}$ and also allows S to come to know that p_n is false, both of which S quite plausibly is not to be in a position to know.

Furthermore, we can assume that in both FALSE TESTIMONY and TRUE TESTIMONY S knows that his knowledge is limited in this way. In FALSE TESTIMONY, S knows that the possession of E prohibits him from coming to know the conjunction $p_1 \wedge \ldots \wedge p_n$. In TRUE TESTIMONY, S's knowledge that one of $p_1, \ldots, p_n$ is false allows him to know that he is not in a position to know the conjunction of all of the truths in the book and it also allows him to know that he is not in a position to know which claim in the book is false. Both cases therefore also give rise to analogues of (β):

---

case of improbable knowing naturally leads to the conclusion that FALSE TESTIMONY is a case of improbable knowing, but the idea that S in FALSE TESTIMONY can know the conjunction $p_1 \wedge \ldots \wedge p_n$ despite expert testimony that $p_1 \wedge \ldots \wedge p_n$ is false is very hard to accept. Finally, I argued above that S in FALSE TESTIMONY plausibly continues to know that he knows each of $p_1,\ldots,p_n$ after he has received evidence that $p_1 \wedge \ldots \wedge p_n$ is false. By a parallel line of reasoning, S in TRUE TESTIMONY plausible also continues to know that he knows each of $p_1,\ldots,p_{n-1}$ even after he has been told that $p_1 \wedge \ldots \wedge p_{n-1}$ is false. If this assumption about S's higher-order knowledge is combined with Williamson's (2000, chap. 9) identification of a subject's evidence with all and only the propositions known by the subject, then it becomes somewhat unclear how to turn TRUE TESTIMONY into a case of improbable knowing.

(β*) The assumption of knowledge closure (MPC) allows S to know certain truths that S knows he is not in a position to know.

There is even reason to think that WALDO and TRUE TESTIMONY share more than similar paradoxical consequences. If the number of pages in Waldo's book with extraordinary facts had been sufficiently high, then it seems that WALDO would in fact have been a special case of TRUE TESTIMONY. The only difference between the cases would be that whereas S in TRUE TESTIMONY is merely told that one of the claims in the book is false, Waldo also has information about the area of the book that contains the falsehood.

If my diagnosis of TRUE TESTIMONY is right, this difference is crucial. Whereas S in TRUE TESTIMONY can know the truth of claims immediately preceding and following the falsehood in the book, Waldo's knowledge of the location of the falsehood would make it epistemically problematic for him to form beliefs about the truth-value of claims on pages close to the falsehood. As such, the fact that Waldo knows roughly where in the book the falsehood is located would prevent him from knowing the truth-value of claims in the book that he could otherwise have known (on the assumption that the number of claims in the book had been sufficiently high[54]).

Because Waldo knows more about the location of the falsehood than S in TRUE TESTIMONY, the mere assumption of *intuitive closure* would not be enough to generate paradoxical consequences such as (α) and (β) in WALDO even if the number of claims in the book had been high. Nonetheless, the relation between WALDO and TRUE

[54] As I mentioned in my initial presentation of WALDO, Waldo plausibly would not have known the truth-value of any of the 50 claims in the book if he had not been informed about roughly where in the book the falsehood is located. The number of claims in *"50 Extraordinary Facts"* is therefore not sufficiently high.

TESTIMONY raises the suspicion that what Williamson's (2000:114-134) *reductio* actually brings out is that WALDO can be turned into a problem case for *intuitive closure* parallel to TRUE TESTIMONY if we add the assumption of (KK) and Waldo's knowledge of his margin for error.

Whereas FALSE TESTIMONY and TRUE TESTIMONY rely on MPC to generate paradoxical consequences analogous to (α) and (β), we can also construct cases that rely on SPC to generate similar paradoxical consequences. Consider the following set of cases:

DRUG (I)[55]. Harry works at a pharmaceutical company that has recently invented a drug that can affect people's ability to make competent deductions. The drug is designed so that it only affects people's ability to make competent deductions from a limited set of propositions (the *reactive* propositions). The drug works in the following way. If a subject affected by the drug considers what follows from a reactive proposition, then she will draw grossly mistaken inferences although they will seem obviously true to the subject. The mistaken inferences always result in falsehoods. Harry is familiar with the drug and knows that it has this effect on exactly 99% of the population. Harry, however, has not been tested and therefore does not know whether he is among the 99%. As a matter of fact, he isn't.

Assume that Harry knows that p and also knows that p is a reactive proposition. On the basis of testimony from a trustworthy colleague, Harry comes to know that some of the other employees have slipped the drug into Harry's morning coffee. Harry starts thinking about what follows from p, competently deduces q and thereby comes to believe q. In fact,

[55] See, e.g., Christensen (2010:187), Hawthorne (2004:4), and Vogel (1990) for somewhat similar cases.

it seems obvious to Harry that q follows from p. According to SPC, Harry knows q. But presumably Harry cannot come to know q given his knowledge that he has been given the drug.

This conclusion is further supported by the following considerations. Plausibly, Harry knows that if the drug affects him and if some proposition seems to him to follow obviously from p, then that proposition is false. Harry also knows that q seems to him to follow obviously from p. If Harry knows that q, then (by *modus tollens*) Harry can come to know that the drug does not affect him and thereby come to know that he is among the 1% of the population that is unaffected by the drug. But presumably he cannot come to know that[56].

DRUG (II). Exactly as in DRUG (I) except that the drug affects 100% of the population. Harry knows this, but unlike before he does not know which propositions are reactive. Assume that Harry is given a piece of paper containing a list of propositions $p_1,\ldots,p_n$ and informed that they are all true. Harry is informed that only one of the propositions on the list is *not* reactive, although he is not told which one. As before, Harry knows that he has been given the drug. Harry now looks at the list and reads one of the propositions, p_3, to himself. Harry thereby comes to know p_3. Coincidently, p_3 is the only non-reactive proposition on the list. Harry starts thinking about what follows from p_3 and competently deduces that q. In fact, it seems obvious to Harry that q follows from p_3. By reasoning

[56] As in the case of TRUE TESTIMONY, one could try to defend the idea that Harry can come to know that he is among the 1% unaffected by the drug by interpreting DRUG (I) as a case of improbable knowing. Just in the case of TRUE TESTIMONY, I suspect that even with the idea of improbable knowing on the table, many will find the conclusion that Harry can come to know he is among the unaffected 1% extremely unpalatable.

analogous to DRUG (I), Harry comes to know that p_3 is the only non-reactive proposition on the list. But presumably he cannot come to know that.

DRUG (III). Exactly as DRUG (II) except that only one of the propositions $p_1,\ldots,p_n$ is reactive and Harry knows this (although he does not know which one). Assume that p_4 is the only reactive proposition. By the same procedure as above, Harry considers every proposition but p_4 and comes to know that they are not reactive. From this, he competently deduces that p_4 must be the reactive proposition on the list. But presumably he cannot come to know that.

DRUG (IV). Exactly as DRUG (III) except that Harry has been told (correctly) that the only reactive proposition is one of propositions at the middle of the list (although Harry does not know which one). Plausibly, Harry knows that the first proposition on the list, p_1, is not reactive. Just as Waldo, Harry also knows that there is a limit to what he can know about the specific location of the reactive proposition on the list. Plausibly, Harry knows that if p_{i+1} is a reactive proposition, then he cannot know that p_i is not a reactive proposition. If we assume (KK), Harry's situation is exactly like that of Waldo. Even without (KK), however, SPC allows Harry to reason as in DRUG (III) and come to know exactly which proposition is reactive, which is incompatible with Harry's margin for error.

DRUG (I) through DRUG (IV) are all cases where the assumption of SPC gives rise to paradoxical consequences parallel to (α) and (β). In all these cases, SPC allows Harry to come to know something that Harry quite plausibly is not in a position to know.

Furthermore, Harry knows that his knowledge is limited in this way and SPC therefore allows Harry to come to know something that he knows he is not in a position to know.

As I argued in section (**5**), it is not entirely clear that the paradoxical consequences (α) and (β) arise in WALDO if we replace Williamson's *intuitive closure* with MPC/SPC. If (α) and (β) do arise, however, then the above cases give us some reason to suspect that it is exactly the assumption of MPC/SPC rather than (KK) that is responsible for the generation of (α) and (β). That is, if MPC and SPC are responsible for the generation of paradoxical consequences such as (α) and (β) in other cases, then there is some attraction to thinking that MPC/SPC rather than (KK) is also responsible if (α) and (β) arise in the case of WALDO. As such, the above cases give us some reason to suspect that (C) rather than (KK) is the problematic assumption in WALDO.

7. An analogue of WALDO? In this section, I provide further support for the above conclusion by considering a case that in various ways is similar to WALDO. Although the case uses the same setup as TRUE TESTIMONY[57], it does not involve knowledge of large conjunctions:

DEDUCTIVE STEPS. Assume that a subject, S, has read the following propositions in a book: p entails r_1, r_1 entails r_2, r_2 entails r_3, ..., r_{n-1} entails r_n, and r_n entails q. The book is considered trustworthy and S therefore believes each of the propositions. Exactly one of the propositions is false, and someone S trusts informs him that this is so. He therefore

[57] Those worried about TRUE TESTIMONY can restate the below case using the setup from FALSE TESTIMONY.

comes to know that one of the propositions is false, although he does not know which one. In fact, the last proposition in the book, r_n entails q, is the only false proposition in the book. Assuming that n is sufficiently high, we can assume (parallel to TRUE TESTIMONY) that S can nonetheless know the true propositions in the book. At the same time, we can assume that S (parallel to TRUE TESTIMONY) knows that he is not in a position to know the conjunction of all the true propositions (forming a true belief in a conjunction of all but one of the propositions that S believes on the basis of having read the book is too lucky to constitute knowledge). Furthermore, we can assume that S knows that he cannot by competent deduction come to know something that requires him to use all the true propositions as premises (again, the truth of the resulting belief would be too lucky to constitute knowledge).

Assume that S knows p with certainty. From his knowledge that p and his knowledge that p entails r_1, S competently deduces that r_1 and thereby comes to believe r_1. If we assume *Intuitive Closure*$_{\text{MPC}}$, S knows r_1. From his knowledge that r_1 entails r_2 and his knowledge that r_1, S can now come to know r_2 by competent deduction. Continuing in this way, S can by a long number of deductive steps come to know that r_n. But as noted above, the truth of S's belief that r_n is too lucky to constitute knowledge and S knows this. The assumption of *Intuitive Closure*$_{\text{MPC}}$ therefore results in paradoxical consequences parallel to (α) and (β) in WALDO.

What if we assume MPC rather than *Intuitive Closure*$_{\text{MPC}}$? Assume as above that S starts out from his belief that p and his belief that p entails r_1, and via a number of competent deductive steps reaches the conclusion that r_n. Does S's belief that r_n amount to knowledge if we replace *intuitive closure* with MPC?

First, we have to determine whether one of S's deductive inferences could result in a loss of knowledge of the premises rather than a loss of knowledge of the conclusion. There is clearly reason to believe that one of the deductive inferences will not lead to knowledge of the conclusion. In fact, S cannot know r_n and knows that he cannot know r_n. The question, however, is whether we have reason to think that any of the deductive inferences result in a loss of knowledge of the premises.

In each deductive inference, the premises of the inference consist of two propositions. The first is a proposition that S has learned by reading the book. In the first inference, the other premise is p. In the remaining inferences, the other premise is a proposition S has come to believe through a number of deductive steps leading back to p and the proposition that p entails r_1 (i.e. the other premise is r_1 or r_2 or … or r_{n-1}) . For each premise, we should now consider whether S's deductive inference could result in a loss of knowledge of that premise.

Could any of the deductive inferences result in a loss of knowledge of the premise that S has learned by reading the book? Plausibly not. If any of the deductive inferences were to destroy S's knowledge of that premise, then it would plausibly also destroy S's knowledge of all the other propotions S has learned by reading the book. This is so, because S is in roughly the same epistemic position and roughly equally confident with respect to each of the propositions S has learned by reading the book, and the deductive inferences do not seem to effect the plausibility of any one of these propositions more than the plausibility of the others. Holding that one of the deductive inferences destroys S's knowledge of all of the propositions S has learned by reading the book does, however, look like an overly skeptical result if we assume that n is sufficiently large. It therefore seems

implausible that any of the deductive inferences should result in S's losing knowledge of the premise of the inference that S has learned by reading the book.

Could any of the deductive inferences result in a loss of knowledge of the other premise (i.e. in p or r_1 or r_2 or … or r_{n-1})? Given that S is certain about p, the deductive inference involving the premise p (and the premise that p entails r_1) could plausibly not result in S's losing his knowledge that p. Could any of the subsequent inferences result in S's losing knowledge that r_1 or r_2 or … r_{n-1} is true?

S's knowledge that one of the propositions in the book is false provides S with evidence that if most of the propositions from the book has been used in reaching a certain conclusion, then it is overwhelmingly likely that the false proposition has been used in reaching that conclusion. It is somewhat unclear, however, whether realizing that the deduced conclusion is too unlikely to be known destroys S's knowledge of the premise that S has *not* come to believe by reading the book (i.e. whether it destroys S's knowledge of r_1 or r_2 or … or r_{n-1}).

Consider both the alternative that it doesn't destroy knowledge of that premise and the alternative that it does. If it doesn't, then the series of deductive inferences will give S knowledge that r_n if we assume MPC. As such, DEDUCTIVE STEPS would have paradoxical consequences parallel to (α) and (β) in WALDO.

If S does in fact lose knowledge of the premise that S does not have from the book (i.e. the premise r_1 or the premise r_2 or … or the premise r_{n-1}), then it seems that S could start over from his knowledge that p and his knowledge that p entails r_1 and through a series of deductive steps once again come to know the relevant premise. Assuming MPC, we can therefore only avoid that S can come S to know r_n if we insist that every time S uses the relevant proposition as a premise, S loses knowledge of the premise instead of

gaining knowledge of the conclusion. Assuming S is extremely persistent, we would therefore have to assume an *unstable* situation where S constantly loses and regains knowledge of the relevant premise (i.e. a situation where S constantly loses and regains knowledge that r_1 or knowledge that r_2 or … or knowledge that r_{n-1}).

We can now note a number of similarities between DEDUCTIVE STEPS and WALDO. Just as Waldo knows that the claim on page 1 is true, S knows that p is true. Through a series of deductive steps, Waldo reaches the conclusion that the claim on page 26 is true whereas S reaches the conclusion that r_n is true. If *intuitive closure* were true, each deductive step would provide Waldo and S with the knowledge necessary to make another deductive step. In both cases, it is somewhat unclear whether replacing *intuitive closure* with MPC has the consequence that some deductive inference results in a loss of knowledge of a premise rather than a gain of knowledge in the conclusion. If the deductive inferences in DEDUCTIVE STEPS still result in S's coming to know that r_n if we replace *intuitive closure* with MPC, then MPC is clearly to blame.

Even though (KK) is necessary to generate the paradoxical consequences (α) and (β) in WALDO, the similarities between WALDO and DEDUCTIVE STEPS provide us with some reason to suspect that if (α) and (β) still arise in WALDO when we replace *intuitive closure* with MPC, then MPC rather than (KK) is to blame. As such, the similarities between WALDO and DEDUCTIVE STEPS give us some reason to suspect that (C) rather than (KK) is the problematic assumption in WALDO.

8. The Surprise Examination and improbable knowing. If it turns out that (C) rather than (KK) is the problem in WALDO, then there are a number of potential epistemological

consequences worth noting. Two of the most straightforward consequences concern the paradox of the *Surprise Examination* and Williamson's (2011, *forthcoming*) argument for the possibility of improbable knowing.

In Chapter 6 of *Knowledge and Its Limits*, Williamson argues that the case of Mr Magoo and the paradox of the *Surprise Examination* belong to a set of paradoxical cases that due to structural similarities should receive a similar diagnosis. Since Williamson takes KK to be the problem in the case of Mr Magoo, he argues that the paradox of the *Surprise Examination* similarly arises from a problematic reliance on KK. If closure rather than (KK) is the problem in both WALDO and Williamson's original Mr Magoo case, then there might be some plausibility to the idea that the paradox of the *Surprise Examination* arises from a problematic reliance on knowledge closure rather than KK.

In his (2011, *forthcoming*), Williamson gives an argument for the possibility of improbable knowing. That is, Williamson argues that cases are possible in which a subject S knows that p although the probability on S's evidence that S knows that p is arbitrarily close to 0. In fact, Williamson (*forthcoming*) argues that cases are possible in which S knows that p while at the same time knowing that on S's evidence the probability that S knows that p is arbitrarily close to 0.

In these arguments the closure of knowledge is not stated as an explicit premise, but follows instead from the assumption of logical omniscience build into the formal models of epistemic logic that Williamson applies to a case of inexact perceptual knowledge to exemplify the phenomenon of improbable knowing. While Williamson is very well aware that logical omniscience is an extreme idealization of any human agent, he takes the idealization to be harmless for the purposes of his argument, because, roughly speaking, "…bad logic cannot compensate for bad eyesight" (Williamson *forthcoming*:6).

The line of thought motivating improbable knowing is, however, an extension of the same style of reasoning that motivates Williamson's (2000) rejection of (KK) in the case of Mr Magoo, and if knowledge closure rather than (KK) is in fact the problem in the Mr Magoo case, then the possibility of improbable knowledge plausibly follows from the assumption of logical omniscience rather than from our need for margins for error in our perceptual knowledge.

9. A familiar object of ridicule? So far I have provided some reasons for thinking that the assumed closure of Waldo's knowledge rather than (KK) is to blame for the paradoxical consequences in WALDO. The central claim of my argument is that the paradoxical consequences (α) and (β) in WALDO might disappear if we replace Williamson's *intuitive closure* with MPC/SPC, but if they don't, then the problem is probably MPC/SPC rather than (KK).

There is, however, a worry for the idea that *if* (α) and (β) still arise in WALDO when we replace *intuitive closure* with MPC/SPC, then MPC/SPC rather than (KK) is to blame.

To spell out the worry, notice that if we attempt to avoid (α) and (β) by giving up both *intuitive closure* and MPC/SPC rather than (KK), then we are forced to claim that one of Waldo's competent deductions fails to give knowledge of the conclusion although Waldo retains knowledge of the premises throughout the deductive process. That is, we are forced to claim that there is a value of i such that Waldo's competent deduction from (KB_i) his knowledge that he knows the claim on page i is true and (LIMIT) his knowledge of his margin for error doesn't provide him with (B_{i+1}) knowledge that the claim on page i+1 is true and doesn't destroy Waldo's knowledge of the premises either.

This conclusion seems to have a rather embarrassing consequence if combined with the idea that knowledge is the norm of assertion[58]. To see why, consider the following line of reasoning.

The deduction from (KB_i) and (LIMIT) to (B_{i+1}) is not very complicated and we can therefore assume that the entailment is obvious and very much transparent to Waldo. Assuming that Waldo adheres to the knowledge norm of assertion, the denial of *intuitive closure* and MPC/SPC results in a rather peculiar behavior by Waldo if asked about the truth-value of claims in the book. Suppose we ask Waldo whether he agrees that (KB_i) and (LIMIT). He asserts that he does: "Yes," he says. We then ask Waldo whether he realizes that (B_{i+1}) follows from (KB_i) and (LIMIT). "Yes," he says. We then ask him whether he agrees that (B_{i+1}). "I'm not agreeing to that," he says. We ask him whether he now wishes to retract his earlier claims. "Oh no," he says. "I'm sticking by my claim that (KB_i) and (LIMIT) and my claim that (KB_i) and (LIMIT) entail (B_{i+1}). I'm just not willing to claim that (B_{i+1})."

Paraphrasing Hawthorne (2005), we can see how Waldo resembles perfectly Lewis Carroll's Tortoise, "…that familiar object of ridicule who was perfectly willing to accept the premises of a *modus ponens* argument, but was unwilling to accept the conclusion" (Hawthorne 2005:32). If we embrace the thesis that knowledge is the norm of assertion, then avoiding (α) and (β) by giving up *intuitive closure* and MPC/SPC will turn Waldo into a facsimile of Carroll's Tortoise.

---

[58] In stating this potential worry I rely heavily on Hawthorne (2005: 32)

It may be suggested that this potential consequence of blaming knowledge closure rather than (KK) for the generation of (α) and (β) in WALDO is a strong point in favour of retaining (C) and giving up (KK). There are, however, a couple of reasons why I don't think we should worry too much about turning Waldo into a facsimile of Carroll's Tortoise.

First of all, FALSE TESTIMONY and TRUE TESTIMONY have exactly the same consequence when combined with the knowledge norm of assertion. In both cases, we can ask the subject of the case whether he agrees that p_1, agrees that p_2, …, agrees that p_{n-2}, and agrees that p_{n-1}. Adhering to the knowledge norm of assertion, he will assert that he does: "Yes," he says. We then ask him whether he realizes that $p_1 \wedge \ldots \wedge p_{n-1}$ follows from $p_1,\ldots,p_{n-1}$. The subject of the case merely has to realize that the truth of a conjunction follows from the truth of the conjuncts, and we can therefore assume that he says "Yes". We then ask him whether he agrees that $p_1 \wedge \ldots \wedge p_{n-1}$. "I'm not agreeing to that," he says. We ask him whether he now wishes to retract his earlier claims. "Oh no," he says. "I am sticking by my claim that p_1, my claim that p_2, …, my claim that p_{n-2}, my claim that p_{n-1}, and my claim that $p_1 \wedge \ldots \wedge p_{n-1}$ follows from these. I'm just not willing to claim that $p_1 \wedge \ldots \wedge p_{n-1}$."

Similarly, we can easily construct versions of the DRUG cases in which a subject adhering to the knowledge norm of assertion will accept the premises of a simple and transparent *modus ponens* argument, but will be unwilling to accept the conclusion despite having competently deduced it from the premises.

Finally, the situation in DEDUCTIVE STEPS is exactly parallel to WALDO. On the one hand, (α) and (β) might not arise in DEDUCTIVE STEPS if we replace *intuitive closure* with MPC, in which case the above worry does not arise. On the other hand, we can't rule

out that (α) and (β) arise on the assumption that MPC is true, in which case the only way of avoiding (α) and (β) will be to claim that one of S's competent deductions does not result in knowledge of the conclusion of a simple and transparent argument, but also doesn't destroy S's knowledge of the two premises. On this second diagnosis of DEDUCTIVE STEPS, S is also turned into a facsimile of Carroll's Tortoise on the assumption that knowledge is the norm of assertion.

Cases where subjects behave roughly like Carroll's Tortoise in an attempt to adhere to the knowledge norm of assertion therefore looks like something we have to come to terms with. And given the similarities between the various problem cases for MPC and SPC on the one hand and WALDO on the other, it should not surprise us if Waldo, like the subjects of those other cases, is led to a rather peculiar behavior in his attempts to adhere to the knowledge norm of assertion.

Another reason not to be too worried about this peculiar potential consequence of giving up (C) rather than (KK) in WALDO is that the denial of (KK) has its own peculiarities when combined with the knowledge norm of assertion. If we give up (KK), then there is some value of *i* such that Waldo knows that the claim on page *i* is true, but does not know that he knows. Suppose I ask Waldo whether he can tell me anything about the truth-value of claims in the book. Adhering to the knowledge norm of assertion he says: "Well, I can tell you that the claim on page *i* is true." I ask him whether he actually knows that the claim on page *i* is true. "I am not sure," he says. I ask him whether this should be seen as a retraction or modification of his previous claim. "Oh no," he says. "I am not retracting anything. The claim on page *i* is true, I am just not sure I know that the claim on page *i* is true". At least to my ear, that sounds somewhat peculiar.

It is also worth noticing that if we accept Williamson's (2000) diagnosis that (KK) is the problem in WALDO, then Williamson's (2011, *forthcoming*) arguments for the possibility of improbable knowing might also look persuasive[59]. That is, giving up (KK) in WALDO seems to go quite naturally with the idea that we can know that p while at the same time knowing that on our evidence the probability that we know that p is arbitrarily close to 0. This, of course, gives rise to even more peculiar results when combined with the knowledge norm of assertion. Suppose again that I ask Waldo if he can tell me something about the truth-value of claims in the book[60]. Adhering to the knowledge norm of assertion he says: "Well, I can tell you that the claim on page *i* is true." I ask him whether he actually knows that the claim on page *i* is true. "Considering my evidence, it really doesn't look like it" he says. I ask him whether this is a retraction of his earlier claim. "Oh no," he says. "I am not retracting anything. The claim on page *i* is true, but given my evidence I am pretty sure I don't know that the claim on page *i* is true".

As Cohen & Comesaña (2013:24-5) notes, there is something almost Moorean about such sentences, and at least in ordinary talk they seem rather odd. Giving up (KK) in WALDO therefore does not avoid a commitment to somewhat peculiar and uncomfortable sounding assertions if knowledge is the norm of assertion.

10. Conclusion. There are compelling reasons for thinking that one of Williamson's (2000) central arguments against the KK principle reveals problems with his preferred closure principle for knowledge rather than problems with KK.

---

[59] See, however, Goodman (2013) for an author who thinks that Williamson's style of argument against KK cannot be straigtforwardly extended to an argument for improbable knowing.

[60] Even if WALDO does not give rise to very improbable knowing, a situation parallel to the one described here will arise in other cases if Williamson (2011, *forthcoming*) is right.

CHAPTER 3

Safety and Defeasibility

1. Introduction. The prevailing trend in contemporary epistemology is to think of knowledge and justification as defeasible: one can lose knowledge or justification for one's beliefs by receiving misleading evidence. Purely externalist theories of knowledge and justification are often thought to have difficulties accounting for our loss of knowledge or justification in the face of misleading evidence. Typically, however, the arguments for the tension between externalism and defeat are not spelled out in detail or depend on features specific to the brand of externalism discussed. An inherent tension between externalism and defeat has not been established.

In this paper I provide a detailed study of how a simple safety account of knowledge interacts with the idea that knowledge is defeasible. In recent years, safety has been one of the more prominent proposals for an externalist condition on knowledge[61], and some have even expressed attraction to the view that knowledge just is belief safe from error[62]. The argument for a tension between safety and defeasibility turns out to be far from straightforward. Despite a range of subtle responses available to the safety theorist, I ultimately conclude that in a number of cases our judgments about defeat cannot be explained by reference to failures of safety.

[61] See, e.g., Pritchard (2005), Sosa (1999, 2000, 2007), Williamson (2000, 2009a,b).
[62] See Lasonen-Aarnio (2010) and Williamson (2009a,b).

2. Some cases motivating defeat. In the following I focus exclusively on knowledge defeat and leave the discussion of justification defeat for another occasion. Let me start by providing two cases[63] to motivate the idea that knowledge is defeasible:

> *Red Lighting*[64] *– Part 1*
>
> At time t_1 Suzy is looking at a red object in front of her and based on her experience she forms the belief that the object in front of her is red. There is nothing abnormal about Suzy's eyesight or the lighting in the room.

> *Premier League Reports*[65] *– Part 1*
>
> *The Times* and *The Guardian* are both extremely reliable in reporting Premier League results and Jack tends to trust them equally. Man U just won an important match against Chelsea and on the front page of *The Guardian* Jack finds the Man U victory reported. He therefore forms the belief that Man U won.

When presented with these cases the inclination is to say that both Suzy's belief and Jack's belief amount to knowledge. Consider now the following developments:

> *Red Lighting – Part 2*
>
> At time t_2 a person, Sally, whom Suzy trusts and considers highly reliable tells Suzy that the object she was looking at at t_1 was illuminated by red lighting that would make objects of any colour look red.

> *Premier League Reports – Part 2*
>
> Shortly after reading *The Guardian*, Jack reads on the front page of *The Times* that Man U lost the match.

---

[63] Since the provided formulations of the cases are fairly schematic, there will of course be ways of filling in the details of the cases that will encourage different readings than the ones I rely on here. On natural ways of filling in the details of the cases, the readings I suggest should, however, be fairly uncontroversial.
[64] This case is a version of Lasonen-Aarnio's (2010:1) *Trick on Suzy*.
[65] This case is a modified version of a case from Hawthorne (2004:71).

In light of Suzy's misleading conversation with Sally and Jack's misleading encounter with *The Times* we are inclined to say that even if Suzy still believes that the object was red and even if Jack still believes that Man U won the match, these beliefs do not amount to knowledge.

The standard way for philosophers to think about cases like *Red Lighting* and *Premier League Reports* is to see them as cases exemplifying the phenomenon of knowledge defeat. That is, philosophers tend to think of Suzy as having lost knowledge as a consequence of the misleading testimony from Sally and to think of Jack as having lost[66] knowledge as a result of his misleading encounter with *The Times*[67].

Purely externalist theorist of knowledge are often thought to have difficulties with accounting for the loss of knowledge in cases like the above[68]. Consider, for example, a simple process reliabilist view according to which knowledge is true belief acquired by a reliable process. How does such a view handle *Red Lighting*? Suzy's belief that the object in front of her is red is acquired by a reliable perceptual process or mechanism. Even if she continues to believe that the object is red after having been misled about the lighting

---

[66] Some readers may object that they were only inclined to attribute knowledge to Jack in *Part 1*, because they assumed that Jack's immediate environment didn't contain other newspaper with false match reports or similar potential sources of misleading evidence. The information in *Part 2* therefore reveals that their initial attribution of knowledge to Jack was false rather than that Jack has lost knowledge as a result of reading *The Times*. Personally, I find this evaluation of the case overly skeptical. The fact that other newspapers might have made a misprint does not prevent me from coming to know by reading a newspaper that hasn't. Those who remain worried about the case should run the below argument with cases that they take to exemplify knowledge defeat more clearly.

[67] Some authors have recently suggested that perhaps knowledge is not lost in the face of misleading evidence of the kinds received by Suzy and Jack in *Part 2* (see Lasonen-Aarnio 2010 and Hawthorne & Srinivasan 2013). In fact, Lasonen-Aarnio (2010) uses the supposed inability of externalist theories of knowledge to accommodate our defeat judgments in cases like the above as a motivation for the idea that those judgments might be mistaken. If my argument in the following is correct, then proponents of a simple safety account would indeed have to claim that our defeat judgments in cases like the above are mistaken.

[68] See, e.g. Alston (1988:238-9), Bonjour (1980:59-60, 1985:34-57), Goldman (1979:18-20, 1986:62-63,111-113), Lasonen-Aarnio (2010:3-8).

conditions, the belief is still acquired via a reliable mechanism and it might therefore seem that Suzy satisfies the process reliabilist conditions on knowledge. Consequently, it might seem that a simple process reliabilist theory of knowledge is unable to accommodate the judgment that Suzy at t_2 loses her knowledge that the object is red.

Externalists have proposed a number of different measures in the attempt to respond to considerations like the above. Some have complicated their externalist conditions on knowledge (e.g., Goldman 1979:20), some have invoked no-defeat clauses (e.g., Goldman 1986:63), some have opted for accounts that combine internalist and externalist conditions on knowledge, and some have taken yet other measures.

While these reactions are often necessary (although perhaps not sufficient), it remains an open question how general the tension between externalism and defeat actually is. Sometimes, the arguments for the tension between an externalist theory and the defeasibility of knowledge trade on features specific to the externalist theory in question (such as a particular combination of reductionist and naturalist commitments). And sometimes, the arguments are not spelled out in detail, leaving unanswered a number of possible responses and objections.

Consider, for example, the worry raised for process reliabilism above. One might have thought that a plausible statement of process reliabilism should involve the requirement that the target belief must be both acquired *and retained* by a reliable process or mechanism. Furthermore, it should be clear that for any belief a subject might hold, there are a variety of different ways of characterizing the process, mechanism or way in which the subject has come to hold the relevant belief. And although the belief is reliably formed and retained according to *one* of these characterizations, this should not be enough to satisfy the process reliabilist conditions on knowledge. In other words, there are ways of

characterizing a subject's belief formation and retention that are relevant to the question of whether the subject's belief satisfies the process reliabilist requirements, and there are plausibly other characterization that are not relevant.

But if that is true, then one might wonder if there isn't a way for process reliabilism to respond to the worry raised by *Red Lighting*? Perhaps Suzy should be characterized as retaining her belief in *Part 2* in a way that involves disregarding strong testimonial evidence, but then it is no longer clear that Suzy's belief is acquired and retained in a reliable way.

This, of course, raises the question of how to identify the process or mechanism relevant to the evaluation of whether a subject's belief is reliable, and as the literature on the generality problem for reliabilism reveals (e.g., Comesana 2006, Conee & Feldman 1998, Goldman 1979), responding to this question has turned out to be very difficult indeed. If fact, it may be that identifying the right way of characterizing a subject's belief formation will have to rely on our judgments about whether or not the subject knows[69]. Consequently, the resulting process reliabilist account of knowledge will not be a reductive analysis, but perhaps a reductive analysis of knowledge is more than we should hope for (see Williamson 2000, 2009a:9).

My aim here is not to determine whether a version of process reliabilism is able to satisfactorily accommodate our judgments in defeat cases. Instead, I want to focus on how a simple safety account of knowledge interacts with the idea that knowledge is defeasible.

---

[69] This thought is not in the spirit of the reductionist ambitions underlying the original statements of reliabilism (e.g. Goldman 1979, 1986). Likewise, my proposed response on behalf of process reliabilism may very well not be in the spirit of the process reliabilist view as originally stated. My point is merely that the details of an argument for a tension between an externalist theory and the defeasibility of knowledge needs to be spelled out in detail.

In recent years, the safety condition has been one of the more prominent proposals for an externalist condition knowledge[70], and some authors have even expressed attraction to a view on which knowledge just is belief safe from error[71] (Lasonen-Aarnio 2010, Williamson 2009a,b). Williamson (2009a), for example, relies on "…a conception of knowledge as safety from error" (2009a:9) in his defense of multi-premise closure for knowledge, although contrary to traditional externalist theories of knowledge, Williamson's (2009a) safety account is not reductionist:

> "I do not intend it [the safety account] to provide necessary and sufficient conditions for knowing in more basic terms. Without reference to knowing, it would be too unclear what sort of safety was in question." (2009a:9)

Elsewhere, he describes the anti-reductionist nature of his proposal by saying that his project is not to give a semantic analysis of 'knowledge' in terms of 'safety', but rather to explore some of the structural features of knowledge given how "Safety is analogous to knowledge" (2009a:0):

> "…the aim is to use safety, in the ordinary sense of 'safety', as a model to help explain the underlying nature of knowledge itself, in the ordinary sense of 'knowledge'." (2009a:20)

In the following I will focus on a simple safety account of knowledge, where safety is understood roughly along the lines of Williamson (2000, 2009a,b).

---

[70] See, e.g., Pritchard (2005), Sosa (1999, 2000, 2007), Williamson (2000, 2009a,b)

[71] While Williamson in his (2009a,b) expresses attraction to the view that knowledge just is belief safe from error, it is unclear whether *Knowledge and Its Limits* only takes safety to be a necessary condition on knowledge. The formulations in *Knowledge and Its Limits* (2000:123-8) suggest that safety is just a necessary condition, but in his (2009a,b) Williamson describes *Knowledge and Its Limits* as advancing "…the conception of knowledge as safely true belief." (2009b:324).

Before considering whether the safety account actually does face a challenge when it comes to our judgments in defeat cases, the notion of safe belief will have to be spelled out in some detail.

3. Introducing safety. The notion of safety is sometimes given the intuitive gloss that one's belief is safe from error just in case *one could not easily have held a relevantly similar false belief* (see Williamson 2000:125). A bit more technically, the idea is that in a case α a subject S is safe from error in believing that p if and only if there is no[72] case close to α in which S has a relevantly similar false belief (Williamson 2000:127). A case is here a

> "…possible total state of a system, the system consisting of an agent at a time paired with an external environment, which may of course contain other subjects. A case is like a possible world, but with a distinguished subject and time: a 'centred world' in the terminology of David Lewis (1979)." (Williamson 2000:52).

In other words, a case is triple consisting of a world, a time, and an agent. There is reason for scepticism when it comes to the prospects of reductive analysis of what it means for a case α to be *close* to a case β. Even so, we need to say a couple of things in order for there to be anything illuminating about a safety conception of knowledge. First, closeness is reflexive: any case is close to itself. If in α one falsely believes p, then there is a case close to α in which one falsely believes p (i.e. α), and one's belief is therefore not safe. The reflexiveness of closeness mirrors the factivity of knowledge (one cannot know p if p is

[72] On some accounts, safe belief is compatible with close cases in which one has a relevantly similar false belief as long as one does not a have a relevantly similar false belief in *most* close cases. See Pritchard (2005, chap. 6) for a view roughly along those lines. Although I will not consider such views here, I see no reason why the arguments below cannot be modified to concern those conceptions of safety as well.

false). Second, closeness is symmetric: if α is close to β, then β is close to α. Third, closeness is not transitive: if α is close to β and β is close to χ, α need not be close to χ (to be safe is not to be safely safe).

While this does not yet tell us very much about what it means for a case α to be close to a case β, the notions of safety and closeness draw part of their intuitive appeal from their connections to the above mentioned notion of *what could easily have happened*. I believe that Obama is the president of the United States. If Obama had been assassinated 5 minutes ago, he would no longer be president, but I would still believe that he was. But since it *could not easily have happened* that Obama had been assassinated 5 minutes ago, the fact that I would have had a relevantly similar false belief if he had been assassinated does not make my belief unsafe. There are no close cases in which Obama has been assassinated 5 minutes ago.

Even if a proposition p that I currently believe *could easily* have been false, that does not make my belief unsafe if I *could not easily have believed p falsely*. Assume that by perception, I have come to believe that a young child is standing on the edge of a cliff[73]. Perhaps it *could easily* have been that the child was not standing on the edge, because the child could easily have fallen. That is, there is a close case in which the child is not standing on the edge. Even so, my belief may be safe, because in close cases where the child has fallen off the edge I do not believe that the child is on the edge.

Finally, even in cases where I believe p, but could easily have believed p falsely, my belief need not be unsafe. The fact that I could easily have believed p falsely does not entail that I could easily have had a *relevantly similar* false belief, because believing p

[73] This case is from Williamson (2000:125)

falsely need not be *relevantly similar* to my believing p truly. Consider a grandmother[74] who has come to believe via perception that her grandson in well. She has good eyesight and can tell whether or not her grandson is well by looking at him. Assume that he could easily have been ill. That is, there are close cases in which he is ill. In these cases, however, he does not come to visit and she is misleadingly informed by his family that he in fact is well. As such, she could easily have believed falsely that her grandson is well. This does not make her belief that he is well unsafe, because her actual belief that he is well is not *relevantly similar* to the false belief she could easily have formed. Beliefs formed by relying on perception and by relying on testimony are not *relevantly similar*, because *the methods of belief formation are importantly different*.

In the remainder of the paper I will focus on a simple safety account of knowledge according to which knowledge just is belief safe from error. The purpose is to investigate what resources the proponent of a simple safety account of knowledge has available when it comes to explaining our judgments in defeat cases.

4. An argument by Lasonen-Aarnio. Two things are required for the simple safety account of knowledge to accommodate our judgments in cases like *Red Lighting* and *Premier League Reports*. First, it has to be the case that Suzy's and Jack's beliefs in *Part 1* are safe, and second, it has to be the case that their beliefs in *Part 2* are unsafe. For present purposes, we will assume that Suzy's and Jack's beliefs are safe in *Part 1*. To determine whether the safety theorist can accommodate the judgment that knowledge is lost in *Part 2*, we should therefore consider whether it is plausible that Suzy's and Jack's beliefs go from being safe in *Part 1* to unsafe in *Part 2*.

[74] This case is a modification of a case from Nozick (1981:179). See also Williamson (2000:153)

The strategy for my discussion will be as follows. In this section, I provide a statement of an argument by Lasonen-Aarnio (2010) to the effect that a simple safety account is unable to accommodate the idea that Suzy and Jack stops knowing in *Part 2*. In the next section (section (**5**)), I highlight some problems for Lasonen-Aarnio's (2010) argument and spell out a number of assumptions required to establish that a simple safety account of knowledge is unable to accommodate defeat. In section (**6**) and (**7**), I provide a detailed discussion and defense of the assumptions. Although the argument turns out to be far from straighforward, I therefore ultimately conclude that a simple safety account of knowledge is unable to satisfactorily accommodate our judgments in defeat cases like *Red Lighting* and *Premier League Reports*.

Here is first a statement of Lasonen-Aarnio's (2010) line of thought. When subjects acquire sufficiently strong misleading evidence that p is false, they typically stop believing that p. Sometimes, however, they don't. For present purposes, we will focus exclusively on subjects who retain their beliefs in the face of misleading evidence. Consider again *Red Lighting* from above. In *Part 1*, Suzy's belief about the colour of the object is assumed to be known and hence safe. Does Suzy's belief fail to be safe if she retains her belief in the face of the misleading evidence acquired in *Part 2*? Lasonen-Aarnio (2010) writes:

> "…presumably, it remains true throughout that Suzy could not easily have initially formed a false belief about the colour of the object. Moreover, there is no reason to think that the mere retention of her belief introduces a failure of safety, assuming that Suzy could not easily have, at t_2, ended up believing the relevant object to be any colour other than red" (2010:5).

Lasonen-Aarnio (2010) therefore finds it difficult to see why Suzy's belief that the object is red must fail to be safe at t_2 if the very same belief was safe at t_1. Given that we do in fact sometimes retain our beliefs in the face of misleading evidence, it therefore looks like the simple safety account is unable to accommodate our judgments about the loss of knowledge in defeat cases like *Red Lighting*.

While I am sympathetic to the conclusion that the simple safety account of knowledge cannot deal satisfactorily with our defeat judgments in cases like *Red Lighting* and *Premier League Reports*, the argument for that conclusion relies on a number of assumptions that I think it is worth bringing out and defending in more detail. That is the project of the following sections.

5. Three assumptions. If we take *Premier League Reports* as our example, the two basic claims of Lasonen-Aarnio's argument are these:

CLAIM A: It remains true throughout that Jack in *Part 1* could not easily have formed a false belief about the result of Man U's match.

CLAIM B: Insofar as Jack could not easily have ended up believing something else in *Part 2* than in *Part 1*, Jack's retention of his belief in the face of misleading evidence does not induce a failure of safety.

By considering these two claims, we can spell out three assumptions that are plausibly required for the success of an argument to the conclusion that a simple safety account of knowledge cannot accommodate our judgments in *Red Lighting* and *Premier League Reports*.

Let us start by considering CLAIM A. As it stands, it is not clear that the safety theorist should take CLAIM A to be true. In the terminology of close cases, CLAIM A states that it remains true throughout that in *Part 1* there are no close cases in which Jack forms a false belief about the result of Man U's match. But why should the safety theorist accept that? Perhaps there is a close case in *Part 1* where Jack does not read *The Guardian*, but is told by his wife that Man U lost. This is not a problem for the safety of Jack's belief in *Part 1* as long as the false belief Jack acquires by talking to his wife is not *relevantly similar* to the true belief he acquires by reading *The Guardian*. What is required for safety is not that there are no close cases in which one has a false belief, but that there are no close cases in which one has a *relevantly similar* false belief.

But what does it mean for two beliefs to be *relevantly similar*? As I noted in section (**3**), the *relevant similarity* of two beliefs is typically cashed out by reference to the methods by which the beliefs are acquired. This is normally done by assuming that two beliefs are relevantly similar *only if* they involve the application of the same method[75]. On this construal, a belief acquired by reading *The Guardian* would therefore not be *relevantly*

---

[75] Lasonen-Aarnio (2010) thinks that this way of spelling out the relation between relevant similarity and methods is problematic, because it "…incorporates two assumptions that cannot be reconciled: the assumption that methods split token belief forming episodes into equivalence classes, and the assumption that any two relevantly similar belief episodes are applications of the same method. An immediate argument against combining these assumptions arises simply from the observation that when we talk about relevant similarity of beliefs, what we are interested in is a similarity relation. But similarity relations are not equivalence relations, since they are not transitive: *a* may be relevantly similar to *b*, and *b* may be relevantly similar to *c*, without *a* being relevantly similar to *c*" (2010:4). It is somewhat unclear why Lasonen-Aarnio (2010) thinks that the difference between similarity relations and equivalence relations is an objection to the idea that sameness of method is a *necessary* condition for relevant similarity. Assume that $belief_1$, $belief_2$, and $belief_3$ are all the result of the application of the *same* method. They therefore satisfy a *necessary* condition for being relevantly similar. This is no objection to the possibility that $belief_1$ is relevantly similar to $belief_2$, and $belief_2$ is relevantly similar to $belief_3$, but $belief_1$ is not relevantly similar to $belief_3$. For present purposes I will therefore set aside Lasonen-Aarnio's (2010) rejection of the standard formulation of safety and work with the standard view according to which sameness of methods is a necessary condition on relevant similarity. I should note that Lasonen-Aarnio's (2010) alternative to the standard formulation of safety makes it very unclear what the safety theorist is actually committed to, and as a consequence her own argument against the unsafety of Suzy's belief in *Part 2* is very far from being decisive.

similar to a belief acquired by talking to one's wife, because the belief-forming methods are different.

As it stands, the safety theorist should therefore not accept CLAIM A. Insofar as it remains true throughout that Jack knew in *Part 1*, the safety theorist should, however, accept the following modification of CLAIM A:

CLAIM A*: It remains true throughout that Jack in *Part 1* could not easily have formed a false belief that is *relevantly similar* to his belief that Man U won.

We are now in a position to spell out three assumptions that are required for CLAIM B to be true given the safety theorist's commitment to CLAIM A*. In other words, we can now spell out three assumptions that are required for it to be true that Jack's *retention* of his belief that Man U won in the face of misleading evidence does not make his belief unsafe in *Part 2*.

The first assumption can be spelled out by consider a way the safety theorist might try to deny CLAIM B.

Imagine that the safety theorist insists that when Jack comes to believe that Man U won in *Part 1*, we should characterize him as using the method of *believing that p if the match report in The Guardian says that p and <u>one has no reason</u> to doubt the match report in The Guardian*.

Imagine further that when Jack in *Part 2* retains his belief that Man U won after having read a match report to the contrary in *The Times*, the safety theorist insists that we should no longer characterize Jack as using the same method. Now, the safety theorist

insists, we should characterize Jack as using the method of *believing that p if the match report in The Guardian says that p and one has good reason to doubt the match report in The Guardian*.

At least initially, these claims about the methods used by Jack in *Part 1* and *Part 2* do not seem outrageous. Imagine finally that the safety theorist insists that in *Part 1* there are in fact close cases where *The Guardian* has made a false match report, but these are all cases in which Jack has good reason to doubt the match report in *The Guardian*.

If this final claim about what cases are close in *Part 1* could be made plausible, then it looks like the safety theorist would have a compelling reason to deny CLAIM B.

Given the method used by Jack in *Part 1*, there are no close cases in which use of the same method gives rise to a false belief (i.e. there are no close cases in which *The Guardian* has made a false match report and Jack has no reason to doubt *The Guardian*). Sameness of methods is assumed to be a necessary condition on relevant similarity, and in *Part 1* there are therefore no close cases in which Jack has a relevantly similar false belief. As such, Jack's belief in *Part 1* is safe.

Given the method used by Jack in *Part 2*, there are in fact close cases in which use of the same method gives rise to a false belief (i.e. there are close cases in which *The Guardian* says that p, but in which Jack has good reason to doubt *The Guardian*). This by itself does not mean that there are close cases in which Jack has a *relevantly similar* false belief, since *sameness of method* is only a necessary condition on *relevant similarity*. Consequently, the closeness of cases in which Jack forms a false belief by use of the same method does not *establish* that his belief in *Part 2* is unsafe.

Nonetheless, a central motivation for accepting CLAIM B would be lost. If the safety theorist could establish that in *Part 2* there are in fact close cases where Jack forms a false belief by use of the same method as the method by which he holds the belief that Man U won, then at least some of these false beliefs will plausibly qualify as *relevantly similar* to Jack's actual belief in *Part 2*. And if some of these beliefs do qualify as *relevantly similar*, the Jack's belief in *Part 2* is not safe.

This means that *if* it could be established that in *Part 1* there are close cases in which *The Guardian* has made a false match report, but these are all cases in which Jack has good reason to doubt *The Guardian*, *then* the safety theorist would have found a compelling reason to deny CLAIM B. In other words, the safety theorist would have found a compelling reason for thinking that Jack's retention of his belief that Man U won in the face of misleading evidence in *Part 2* does in fact induce a failure of safety.

This points to the first assumption we should make to defend CLAIM B. In the case of *Premier League Reports*, we can avoid the above objection to CLAIM B by assuming that in *Part 1* there are no close cases in which *The Guardian* has made a false[76] match report.

If in *Part 1* there are no close cases in which *The Guardian* has made a false match report, then there are also no close cases in which *The Guardian* has made a false match report while Jack has good reason to doubt *The Guardian*. Consequently, there are no close cases in which Jack acquires a false belief by using the method of *believing that p if the match report in The Guardian says that p and* <u>*one has good reason*</u> *to doubt the match report in The Guardian*.

[76] Of course, if *The Guardian* had reported that Man U had lost 15-0 to Chelsea, then Jack would not have trusted *The Guardian*. Whenever I talk of false match reports in *The Guardian* in the following, it should be seen as referring to false *but sensible* match reports.

Even if Jack should be characterized as using this method in *Part 2*, his belief is not unsafe if there are no close cases in which the same method gives rise to a false belief. The safety theoretic attempt to deny CLAIM B therefore cannot get off the ground if in *Part 1* there are no close cases in which *The Guardian* has made a false match report.

A natural way of defending CLAIM B therefore involves taking on board the following assumption:

ASSUMPTION 1: If Jack's belief that Man U won is safe in *Part 1*, then in *Part 1* there are no close cases in which *The Guardian* has made a false match report.

While Lasonen-Aarnio (2010) neither mentions nor defends this kind of assumption about which cases are close in *Part 1*, there are in fact good reasons to accept it. To see why, consider the following line of thought. Presumably, cases in which *The Guardian* has made a false match report and Jack has *no* reason to doubt *The Guardian* are intuitively closer than cases in which *The Guardian* has made a false match report but Jack *does* in fact have good reason to doubt *The Guardian*. The closest cases in which *The Guardian* has made a false match report are plausibly cases in which everything[77] appears the same to Jack.

If that is true, then the claim that there are close cases in which *The Guardian* has made a false match report, but Jack has good reason to doubt *The Guardian* naturally leads to the claim that there are close cases in which *The Guardian* has made a false match report and Jack has no reason to doubt *The Guardian*.

[77] Or at least *almost* everything appears the same. There might be a small change in the match report in *The Guardian*, since the close cases in which *The Guardian* has made false match report might be cases in which *The Guardia*n has reported a Man U loss or draw.

But if there are close cases in which *The Guardian* has made a false match report and Jack has no reason to doubt *The Guardian*, then it looks like Jack's belief was never safe to start with. The method of *believing that p if the match report in The Guardian says that p and one has no reason to doubt the match report in The Guardian* would lead to false beliefs in close cases. And at least some of these false beliefs would plausibly count as *relevantly similar* to Jack's belief in *Part 1*.

Consequently, it is plausible to assume that if Jack's belief that Man U won is safe in *Part 1*, this is because there are no close cases in which *The Guardian* has made a false match report. ASSUMPTION 1 is therefore reasonable.

It might be objected that in fact ASSUMPTION 1 is completely unnecessary for the plausibility of CLAIM B. The idea behind this objection could be the following. The suggested safety theoretic objection to CLAIM B relies on the idea that when Jack in *Part 2 retains* his belief in the face of misleading evidence, Jack's belief should be seen as *relevantly similar* to false beliefs *formed* by relying on the match report in *The Guardian* in the face of evidence that *The Guardian* is mistaken.

It might, however, be objected that the method involved in *retaining* a belief in the face of contrary evidence is not strictly speaking the *same* as the method involved in *forming* a belief in the face of contrary evidence. Since the sameness of method is a necessary condition on relevant similarity, the belief Jack *retains* in the face of evidence that *The Guardian* is mistaken is therefore not *relevantly similar* to a false belief *formed* by relying on *The Guardian* in the face of evidence that *The Guardian* is mistaken.

Even if there are close cases in which Jack *forms* a false belief by relying on the match report in *The Guardian* in the face of evidence that *The Guardian* is mistaken, this

does not mean that in *Part 2* there are close cases in which Jack comes to hold false beliefs that are *relevantly similar* to his actual belief that Man U won.

ASSUMPTION 1 is therefore superfluous.

Personally, I don't find this line of response particularly convincing. Insisting that we should sharply distinguish the methods involved in *retaining* and *forming* a belief in a case like *Premier League Reports* is simply not giving the simple safety account of knowledge a run for its money.

Furthermore, the suggested line of response relies heavily on the thought that the simple safety account should be spelled out in terms of *sameness* of methods. It is not clear, however, that this is ultimately the most plausible way of cashing out the notion of safe belief.

One problem with understanding safety in terms of *sameness* of methods is that it seems to presuppose a fairly coarse-grained conception of the methods we rely on in our belief formation. As was brought out by the grandmother example in section (**3**), the notion of methods is meant to highlight that there are different ways of coming to believe the same proposition. There are, however, difficult questions about how finely methods should be individuated.

At one end of the spectrum, one might think that a reasonable coarse-grained characterization of how Jack has come to believe that Man U won is by use of the method of *believing that p if The Guardian says that p* or perhaps the method of *believing that p if The Guardian says that p and one has no reason to doubt The Guardian.*

Some might think, however, that talk of *methods* in connection with a safety condition on knowledge should be understood in a much more fine-grained way that

incorporates various details about the kind of situation that a subject is in as she comes to believe that something is the case. On this latter way of thinking about methods, a subject's method of belief formation cannot be captured by simple slogans such a *believing that p if The Guardian says that p*.

On a very fine-grained conception of methods, there may be few (if any) close cases in which a subject forms a belief by use of *exactly* the same method as in the actual case. If two beliefs count as *relevantly similar* only if they involve the application of the *same* method, then there may only be few (if any) close cases in which a subject forms a belief that is relevantly similar if we assume a very fine-grained conception of methods. On a very fine-grained conception of methods, requiring two beliefs to be formed by use of the *same* method for them to be *relevantly similar* therefore makes the safety conception of knowledge somewhat implausible, since any true beliefs almost automatically will be safe[78] (the false beliefs formed in close cases are not relevantly similar, because the method used is slightly different).

In order to remain as neutral as possible on how methods should be individuated, it may therefore be better to say that two beliefs are *relevantly similar* only if they are formed and retained by use of *sufficiently similar* methods (rather than by use of the *same* method).

If we understand beliefs as *relevantly similar* only if they involve the application of *sufficiently similar* methods, then it looks even more problematic to claim that a belief

[78] This is not merely a problem for proponents of a view on which knowledge just is belief safe from error, but also for defenders of the idea that safety is a necessary condition on knowledge. If two beliefs are relevantly similar only if they involve the application of the *same* method and if methods should be individuated very finely, then the safety condition is in danger of collapsing into the claim that knowledge is factive.

retained in the face of strong evidence that the belief is false is not relevantly similar to a belief *formed* in the face of strong evidence that the belief is false.

Consequently, distinguishing between the methods involved in *forming* and *retaining* a belief in a case like *Premier League Reports* does not provide a very plausible strategy for responding to the safety theoretic objection to CLAIM B suggested above.

The most natural way of defending CLAIM B from the suggested safety theoretic objection is therefore by taking on board ASSUMPTION 1. And as argued above, ASSUMPTION 1 is ultimately rather plausible.

We can now turn our attention to two further assumptions that the plausibility of CLAIM B seems to require. That is, there are two additional assumptions that seem to be required for it to be plausible that Jack's retention of his belief in the face of misleading evidence in *Part 2* does not induce a failure of safety.

The first of the two assumptions is an extension of ASSUMPTION 1 (i.e. the assumption that in *Part 1* there are no close cases in which *The Guardian* has made a false match report). Specifically, the assumption required is that if in *Part 1* there are no close cases in which *The Guardian* has made a false match report, then in *Part 2* there are no close cases in which *The Guardian* has made a false match report. That is, CLAIM B seems to be plausibly only if receiving evidence that *The Guardian* has made a false match report does not make cases close in which *The Guardian* has in fact made a false match report.

The rationale for this assumption should be rather obvious. In *Part 1* Jack's belief is safe because there are no close cases in which *The Guardian* has made a false match report and Jack is using a method that involves relying on the match report in *The*

Guardian. In *Part 2*, Jack is plausibly still using a method that involves relying on the match report in *The Guardian*, but if in *Part 2* there are close cases in which *The Guardian* has made a false match report, then we can easily imagine a safety theoretic argument to the effect that Jack's belief is unsafe in *Part 2*.

In particular, the safety theorist might insist that Jack's belief is safe in *Part 1*, but fails to be safe in *Part 2*, because there is a difference in what cases count as close. To avoid this kind of objection to CLAIM B, we should therefore make the following assumption about what cases count as close:

ASSUMPTION 2: If in *Part 1* there are no close cases in which *The Guardian* has made a false match report, then in *Part 2* there are no close cases in which *The Guardian* has made a false match report.

While I think this assumption is not at all unreasonable, it may nonetheless be worth thinking more carefully about its plausibility. That is the project of section (**6**).

The final assumption that seems required for CLAIM B to be plausibly concerns which beliefs count as *relevantly similar* to Jack's belief in *Part 1* and in *Part 2*.

As we have seen above, it is quite plausible for the safety theorist to insist that the method Jack should be characterized as using in *Part 2* is importantly different from the method he should be characterized as using in *Part 1*. Whether or not we opt for a coarse-grained or a fine-grained individuation of methods, it is plausible to assume that in *Part 2* the method Jack is relying on should in one way or another be characterized so that mentions how he disregards evidence.

As a consequence of the important difference in the methods Jack relies on in *Part 1* and *Part 2*, there is quite plausibly a difference in the methods that count as *sufficiently similar* to the method Jack relies on in *Part 1* and the method he relies on in *Part 2*. By extension, there is an important difference in what beliefs count as relevantly similar to Jack's belief in *Part 1* and Jack's belief in *Part 2*.

The safety theorist could try to argue that this difference can be invoked to explain why Jack's belief is safe in *Part 1* but fails to be safe in *Part 2*. Even if we grant that *The Guardian* could not easily have made a false match report in *Part 1* and *Part 2*, there is still a whole range of other false beliefs that Jack could easily come to hold in both *Part 1* and *Part 2* if he relies on methods that involve disregarding evidence. Only in *Part 2*, however, is this range of false beliefs relevantly similar to Jack's belief about the outcome of Man U's match, because he only disregards evidence in *Part 2*. We should therefore see Jack's belief in *Part 2* as unsafe even though his belief in *Part 1* is safe.

To avoid this kind objection to CLAIM B, we therefore need to assume that the difference in what beliefs count as relevantly similar to Jack's belief in *Part 1* and *Part 2* cannot be invoked to defend the idea that Jack's retention of his belief in *Part 2* induces a failure of safety:

ASSUMPTION 3: If *The Guardian* could not easily have made a false match report in *Part 1* and *Part 2*, then the difference in what beliefs count as *relevantly similar* to Jack's belief in *Part 1* and *Part 2* does not make Jack's belief safe in *Part 1* and unsafe in *Part 2*.

Again, I think the assumption is ultimately plausible, but nonetheless worth considering in more detail. That is the project of section (**7**).

6. A *flexible* understanding of closeness. As I argued above, one way the safety theorist might try to deny CLAIM B is by insisting that CLAIM B relies on a misunderstanding of what it means for a case α to be *close* to a case β. CLAIM B plausibly relies on the idea that if there are no close cases in which *The Guardian* has made a false match report before Jack's encounter with *The Times*, then there are no close cases in which *The Guardian* has made a false match report after Jack's encounter with *The Times*.

The proponent of a safety account of knowledge could, however, deny that this is really what falls out of a proper understanding of the notion of *close cases*. She could emphasize that since cases are triples consisting of worlds, agents and times, the mere time difference should make everyone acknowledge that there is *some* difference in what cases count as close before and after Jack has read *The Times*. Acknowledging this difference, the safety theorist could insist that the cases that count as close before and after Jack has read *The Times* differ in a way that explains why Jack's belief isn't safe after he has read *The Times* even if it was safe before.

To repeat how this might work, assume that the methods that count as *sufficiently similar* to the methods used by Jack before and after he has read *The Times* in some way involve *relying on the match report in The Guardian*. The idea is now that while in *Part 1* there are no close cases in which *The Guardian* has made a false match report, *Part 2* differs from *Part 1* in a way that makes such cases close. While it is therefore true that in *Part 1* there are no close cases in which Jack forms a false belief by use of a sufficiently similar method, it is also true that in *Part 2* there are close cases in which Jack forms a false belief by use of a sufficiently similar method. Assuming that some of these false beliefs count as *relevantly similar* to Jack's belief in *Part 2*, his belief in *Part 2* will therefore be unsafe (even though his belief in *Part 1* is safe).

Since *Part 1* and *Part 2* are themselves very different cases, there is nothing surprising about the fact that they differ in what cases count as close. Although it may initially seem like a slight stretch, we should therefore accept that while at the earlier time, t_{Part1}, Jack could not easily have formed a relevantly similar false belief, he could at the later time, t_{Part2}, easily have held a relevantly similar false belief.

Let us call the understanding of closeness at work in this kind of explanation a *flexible* understanding of closeness[79].

While the flexible understanding of closeness may seem to provide the safety theorist with the resources required to explain our judgments in defeat cases, the account is not without its problems. The central worry for a flexible understanding of closeness is that it is at odds with our intuitive understanding of safety in terms of what *could easily have happened*. Since at least part of the attraction of the safety account derives from our pre-theoretic grip on the notion of what *could easily have happened*, this is an undesirable consequence of the flexible understanding of closeness.

A couple of examples can help to illustrate the worry.

The difference in time between *Part 1* and *Part 2* played some role in initially motivating flexible closeness. Consider now flexible closeness in an interpersonal case where the time is kept fixed. Assume again that Jack has read both *The Guardian* and *The Times*. Given

[79] When considering the merits of this idea below, it is important to note that the cases in which *The Guardian* has made a false match report need not be cases in which *The Guardian* says that Man U won although in fact they lost or drew. It can also be cases in which *The Guardian* has reported a Man U loss or a Man U draw while in fact Man U won. For a belief to count as relevantly similar it need not have the same content as Jack's actual belief. In fact, if that was a requirement, it would be difficult to explain how believing a complicated truth of first order logic on the basis of a lucky guess could ever fail to be safe.

flexible closeness, there is a close case in which *The Guardian* has made a false match report. In other words, *The Guardian could easily* have made a false match report. That, in effect, is what prohibits Jack from knowing that Man U won by relying on *The Guardian*. Consider now Jack's colleague, Jill, who (at time t_{Part2} where Jack has read both newspapers) has read neither *The Guardian* nor the *The Times*. Assuming that Jack knew that Man U won prior to reading *The Times*, it seems that if Jill were to read *The Guardian*, she could thereby come to know that Man U won. Given that Jill would also have trusted a false match report, it therefore seems that (from Jill's perspective) *The Guardian could not easily have made a false match report*.

In and of itself, this seems like a somewhat peculiar result. Given that Jack and Jill are different people, it is surely true that what *could easily* have happened to Jack may not easily have happened to Jill (just as what could easily have happened at time t_{Part1} varies from what could easily have happened at time t_{Part2}). Even so, the idea that *The Guardian*'s making a false match report is one of the things that could easily have happened to Jack, but not to Jill, is at odds with our ordinary understanding of how facts about what *could easily have happened* vary from one person to another.

Additionally, one might worry that these observations about the interpersonal case have certain undesirable consequences for the safety condition when it comes to knowledge via testimony. As the case above is described, Jill can come to know that Man U won by reading *The Guardian*. Assume that she does not, but decides to ask Jack. Assume further that she knows that if Jack has anything to say about the match, he will have read it in *The Guardian*. Jack says 'Man U won'. Jill forms the belief that Man U won. Does she know? Our initial inclination is to say that she does not, but it is not clear that flexible closeness can explain this verdict. In Jill's situation, *The Guardian* could not

easily have made a mistake, and so there is no case close to Jill's in which Jack's testimony is false. If this is right, then there is presumably no case close to Jill's in which she forms a relevantly similar false belief on the basis of Jack's say-so. In other words, it seems that her belief is safe.

If this is in fact a consequence of flexible closeness, that is surely an undesirable result.

Consider now a case of how memory loss can induce evidence loss. Assume again that Jack has read both *The Guardian* and *The Times*. According to flexible closeness, Jack does not know because there is a close case in which *The Guardian* has made a false match report. Or less technically, *The Guardian could easily* have made a false match report.

Imagine that Jack is involved in an accident and that as a consequence he suffers from severe retrograde amnesia. He forgets who he is, where he is from and what has happened prior to the accident (including his reading of *The Guardian* and *The Times*). Years later, Jack finds an old copy of *The Guardian* that he (unbeknowst to himself) has read many years ago. He reads in *The Guardian* that Man U won and comes to believe that they did.

There is some inclination to say that if Jack knew that Man U won after he read *The Guardian* years ago, he also comes to know that they won by reading *The Guardian* again. On flexible closeness, this will have to be explained by the idea that there is no close case in which *The Guardian* has made a false match report. In other words, *The Guardian could not easily* have made a false match report.

Consequently, Jack's amnesia has made a difference to whether or not cases in which *The Guardian* makes a false match report are close. Whereas this particular copy of

The Guardian could easily have contained a false match report *prior* to Jack's memory loss, that same copy could *not* easily have contained a false match report *after* Jack's memory loss.

At least on an ordinary understanding of what *could easily* have happened, amnesia (or memory loss more generally) does not seem like the kind of thing that should make a difference to how easily the seemingly independent event of a false match report in *The Guardian* could have occurred.

The proponent of flexible closeness may insist that even though her account disconnects the safety condition from our ordinary understanding of what *could easily have happened*, the account still has some appeal.

The idea could be that if we consider Jack's situations before and after he has read *The Times*, there is a difference in what cases we *should* consider close in the more theoretical sense of being *epistemically relevant* to the evaluation of Jack's situation. That is, there is an important sense in which cases where *The Guardian* has made a false match report are not relevant to the epistemic evaluation of Jack's situation *prior* to his reading of *The Times*, whereas some of these cases may be relevant to evaluating Jack's epistemic situation *after* he has read *The Times*.

Even if we set aside the somewhat theoretical and technical nature of this way of understanding the notion *close cases*, there are other worries with the account. One such worry arises if we consider a version of a well-known challenge for the idea that knowledge is closed under competent deduction (see Hawthorne 2004:70-73). To

formulate the challenge, I will rely on the following formulation of single-premise closure for knowledge (SPC):

SPC: Necessarily, if S knows *p*, competently deduces *q*, and thereby comes to believe *q*, while retaining knowledge of *p* throughout, then S knows *q*. (Hawthorne 2004:34)

Consider again Jack prior to his encounter with *The Times*. As before, Jack has come to know that Man U won by reading *The Guardian*. Assume that Jack from his knowledge that Man U won competently deduces that *if The Times has reported that Man U lost, The Times is mistaken*. If Jack believes what he has deduced, then according to SPC Jack knows that *if The Times has reported that Man U lost, The Times is mistaken*.

The result that Jack can come to know this conditional is sometimes considered a problem for SPC, since people are not inclined to say that Jack can know the deduced conditional in a situation where he has come to know the premise of his deductive inference by reading *The Guardian*. In particularly, people worry that Jack's knowledge of the deduced conditional would allow him to infer (and thereby come to know) that *The Times* has made a mistake when he reads in *The Times* that Man U lost.

A popular closure preserving response to the worry relies on the phenomenon of knowledge defeat (see Hawthorne 2004:72-3). The idea is that while Jack does indeed know the deduced conditional prior to reading *The Times*, this knowledge cannot be used to infer that *The Times* has made a mistake, because Jack's knowledge of the conditional is defeated when he reads the diverging match report in *The Times*.

While there may be independent reasons for finding this an unsatisfactory response to the problem, I think the response looks particularly unattractive if we assume flexible closeness. Recall that according to flexible closeness, Jack knows that Man U won prior to his encounter with *The Times* only because there is no close case in which *The Guardian* has made a false match report. Or on the current way of understanding flexible closeness, the cases where *The Guardian* has made a false match report are *not relevant to the epistemic evaluation of Jack's situation*.

On flexible closeness, competent deduction therefore allows Jack to know that *if The Times has reported that Man U lost, The Times is mistaken* because cases where *The Guardian* has made a false match report are *not relevant to the epistemic evaluation of Jack's situation*. In other words, when Jack is *actively* thinking about whether *The Times* has made a misprint and when he concludes on the basis of information from *The Guardian* that if *The Times says otherwise, The Times is mistaken*, the cases in which *The Guardian* has made a false match report are *not* relevant to an epistemic evaluation of Jack's situation. As soon as Jack reads in *The Times* that Man U lost, however, the cases in which *The Guardian* has made a false match report *do* become relevant to the evaluation of Jack's situation.

This commitment to when cases become *epistemically relevant* looks gerrymandered and unmotivated, and much of the initial attraction of the safety account therefore seems compromised by flexible closeness[80].

[80] As Williamson (2009b) himself notes: "…we do not want closeness to be *too* unnatural." (2009b:329).

7. Determining what counts as relevantly similar. As we have seen, there are undesirable consequences of explaining knowledge defeat by reference to variations in what cases count as close. The proponent of a simple safety account of knowledge might therefore attempt to explain our judgments about *Premier League Reports* by reference to differences in which beliefs count as *relevantly similar* before and after Jack has read *The Times*.

Here is one initial reason why one might be skeptical of this kind of response. Both before and after Jack has read *The Times*, it seems that Jack has come to believe that Man U won by *relying on the match report in The Guardian*. Both before and after Jack has read *The Times*, it also seems that *The Guardian* could not easily have made a false match report. That is, unless we assume the flexible account of closeness discussed in the last section, it seems that both before and after Jack has read *The Times* there are no close cases in which *The Guardian* has made a false match report. If all the beliefs that count as relevantly similar to Jack's actual belief are beliefs formed by *relying on the match report in The Guardian*, it therefore seems that even as Jack retains his belief in light of *The Times*' match report to the contrary, Jack could not easily have had a relevantly similar false belief. In the terminology of cases, there are no close cases in which Jack has a relevantly similar false belief.

The proponent of explaining knowledge defeat by reference to safety failure could object that the above argument relies on a notion of *relevantly similar beliefs* that is too strict. In *Part 2* Jack retains his belief that Man U won despite reading a match report to the contrary in *The Times*. In such a case, the relevantly similar beliefs are not merely beliefs that involve relying on *The Guardian*'s match report while ignoring *The Times*' match report. For someone who, like Jack, ignores *The Times*' match report to the

contrary, the relevantly similar beliefs also involve all kinds of other non-football related beliefs that Jack could have held by relying on *The Guardian* while ignoring *The Times*. As soon as we understand relevantly similar beliefs in this slightly more relaxed way, it seems much more plausible to think that Jack *could easily have had a relevantly similar false belief*. It seems reasonable to suppose that the more general strategy of relying on *The Guardian* despite contrary evidence in *The Times* is a strategy that could easily lead to a false belief. It may therefore seem that slightly relaxing our understanding of which beliefs count as relevantly similar provides the safety theorist with the resources to explain why Jack fails to know after he has read *The Times*.

There are two related worries for this strategy. Both of them take as their starting point the observation that if we allow for a more relaxed understanding of what beliefs count as relevantly similar *after* Jack has read *The Times*, then it is natural to allow for a more relaxed understanding of what beliefs count as relevantly similar *before* Jack has read *The Times*. It was proposed above that all kinds of other non-football related beliefs that Jack could have held by relying on *The Guardian* while ignoring contrary evidence in *The Times* are beliefs that count as relevantly similar to Jack's actual belief in *Part 2*. If this is plausible, then other non-football related beliefs that Jack could have formed by relying on *The Guardian* prior to reading *The Times* presumably count as relevantly similar to Jack's actual belief in *Part 1*.

The first worry is that this understanding of what beliefs count as relevantly similar is too relaxed. Newspapers often make mistakes, and on the current understanding of relevantly similar beliefs, the beliefs that Jack could have easily have formed on the basis of such mistakes count as relevantly similar. For Jack to believe safely that Man U won by reading a correct match report in *The Guardian* in *Part 1*, the current account requires that

there could not easily have been a misprint somewhere else in *The Guardian*. If that is right, the one might worry that Jack's belief that Man U won is not very likely to be safe even before he has read *The Times*. If there could easily have been a factual mistake in one of the sections of *The Guardian*, then on the current account Jack could easily have had a relevantly similar false belief. As such, one might worry that relaxing the requirement on what beliefs count as relevantly similar will have sceptical consequences.

The second (and perhaps more serious) worry arises if the proponent of safety maintains that in order for Jack's belief to be safe it has to be the case that *The Guardian* could not easily have made any misprints. The worry is that in a situation where Jack's belief that Man U won actually is safe, it is difficult to see how even the relaxed understanding of relevantly similar beliefs can provide the conclusion that Jack's belief that Man U won is unsafe after he has read *The Times*. If none of the other beliefs that Jack could have formed by relying on *The Guardian* are false (because *The Guardian* could not easily have made any misprints), then none of the other beliefs that Jack could have formed by relying on *The Guardian* despite contrary evidence from *The Times* are false. The latter beliefs are a subset of the former.

My response above relies on the thought that if we are relaxed about what *relevant similarity* amounts to after Jack has read *The Times*, then we should also be relaxed about what *relevant similarity* amount to before Jack has read *The Times*. If the proponent of the simple safety account of knowledge wants to explain knowledge defeat in a case like *Premier League Reports* by reference to what beliefs count as *relevantly similar* before and after Jack has read *The Times*, it therefore seems that what is required is an argument for thinking that our standards for *relevant similarity* should be strict before Jack has read *The*

Times and relaxed after he has read *The Times*. Before such an argument has been provided, I think we should be sceptical about the prospects of explaining defeat by reference to what beliefs count as *relevantly similar*[81].

8. Conclusion. Although there are a number of subtle strategies that the proponent of a simple safety account of knowledge might invoke in an attempt to handle knowledge defeat, I think we should ultimately conclude that our judgments in defeat cases like *Red Lighting* and *Premier League Reports* cannot be explained by a simple safety account of knowledge. Those attracted to the idea that knowledge involves safe belief should therefore find other ways of dealing with our judgments in defeat cases. Deciding whether this involves an error theory about those judgments (see Lasonen-Aarnio 2010) or a different and potentially more complicated account of knowledge is beyond the scope of this paper.

[81] One might, of course, claim that what beliefs count as *relevantly similar* is to be determined by our judgments about whether the subject knows. If our judgments that Jack knows in *Part 1* and fails to know in *Part 2* only line up with safety if we are strict about what counts as relevantly similar in *Part 1* and relaxed about what counts as relevantly similar in *Part 2*, then we should be strict in *Part 1* and relaxed in *Part 2*. Personally, I think this kind of maneuver risks compromising whatever grip we might have on relevant similarity, and I suspect that many will not find it satisfying as an explanation for why Jack's belief goes from being safe in *Part 1* to unsafe in *Part 2*.

CHAPTER 4

Defeat by Higher-Order Evidence

1. Introduction. The notion of higher-order evidence has received much attention in the recent literature. A number of arguments have been presented in defense of various surprising claims about how and when higher-order evidence affects the epistemic status of our beliefs[82]. Below I focus on Christensen's (2010a) and Lasonen-Aarnio's (2014a) arguments for thinking that we violate rational ideals or rules of rational belief formation by revising our beliefs in the face of certain kinds of misleading higher-order evidence. Christensen's (2010a) and Lasonen-Aarnio's (2014a) main arguments for the occurance of these violations proceed by pointing out supposedly unattractive consequences of rejecting that such violations occur. Christensen (2010a) argues that rejecting his diagnosis involves giving up a very attractive picture of what is distinctive about higher-order evidence. Lasonen-Aarnio (2014a) argues that rejecting her diagnosis involves giving up a very attractive picture of how we manage our beliefs by following epistemic rules. In what follows, I argue that if we violate rational ideals or rules of rational belief formation in the kinds of cases that Christensen (2010a) and Lasonen-Aarnio (2014a) discuss, then Christensen's (2010a) proposal for what is distinctive about higher-order evidence and Lasonen-Aarnio's (2014a) proposal for how we manage our beliefs will have to be given up anyway. Their arguments do therefore not provide us with a reason for thinking that rational ideals or rules of rational belief formation are violated in the relevant kinds of cases.

---

[82] See, e.g., Christensen (2007, 2010a,b), Elga (2013, *unpublished*), Lasonen-Aarnio (2014a), Williamson (2011, *forthcoming*).

2. The notion of higher-order evidence. It is a familiar idea that a belief's status of being known, justified or rational is defeasible. That is, a belief can cease to be known, justified or rational if the subject holding the belief receives sufficiently strong misleading evidence. A few cases will help to illustrate the idea:

> *Red Lighting*[83] *– Part 1*
>
> At time t_1 Suzy is looking at a red object in front of her and based on her experience she forms the belief that the object in front of her is red. There is nothing abnormal about Suzy's eyesight or the lighting in the room.

> *Premier League Reports*[84] *– Part 1*
>
> *The Times* and *The Guardian* are both extremely reliable in reporting Premier League results and Jack tends to trust them equally. Man U just won an important match against Chelsea and on the front page of *The Guardian* Jack finds the Man U victory reported. He therefore forms the belief that Man U won.

When presented with these cases the inclination is to say that both Suzy's belief and Jack's belief are known, justified and rational. Consider now the following developments:

> *Red Lighting – Part 2*
>
> At time t_2 a person, Sally, whom Suzy trusts and considers highly reliable tells Suzy that the object she is looking at is illuminated by red lighting that would make objects of any colour look red.

> *Premier League Reports – Part 2*
>
> Shortly after reading *The Guardian*, Jack reads on the front page of *The Times* that Man U lost the match.

[83] This case is a version of Lasonen-Aarnio's (2010:1) *Trick on Suzy*.
[84] This case is a modified version of a case from Hawthorne (2004:71).

In light of Suzy's misleading conversation with Sally and Jack's misleading encounter with *The Times* we are inclined to say that even if Suzy still believes that the object is red and even if Jack still believes that Man U won the match, these beliefs are neither known, justified or rational.

The standard way for philosophers to think about cases like *Red Lighting* and *Premier League Reports* is to see them as cases exemplifying the defeasibility of knowledge, justification and rationality. That is, philosophers tend to think that as a consequence of receiving misleading evidence[85], Suzy's and Jack's beliefs have lost the epistemic status of being known[86], justified and rational.

In recent years, special attention has been paid to so-called *higher-order* evidence and the conditions under which receiving misleading higher-order evidence can make a belief lose its status of being known, justified or rational[87].

While the notion of higher-order evidence is rarely precisely defined, the general idea is that higher-order evidence with respect to a proposition p is evidence about the way in which one has come to believe p. As such, misleading higher-order evidence with respect to p might be evidence that the belief that p is the result of a flawed belief-forming

---

[85] Philosophers sometimes distinguish between *rebutting* and *undercutting* defeaters (see Pollock 1986:38-39). A rebutting defeater for a belief that p is typically specified a reason for believing the negation of p. An undercutting defeater for a belief that p is typically specified as reason for no longer believing p without hereby being a reason for believing the negation. Philosophers attracted to this distinction typically think of the evidence Suzy receives in *Part 2* as an undercutting defeater and the evidence Jack receives in *Part 2* as a rebutting defeater.

[86] Recently, some authors have suggested that acquiring misleading evidence does not defeat knowledge (and rationality) in as broad a range of cases as epistemologists have tended to think. Later in this paper I discuss one of Lasonen-Aarnio's (2014a) arguments for thinking that a belief can continue to be rational in the face of strong misleading *higher-order* evidence. See Lasonen-Aarnio (2010) and Hawthorne & Srinivasan (2013) for authors attracted to the idea that a belief can continue to be known even when retained in the face of strong misleading *first-order* evidence.

[87] See, e.g., Christensen (2007, 2010a,b), Elga (2013, *unpublished*), Lasonen-Aarnio (2014a), Williamson (2011, *forthcoming*).

process when in fact it is the result of an impeccable belief-forming process. Standard examples include evidence that one has made a simple calculation error, evidence that one is subject to an undetectable cognitive malfunction or evidence that one has failed to appreciate the import of one's evidence. I might, for instance, receive false testimony that I have been given a reasoning-distorting drug or I might learn that someone I consider an epistemic peer has reached a different conclusion than me by reasoning from the same evidence. In these cases the evidence I receive is evidence that a belief of mine is the result of a flawed process, which is higher-order evidence with respect to the proposition believed.

The notion of higher-order evidence gives rise to a number of interesting issues. One such issue concerns the conditions under which higher-order evidence with respect to p affects the rationality of believing p. It is a natural thought that if I receive strong evidence that a belief of mine is the result of a flawed process, then it is not rational for me to hold on to that belief even if the evidence is misleading and my belief in fact is the result of an impeccable belief-forming process. In other words, there is an inclination to think that in such circumstances strong misleading higher-order evidence with respect to p has defeating force and therefore makes it irrational for me to continue believing p.

Recently, however, some authors have raised a number of challenges for this idea (see, e.g., Hawthorne & Srinivasan (2013), Lasonen-Aarnio (2014a), Williamson (*forthcoming*)). That is, some have suggested that there are circumstances in which it is perfectly rational for me to continue believing p even though I have strong misleading evidence that my belief that p is the result of a flawed process.

Another interesting issue that the notion of higher-order evidence gives rise to is the question of when evidence for the falsity of p is also evidence that one's belief that p was acquired in an irrational manner. Many are inclined to think that evidence for the falsity of p need not be evidence that one's belief that p was irrational prior to one's acquisition of the relevant evidence. Many are, for example, inclined to think that the evidence received by Suzy and Jack in *Part 2* of *Red Lighting* and *Premier League Reports* is evidence that the beliefs they formed in *Part 1* are false, but not evidence that those beliefs were irrational. These claims about rationality are, however, not entirely uncontroversial. If, for example, one is attracted to a view on which a belief is rational or justified only if (or just in case) it is true[88] (or known[89]), then the evidence Jack receives in *Part 2* is evidence that his belief in *Part 1* was not rational or justified.

Although the above issues have some relevance for my discussion in the following, they are not my primary concern. Instead, I want to focus on arguments by Christensen (2010a) and Lasonen-Aarnio (2014a) to the effect that subjects violate rational ideals or rules of rational belief formation by revising their beliefs in the face of certain kinds of misleading higher-order evidence.

3. Violating rational ideals or epistemic rules. Before engaging with Christensen's (2010a) and Lasonen-Aarnio's (2014a) arguments, it will be helpful to get clear on what they are arguing for. To spell out Christensen's (2010a) view, let us start by considering the following case:

[88] See, e.g., Littlejohn (2012).
[89] See, e.g., Sutton (2005).

DRUG[90]. Harry works at a pharmaceutical company that has recently invented a drug that can affect people's ability to make competent deductions. The drug is designed so that it only affects people's ability to make competent deductions from a limited set of propositions (the *reactive* propositions). The drug works in the following way. If a subject affected by the drug considers what follows from a reactive proposition, she will draw grossly mistaken inferences although they will seem obviously true to her. Harry is familiar with the drug and knows that it has this effect. Assume that p is part of Harry's total evidence and that Harry knows that p is a reactive proposition. On the basis of testimony from a colleague, Harry comes to believe falsely that one of the other employees has slipped the drug into Harry's morning coffee. Harry starts thinking about what follows from p, competently deduces q and thereby comes to believe q. In fact, it seems obvious to Harry that q follows from p.

Just as in the case of *Red Lighting* and *Premier League Reports*, many are inclined to think that the misleading evidence Harry receives makes it irrational for him to believe q. What Christensen (2010a) notes, however, is that there are certain respects in which Harry's situation seems to differ from Suzy's and Jack's.

In *Red Lighting* and *Premier League Reports* it seems that once Suzy and Jack receive misleading evidence in *Part 2*, there is a very straighforward sense in which their total evidence no longer provides the same support for the propositions that Suzy and Jack came to believe in *Part 1*[91].

---

[90] This is a slightly modified version of Christensen's (2010a:187) own case.

[91] I should note that this evaluation of Suzy's and Jack's situation is not entirely uncontroversial. If, for example, one thinks that Jack can continue to know in *Part 2*, and if one thinks that a subject's evidence consists of all and only the propositions known by the subject, then the proposition that Man U won the

In DRUG, however, the change in evidential support is much less straightforward. The false testimony Harry receives about having been drugged does not put into question whether p is true and it is therefore natural to think that p is still part of Harry's evidence after he has been told that he has been drugged. P entails q, and q is therefore plausibly still entailed by Harry's total evidence after he has been falsely informed that he has been drugged. If we were to spell out evidential support in terms of the probability that a proposition has on a subject's evidence, it looks as if q has probability 1 on Harry's total evidence even after Harry has been falsely informed about being drugged. And since Harry hasn't in fact been drugged, he is fully capable of tracing out the consequences of his evidence by competently deducing q from p.

According to Christensen (2010a), strong misleading higher-order evidence can also leave a subject's evidential support for a proposition unaffected in cases where the subject's total evidence does not entail the target proposition.

Imagine, for example, a situation in which my evidence is best explained by the truth of q although q does not follow from my evidence. Imagine, however, that I am falsely informed that I am prone to misevaluate what my evidence supports or what would best explain my evidence.

In this situation, Christensen (2010a) thinks that even though I have been misled about my ability to evaluate my evidence, my total evidence still supports q or is best explained by q, and I am still capable of competently reasoning from my evidence to q.

---

match is still part of Jack's evidence in *Part 2*, and will therefore still be entailed by Jack's evidence in *Part 2*. I will set such complications aside for present purposes.

In cases like the above, Christensen (2010a:195) thinks that there is a sense in which the misleading higher-order evidence forces us to 'bracket' some of our first-order reasons and makes it impossible for us to give a certain part of our evidence 'its due'. By this, Christensen (2010a) means that higher-order evidence can require subjects or agents:

> "…to violate or compromise certain rational ideals. (...) HOE [higher-order evidence] can put agents in a position where they cannot trust their appreciation of the first-order reasons. So even if they see clearly what a certain epistemic ideal - such as respecting logic or IBE [Inference to the best explanation] requires, they cannot simply allow their beliefs to obey these requirements. That is what is meant by saying they must (to at least some extent) bracket their first-order reasons. And the result of this, of course, is that their beliefs end up falling short of the relevant ideals." (Christensen 2010a:198)

Instead of *rational ideals*, Lasonen-Aarnio (2014a) talks of *epistemic rules* or *rules of rational belief formation*. One candidate for an epistemic rule is a principle telling one to update one's credences by conditionalization. Another example of a candidate epistemic rule is the following rule concerning the acquisition of beliefs by perception:

Perception: Believe p when it perceptually seems as if p and there is no (sufficiently good) reason to think that in the present situation, how things seem isn't a good guide to how they are (Lasonen-Aarnio 2014a:322)

Just as Christensen (2010a) claims that Harry would violate a rational ideal by not believing q in the face of misleading evidence that he has been drugged, Lasonen-Aarnio (2014a) claims that there are a number of cases in which subjects violate epistemic rules or rules of rational belief formation by revising their beliefs in the face of misleading higher-order evidence.

As an illustration of the idea, Lasonen-Aarnio (2014a) notes that **Perception** has a build-in proviso to deal with defeat cases like *Red Lighting*, but has no provisos concerning misleading higher-order evidence:

> "…consider a situation in which I believe that it is snowing outside based on its perceptually seeming to me as if it is snowing, but I also have strong but misleading evidence that I have been given a drug that seriously impairs my ability to form beliefs based on following Perception: the drug makes me prone to misjudge how things seem to me. This is not evidence that, in my present circumstances, how things seem is not a good guide to how they are. Moreover, because the higher-order evidence is misleading, it does in fact perceptually seem to me as if it is snowing. Hence, in my circumstances Perception still recommends believing p." (Lasonen-Aarnio 2014a:322)

One might, of course, think that this reveals that **Perception** is not a correct epistemic rule, but Lasonen-Aarnio (2014a) thinks that we ultimately have to accept that there are a number of cases in which subjects violate correct epistemic rules by revising their beliefs in the face of misleading higher-order evidence.

Lasonen-Aarnio (2014a) therefore agrees with Christensen (2010a) that there are a number of cases in which subjects violate certain principles of rationality by revising their beliefs in the face of strong misleading higher-order evidence.

She does not, however, agree with Christensen (2010a) about what implications such cases should have for our theory of rationality. Lasonen-Aarnio (2014a) is attracted to the idea that a belief is rational when formed in accordance with a correct epistemic rule, and she is therefore attracted to the thought that a subject S can rationally believe p although S has strong misleading evidence that the belief that p is the result of a flawed belief-forming process. As such, beliefs formed by following the recommendations of

Perception might be rational even though the subject holding those beliefs has strong misleading evidence that she is prone to misjudge how things seem to her. Lasonen-Aarnio (2014a) therefore thinks we should reconsider when misleading higher-order evidence with respect to a proposition p can defeat the rationality of believing p.

Christensen (2010a), on the other hand, is not attracted to this *indefeasibilist* line of thought. When Christensen (2010a) writes that higher-order order evidence is "…prone to being rationally toxic" (2010a:212) he does not mean that we are rationally required or permitted to retain our beliefs in the face of misleading higher-order evidence. Rather, he thinks that cases like DRUG force us into situations where the *most* rational or rationally *best* response is not rationally *perfect*. In DRUG, Harry is therefore not rationally permitted to believe the conclusion q although *not* believing q (e.g. by withholding on q) is rationally suboptimal, because it involves violating certain rational ideals (such as believing what one has competently deduced from one's evidence).

While those friendly to defeat by higher-order evidence are likely to find Christensen's (2010) view more palatable than Lasonen-Aarnio's (2014a), both views involve surprising and somewhat controversial claims about the rational status of beliefs revised in the face of certain kinds of higher-order evidence. As both authors realize, some might therefore attempt to resist the claim that we violate rational ideals or epistemic rules in the kinds of cases Christensen (2010) and Lasonen-Aarnio (2014a) focus on:

> "Why see these cases as showing that agents must violate rational ideals, rather than as showing that the deductive or inductive principles in question were not rational ideals to begin with?" (Christensen 2010a:203)

> "Above I claimed that correct epistemic rules need not and don't make provisions for all possible higher-order defeaters. But maybe I have failed to consider the kinds of epistemic rules that are ultimately correct. Someone might object: "If there are situations in which it is no longer rational to apply a rule R, then this shows that it was never a correct rule to start out with (…)"." (Lasonen-Aarnio 2014a:330).

Christensen (2010a) and Lasonen-Aarnio (2014a) therefore present a number of arguments in defense of their diagnosis. Christensen's (2010a) and Lasonen-Aarnio's (2014a) main arguments for the occurance of these violations proceed by pointing out supposedly unattractive consequences of rejecting that such violations take place[92].

Christensen (2010a) argues that rejecting his diagnosis involves giving up a very attractive picture of what is distinctive about higher-order evidence. Lasonen-Aarnio (2014a) argues that rejecting her diagnosis involves giving up a very attractive picture of how we manage our beliefs by following epistemic rules.

In what follows, I argue that if we violate rational ideals or rules of rational belief formation in the kinds of cases that Christensen (2010a) and Lasonen-Aarnio (2014a) seem attracted to thinking that we do, then Christensen's (2010a) proposal for what is distinctive about higher-order evidence and Lasonen-Aarnio's (2014a) proposal for how we manage our beliefs will have to be given up anyway. Their arguments do therefore not provide us with a reason for thinking that rational ideals or rules of rational belief formation are violated in the relevant kinds of cases[93].

---

[92] In what follows, I focus only on what I take to be their main arguments and leave other considerations they present in defense of their diagnosis for another occasion.

[93] In what follows, I will remain silent on whether we actually do violate rational ideals or rules of rational belief formation in the kinds of cases Christensen (2010a) and Lasonen-Aarnio (2014a) seem attracted to thinking that we do. My focus here is merely on whether the arguments they provide give us a reason for accepting this conclusion. I will also not discuss whether their arguments provide us with a reason for thinking that we *sometimes* violate rational ideals or rules of rational belief formation. In what follows I am

The plan for the remaining sections is as follows. In sections (**4**) and (**5**) I focus on Christensen's (2010a) argument from the supposed distinctiveness of higher-order evidence. In section (**6**) and (**7**) I turn to Lasonen-Aarnio's (2014a) argument for thinking that the rejection of her diagnosis is incompatible with what she calls the *rule-driven picture of epistemic cognition*.

4. The distinctiveness of defeat by higher-order evidence. Christensen's (2010a:186-7) discussion suggests that he takes the violation of rational ideals to occur in a number of different cases involving misleading higher-order evidence. His examples include the following kinds of cases:

> A: After having (competently) diagnosed a patient and prescribed medication you realize that you have been awake for 36 hours. As a result, you start to worry about possible diagnostic errors given your sleep-deprived state.
>
> B: You realize that someone you consider an epistemic peer has reached a different conclusion than you by reasoning from the same body of evidence (in fact your perceived peer has made a mistake).
>
> C: Someone you trust (falsely) informs you that you have been given a drug that makes you prone to make errors in reasoning.

In such cases, he takes subjects who revise their beliefs in the face of the misleading higher-order evidence to violate rational ideals. He is aware that some might resist his diagnosis of the relevant cases by claiming that his proposed[94] rational ideals are not the true rational ideals, but he argues that this way of resisting his diagnosis runs the risk of

---

merely interested in whether their arguments provide us with a reason for thinking that rational ideals or rules of rational belief formation are violated in a certain class of cases.

[94] Christensen (2010a) never provides a clear statement of any specific rational ideals, but one of them seems to be something like *believing what one has competently deduced from one's evidence*.

overlooking an "…interesting feature of the structure of epistemic rationality…" (2010a:204) by defining away an important difference between higher-order and first-order evidence. In particular, the proposed rejection of Christensen's (2010a) diagnosis:

> "…would obscure an interesting point about the special epistemic role played by higher-order considerations. We've seen that certain intuitively attractive epistemic ideals have to be compromised in particular situations. But when we've seen this occur, the evidence that requires this sort of compromise is higher-order. If this is right, then it turns out that HOE [higher-order evidence] (and perhaps, more broadly, an agent's higher-order beliefs) play an important and distinctive role, a role that would be hidden (…). We would miss out on seeing how an agent's beliefs about herself have a sort of broad systematic effect on epistemology that's not produced by, e.g., an agent's beliefs about trees, or numbers, or other agents." (2010a:204)

In other words, Christensen (2010a) thinks we should acknowledge that there are a number of cases in which belief-revision in the face of higher-order evidence leads to the violation of rational ideals, because this diagnosis of the relevant cases helps us to see what is distinctive about higher-order evidence[95,96].

In this and the following section I argue that if subjects violate rational ideals in the cases that Christensen (2010a) seems to think that they do, then there are plausibly also cases where subjects violate rational ideals by revising their beliefs in the face of misleading first-order evidence.

---

[95] In section (**2**) I mentioned the possibility of a view on which evidence that one's belief that p is false is always evidence that one has been irrational in believing that p. It may be that on such a view every case of defeat by first-order evidence is also at the same time a case of defeat by higher-order evidence. If this is right, then it will be difficult to make sense of the idea that there is something distinctive about defeat by higher-order evidence. For present purposes, I will set such potential complications for Christensen's (2010a) diagnosis aside.

[96] In his response to a self-undermining objection to conciliationist views of disagreement, Christensen (2013) provides an argument for the more general claim that responding rationally to our evidence *sometimes* involves violating rational ideals. For present purposes I will only focus on his (2010a) argument for thinking that belief-revision in cases like A, B, and C involves the violation of rational ideals.

Specifically, I argue that if Harry in DRUG violates a rational ideal by not believing what he has competently deduced from his evidence, then there are plausibly cases in which misleading first-order evidence makes subjects violate rational ideals by not believing what they have competently deduced from their evidence.

Consequently, Christensen's (2010a) claim that only higher-order evidence forces us to violate rational ideals is incompatible with thinking that Harry in DRUG violates rational ideals by not believing q. If Christensen (2010a) is right that the violation of rational ideals is a distinctive feature of rationally accommodating higher-order evidence, he is therefore wrong about the cases in which such violations occur.

For the purpose of my argument it will be convenient to start by assuming Williamson's (2000:184-208) equation of a subject's total evidence with all and only the propositions known by the subject (E=K). We can then later consider how the argument generalizes to other conceptions of evidence.

According to Christensen (2010a), DRUG involves the violation of rational ideals because Harry in DRUG is prevented from believing what he has competently deduced from his evidence. On the assumption of E=K, a similar situation arises in the following case involving defeat by first-order evidence:

CONJUNCTION[97]**.** Assume that a subject, S, knows each of the propositions p_1, …, p_n. Assume that S is in roughly the same epistemic position with respect to each of the propositions and roughly equally confident in all of them. Imagine, for example, that p_1, …, p_n are a number of propositions that S has learned by reading a book that is widely

[97] See Chapter 1 for a more detailed version of this kind of case.

accepted as trustworthy on the relevant issues. Since S knows p_1, …, p_n, the propositions are all part of S's total evidence according to E=K.

Suppose S competently deduces the conjunction $p_1 \wedge \ldots \wedge p_n$ and on that basis forms a belief in the conjunction. We can assume that S knows that the truth of the conjunction follows from the truth of the conjuncts. Consequently, $p_1 \wedge \ldots \wedge p_n$ follows from propositions that all belong to S's total evidence and S knows that $p_1 \wedge \ldots \wedge p_n$ follows from these propositions.

Imagine now that S receives strong misleading evidence that the conjunction $p_1 \wedge \ldots \wedge p_n$ is false. Imagine, for example, that someone S considers highly trustworthy convincingly informs S that the conjunction is false because exactly one of the conjuncts is false. The informant, however, provides no indication of which conjunct. Upon receiving this piece of misleading evidence, we would expect S to give up her belief in the conjunction. If S retains her belief in the conjunction, we are inclined to think that S's belief is irrational.

What is especially interesting for our purposes is whether the propositions p_1, …, p_n all continue to be part of S's total evidence after S has been told that one of the propositions is false? Given E=K, the question is whether S continues to know each of the propositions p_1, …, p_n.

If n is very low, we might expect S to give up belief in at least some (if not all) of the conjuncts upon receiving the strong misleading evidence. Even if S were to retain belief in all of the conjuncts it is compelling to think that the misleading evidence S has received will defeat S's knowledge of at least some (if not all) of the conjuncts. Whether or not S continues to believe all of the conjuncts, it is therefore plausible to think that if n is

very low, the misleading evidence that defeats knowledge of the conjunctive conclusion also results in a loss of knowledge of at least some of the premises. According to E=K, at least some of the propositions $p_1, \ldots, p_n$ therefore cease to be part of S's evidence when S receives misleading evidence that the conjunction $p_1 \wedge \ldots \wedge p_n$ is false.

What if n is very high? If n is sufficiently high we would not expect S to give up belief in all of the propositions $p_1, \ldots, p_n$ upon being told that one of them is false. In fact, given that S is roughly as confident with respect to each of the propositions $p_1, \ldots, p_n$ and given that the misleading evidence is not targeted at any particular proposition, we would not expect S to stop believing any of the propositions $p_1, \ldots, p_n$. As such, we would expect S to continue believing all of the propositions $p_1, \ldots, p_n$ if n is very high.

Does S also retain knowledge of all of the propositions $p_1, \ldots, p_n$ upon receiving evidence that one of the propositions is false? It does not seem particularly plausible to think that receiving misleading evidence defeats S's knowledge of some of the propositions while allowing S to retain knowledge of some of the other propositions. After all, the misleading evidence is not targeted at any particular proposition and S was in roughly the same epistemic position with respect to each of the propositions $p_1, \ldots, p_n$ before receiving E. At the same time, the claim that S loses knowledge of all of the propositions $p_1, \ldots, p_n$ looks overly skeptical. If n is very high and if S knew all of the propositions $p_1, \ldots, p_n$ before being told that the conjunction is false, then we are not inclined to think that S stops knowing all of the propositions upon receiving misleading evidence that one of them is false. As such, there is some attraction to the thought that S does in fact retain knowledge of all of the propositions $p_1, \ldots, p_n$ upon receiving misleading evidence that the conjunction is false. Consequently, there is some attraction to

the thought that on E=K all of the propositions p_1, …, p_n are still part of S's total evidence after S has received misleading evidence that the conjunction is false.

Plausibly, S also still knows that the conjunction follows from the conjuncts p_1, …, p_n. Consequently, $p_1 \wedge \ldots \wedge p_n$ follows from propositions that all belong to S's total evidence and S knows that $p_1 \wedge \ldots \wedge p_n$ follows from these propositions. At the same time, S has misleading evidence that makes it irrational for S to believe the conjunction $p_1 \wedge \ldots \wedge p_n$.

If Christensen (2010a) is right that Harry in DRUG violates rational ideals because he is rationally prevented from believing what he has competently deduced from his evidence, then S in CONJUNCTION similarly violates rational ideals because S is rationally prevented from believing what S has competently deduced from her evidence.

In CONJUNCTION, however, it seems that what makes it irrational for S to believe the conjunction $p_1 \wedge \ldots \wedge p_n$ is first-order evidence that $p_1 \wedge \ldots \wedge p_n$ is false[98]. If DRUG is a case

[98] Of course, evidence that $p_1 \wedge \ldots \wedge p_n$ is false is in some sense evidence that S's belief that $p_1 \wedge \ldots \wedge p_n$ is flawed. In particular, it might be suggested that evidence that $p_1 \wedge \ldots \wedge p_n$ is false is evidence that S has relied on a proposition in coming to believe $p_1 \wedge \ldots \wedge p_n$ that is not in fact part of S's evidence. As such, some might try to defend Christensen (2010a) by interpreting CONJUNCTION as a case of defeat by higher-order evidence. There are, however, several problems with diagnosing CONJUNCTION as a case of defeat by higher-order evidence. First of all, evidence that p is false is always in *some* sense evidence that the belief that p is flawed. Christensen (2010a), however, needs there to be a distinction between defeat by first-order evidence and higher-order evidence, and he therefore cannot simple interpret any defeat case as a case of defeat by higher-order evidence. And if it does not count as defeat by first-order evidence when expert testimony that $p_1 \wedge \ldots \wedge p_n$ is false makes it irrational to believe $p_1 \wedge \ldots \wedge p_n$, then we are at least owed an account of what defeat by first-order evidence is.
Secondly, diagnosing CONJUNCTION as a case of defeat by higher-order evidence seems to turn too many cases into cases of defeat by higher-order evidence. Consider, for example, *Premier League Reports* from section (**2**). In *Part 1* Jack has come to know that Man U won an important match against Chelsea. Suppose Jack from this knowledge competently deduces that Chelsea lost the match. If Jack now reads in *The Times* that Chelsea won, then this looks like a paradigmatic case of defeat by first-order evidence. Of course, reading *The Times* has also provided Jack with evidence that the proposition that Man U won was never known and therefore never part of Jack's evidence (assuming E=K). As such, reading *The Times* has provided Jack with evidence that he has relied on a proposition that was not actually part of his evidence in coming to the conclusion that Chelsea lost. But if that is correct, then Jack's situation looks parallel to CONJUNCTION. If CONJUNCTION is a case of defeat by higher-order evidence, then presumably Jack's situation is also a case of defeat by higher-order evidence. But presumably Christensen wouldn't want that?

in which the attempt to accommodate misleading higher-order evidence involves the violation of certain rational ideals, it therefore looks like CONJUNCTION is a case in which the attempt to accommodate misleading first-order evidence involves the violation of certain rational ideals (on the assumption that E=K).

On the assumption of E=K, Christensen's (2010a) diagnosis of DRUG is therefore incompatible with the idea that only higher-order evidence forces us to violate rational ideals.

5. Generalizing to other conceptions of evidence. Does the worry raised for Christensen's (2010a) argument arise on other conceptions of evidence? To answer that question, we should consider which features of E=K a case like CONJUNCTION trades on.

Plausibly, cases like CONJUNCTION are possible only if evidence is propositional. It is very difficult to make sense of the idea that the conjunction $p_1 \wedge \ldots \wedge p_n$ is *entailed by* or *follows from* the subject's evidence if evidence is thought of as non-propositional. There is, however, good reason to think that evidence is propositional (see, e.g., Williamson 2000:194-200), and in any case Christensen (2010a) should be happy to grant this assumption as his own diagnosis of DRUG relies on it.

What other characteristics must evidence have in order for a case like CONJUNCTION to arise? One suggestion is that CONJUNCTION trades on a failure of KK (i.e. that CONJUNCTION trades on the idea that we can know that p without knowing that we know p). That is, it might be suggested that once S in CONJUNCTION receives strong misleading evidence that the conjunction $p_1 \wedge \ldots \wedge p_n$ is false, S still knows p_1, …, p_n, but no longer knows that she knows p_1, …, p_n. Given E=K, S therefore no longer knows that the

propositions p_1, …, p_n are part of her total evidence and this is why it is not rational for S to believe the conjunction $p_1 \wedge \ldots \wedge p_n$ despite her ability to competently deduce the conjunction from propositions that are part of her total evidence. If this diagnosis is right, then one might think that situations like CONJUNCTION only arise on those conceptions of evidence on which we are not always in a position to know what our evidence is.

How far does the problem for Christensen (2010a) generalize if CONJUNCTION does indeed trade on a failure of KK? According to Williamson's (2000:93-113) famous anti-luminosity argument, there is no non-trivial condition such that if it obtains, we are always in a position to know that it obtains[99]. If that argument is correct, then the idea that we cannot always know what our evidence is will not be a peculiarity of E=K. In fact, if Williamson's (2000) anti-luminosity argument is correct, then we should expect that on more or less any conception of evidence we will not always be in a position to know what our evidence is. If that is true, and if CONJUNCTION trades on the idea that we cannot always know what we know, then it looks like situations parallel to CONJUNCTION might arise on many different conceptions of evidence.

There is reason to suspect, however, that CONJUNCTION does not rely on a failure of KK. In fact, it is not at all clear that S does not know that she knows each of the propositions p_1, …, p_n after she has received evidence that the conjunction $p_1 \wedge \ldots \wedge p_n$ is false. After all, we know that long books very often contain errors and misprints, and we therefore often have strong evidence that at least one of the claims of a long book is false. Likewise, we know that information received from sources such as TV, the Internet, our memory and the

[99] See, e.g., Cohen (2010), Vogel (2010) and Weatherson (2004) for objections to Williamson's (2000) anti-luminosity argument, and see Srinivasan (2013) for a response to those and other objections.

testimony of our friends is occasionally false. Yet, it seems implausible to claim that in all these cases we never or rarely know that we know[100].

The implausibility of denying higher-order knowledge in CONJUNCTION as well as these other cases is also brought out by considering the consequences of this denial when combined with the idea that knowledge is the norm of assertion (see Williamson 2000:238-269). According to the knowledge norm, one should assert p only if one knows that p, but if S in CONJUNCTION does not know for each of the propositions p_1, …, p_n that she knows it, then it is never permissible for her to assert "I know that p_1". Generalized to our other sources of knowledge, we would often not be permitted to assert "I know that p" for some p that we have come to know by watching TV, by reading on the Internet etc. But that is highly revisionary with respect to our actual practices of assertion and looks rather implausible. As such, there is much plausibility to the thought that S does not merely know each of the propositions p_1, …, p_n, but also knows that she knows p_1, …, p_n even after she has received evidence that $p_1 \wedge \ldots \wedge p_n$ is false.

While CONJUNCTION does not seem to rely on a failure of KK, there is some attraction to the thought that it relies on the compatibility of knowledge and uncertainty. When receiving evidence that the conjunction $p_1 \wedge \ldots \wedge p_n$ is false we would expect S to be less than certain about the truth of each of the propositions p_1, …, p_n. If having p as part of one's evidence requires being certain that p, then the propositions p_1, …, p_n would

[100] Notice also that there is a strong inclination to think S at least knew that she knew p_1, …, p_n prior to receiving evidence that the conjunction is false. The argument for thinking that S can continue to know each of p_1, …, p_n after having received the relevant piece of misleading evidence can presumably be restated as argument for thinking that S can continue to know that she knows each of p_1, …, p_n. If we accept that S can continue to know p_1, …, p_n after having acquired evidence that the conjunction is false, we are therefore under significant pressure to accept that S can likewise continue to know that she knows.

plausibly not be part of S's evidence after S has received evidence that $p_1 \wedge \ldots \wedge p_n$ is false. Since knowing that p is compatible with being less than certain that p[101], the fact that S is less than certain about the truth of the propositions p_1, …, p_n after she has received evidence that the conjunction is false does not mean that S does not know the truth of p_1, …, p_n. Given E=K, S's uncertainty about the truth of p_1, …, p_n does not prevent the propositions p_1, …, p_n from being part of S's evidence.

If this is right, then CONJUNCTION plausibly trades on the compatibility of knowledge and uncertainty, and cases similar to CONJUNCTION will therefore plausibly be restricted to conceptions of evidence on which having a proposition p as part of one's evidence is compatible with being less than certain about p.

There are views on which evidence is required to be certain, and these views might therefore seem to avoid cases like CONJUNCTION. Standard subjective Bayesianism is perhaps the most prominent example[102]. According to standard subjective Bayesianism, a subject S's doxastic attitude towards a proposition p is characterized as the subjective probability or credence S assigns to p, where the probability or credence is represented by a number $x \in [0;1]$. The propositions that constitute a subject's evidence are assigned subjective probability or credence 1, where assigning credence 1 to a proposition p is standardly understood as being certain that p. If standard subjective Bayesianism can

[101] The compatibility of knowledge and uncertainty should not be confused with a view sometimes called *fallibilism* about knowledge. According to fallibilism, a subject S can know p although S's evidence does not entail p (see, e.g., Dougherty & Rysiew (2009) and Littlejohn (2011) for such statements of fallibilism about knowledge). Acknowledging the compatibility of knowledge and uncertainty does not commit one to fallibilism about knowledge. One might, for example, accept that knowledge is compatible with uncertainty while holding that E=K, in which case p is entailed by one's evidence if one knows p.

[102] On Jeffrey conditionalization things are more complicated (Jeffrey 1983). See, e.g., Williamson (2000, chap. 10) for a discussion.

provide a plausible account of what propositions could play the role of evidence in the theory, it therefore seems that Bayesians would have a way of avoiding problem cases parallel to CONJUNCTION.

It is difficult to see, however, what propositions Bayesians or anyone else could invoke as evidence that we are always certain about. It might be tempting to suggest that we cannot be uncertain about how things seem or appear to us, but there are substantial reasons to think that this is not true.

As I have already noted, Williamson's (2000) anti-luminosity argument raises serious worries about whether we are always in position to know how things appear to us, and if we are not always in position to know how things appear to us, then presumably we aren't (and shouldn't) always be certain about how things appear to us.

Even without Williamson's (2000) highly sophisticated anti-luminosity argument there are reasons to think that we are not (and should not) always be certain of how things seem or appear to us. Consider a couple of intuitive examples of how we form beliefs about the world on the basis of how things seem or appear. I form the belief that there is a ruby red object in front of me because it looks like there is a ruby red object in front of me. It sounds like someone is playing the clarinet, so I form the belief that someone is. The beer I am drinking tastes like weissbier, so I form the belief that it is weissbier. These seem like prime examples of beliefs formed on the basis of appearances.

While I know how ruby red looks, how the clarinet sounds and how weissbier tastes, I am not absolutely certain. In fact, I used to be confused about what ruby red looks like, how the clarinet sounds and what weissbier tastes like, and as a consequence of my previous confusion I am rarely (if ever) certain about how things appear to me.

For example, I used to mistake magenta red for ruby red and so took objects to look ruby red when in fact they were magenta red. It wasn't that magenta red objects looked ruby red to me. Rather, I didn't know how magenta red and ruby red look and therefore took objects to look ruby red when in fact they didn't[103]. Now I know how ruby red and magenta red look, but given my former confusion, I realize that my judgments about how things look are not infallible and I am therefore rarely (if ever) certain about how things look.

Similarly, I used to mistake oboes for clarinets and used to mistake lager beer for weissbier. As such, I mistook the sound of an oboe for the sound of a clarinet and the taste of a lager for the taste of a weissbier. It wasn't that oboes sounded like clarinets to me, or that lagers tasted like weissbier. Rather, I didn't know how oboes and clarinets sound, or how lagers and weissbier taste. I therefore took instruments to sound like the clarinet and beers to taste like weissbier when in fact they didn't. Now I know what a clarinet sounds like and what a weissbier tastes like. I realize, however, that how things sound or taste is not always transparent to me, and so I am rarely (if ever) certain about how things sound or taste.

Perhaps I am not considering the kind of propositions about appearances that the proponents of certainty about evidence have in mind, or perhaps there are other proposals for propositions that can serve as evidence that we are always certain about. There is, however, reason for skepticism about the prospects of giving a plausible account of

[103] The argument here relies on the possibility of possessing a colour concept like RUBY RED even if one is mistaken about exactly what the extension of the concept is. As such, the argument relies on some form of externalism about the concepts and mental contents we should characterize an individual as possessing or entertaining. See, e.g., Burge (1979, 1982) for a more detailed description of a version of externalism about mental content individuation.

evidence on which it is both certain and capable of justifying or rationalizing our beliefs about the external world. Unless it can be established that CONJUNCTION depends on some other peculiarity of E=K, we should therefore suspect that cases parallel to CONJUNCTION are possible on a range of different conceptions of evidence.

The idea that it is a distinctive feature of higher-order evidence that it sometimes forces us to violate rational ideals cannot motivate the idea that Harry in DRUG violates rational ideals by not believing q. If Christensen (2010a) is right that only higher-order evidence forces us to violate rational ideals, then he is mistaken about the cases in which such violations occur.

6. The rule-driven picture of epistemic cognition. Parallel to Christensen (2010a), Lasonen-Aarnio (2014a) presents a number of cases that she seems to think of as potential candidates for cases in which we violate epistemic rules by revising our beliefs in the face of misleading higher-order evidence. Her examples include cases like the following:

> B: You realize that someone you consider an epistemic peer has reached a different conclusion than you by reasoning from the same body of evidence (in fact your perceived peer has made a mistake).
>
> C: Someone you trust (falsely) informs you that you have been given a drug that makes you prone to make errors in reasoning.
>
> D: After having (competently) gone through a sequence of reasoning to check that everything is set for your climb down the mountain, you acquire strong evidence that you are in serious danger of suffering from hypoxia[104] caused by the high altitude (in fact you are not hypoxic).

[104] Hypoxia is a condition caused by oxygen deprivation. The condition can cause hard-to-detect cognitive impairments.

Lasonen-Aarnio does not explicitly argue that these are cases in which we violate epistemic rules by revising our beliefs in the face of misleading higher-order evidence. Instead, the conclusion of Lasonen-Aarnio's (2014a) argument is the more general claim that there *are* circumstances in which we violate the correct epistemic rules by revising our beliefs in the face of strong misleading higher-order evidence. She does, however, suggest that this general conclusion makes cases like B, C, and D potential candidates for cases in which we violate epistemic rules by revising our beliefs in the face of the relevant kinds of higher-order evidence.

Below I provide some reasons for thinking that her argument does not support the idea B, C, and D are cases in which such violations occur[105].

Let me, however, start by presenting Lasonen-Aarnio's (2014a) argument for the claim that there are circumstances in which we violate the correct epistemic rules by revising our beliefs in the face of misleading higher-order evidence[106]. Her argument starts from the thought that it is very difficult to provide informative statements of epistemic rules that take all higher-order evidence into account.

Consider again **Perception** from above:

Perception: Believe p when it perceptually seems as if p and there is no (sufficiently good) reason to think that in the present situation, how things seem isn't a good guide to how they are.

---

[105] I do hereby take a stand on whether Lasonen-Aarnio's (2014a) argument gives us reason to believe that there are *some* cases in which we violate correct epistemic rules by trying to accommodate strong misleading higher-order evidence. My focus here is on cases like B, C, and D.

[106] In what follows I provide a reconstruction of what I take to be Lasonen-Aarnio's (2014a) main argument. She does, however, make a number of interesting observations throughout her discussion that I cannot engage with or do justice to here.

As was noted in section (**3**)**, Perception** does not have a build-in proviso to deal with higher-order evidence. If I am falsely informed that I have been given a drug that makes me prone to misjudge how things seem, **Perception** still recommends belief. **Perception** is therefore not a correct epistemic rule according to a view on which correct epistemic rules have build-in provisos that allow them to accommodate higher-order evidence.

As an alternative, we could suggest the following modification of **Perception**:

Perception*: Believe p when it perceptually seems as if p, there is no (sufficiently good) reason to think that things are not how they seem and no (sufficiently good) reason to think that one is disposed to midjudge how things seem.

Perception* faces the same problem as **Perception**; one can receive misleading evidence that one has misapplied **Perception*** or that **Perception*** is not a correct epistemic rule.

I might, for example, receive misleading evidence that I have been given a drug that makes me prone to ignore my reasons for thinking that things are not how they seem. If someone told me that an object I am looking at is illuminated by red light, the drug would make me block out that reason for thinking that the object is not in fact red. Despite the misleading higher-order evidence, **Perception*** still recommends belief in my current situation.

These problems for **Perception** and **Perception*** raise the following worry:

> "…even when a rule has some (…) provisos built into it, it need not have provisos for all possible higher-order defeaters built into (…) it—and it is at least a prima facie challenge to show how this would even be possible." (Lasonen-Aarnio 2014a:323-324)

Given the problems for **Perception** and **Perception***, Lasonen-Aarnio (2014a:332) thinks that we should be skeptical about the prospects of providing a finite and informative statement of a rule of perceptual belief-formation that takes into account all kinds of higher-order evidence. She worries that for any modification of **Perception** that is finite and informative we can come up with circumstances in which the modified rule recommends belief even when we have misleading higher-order evidence that the rule has been misapplied or is incorrect.

She takes the problem to generalize to other candidate epistemic rules as well.

One reaction to Lasonen-Aarnio's (2014a) argument is to give up the attempt to provide short and simple statements of epistemic rules along the lines of **Perception**. Instead, one might propose an infinitely long (or at least incredibly complicated) epistemic rule that takes into account any possible situation involving higher-order evidence.

While Lasonen-Aarnio (2014a) acknowledges that such a rule may be possible, she thinks that accommodating higher-order evidence by providing such a complicated epistemic rule is in tension with some of the central motivations for what she calls the *rule-driven picture of epistemic cognition*. Spelling out the worry requires a bit of background.

As Lasonen-Aarnio (2014a:333) points out, a number of philosophers[107] have been attracted to the thought that beliefs are rational or justified when they are formed and retained in a way that is appropriately tied to the goal or aim of belief, where the aim of belief is typically taken to be truth or knowledge[108]. By this some have understood that a

[107] See, e.g., Bonjour (1985:7-8), Chisholm (1982:4), Conee (1992), Wedgwood (2002).
[108] There are intricate questions about what we should understand by the idea that beliefs aim at truth or knowledge and how this idea relates to the notions of rationality and justification. I cannot hope to do justice to the complicated epistemological discussions of these questions here. For present purposes, I focus

belief is rational or justified only if (or just in case) it is true (or known[109]). Often, however, the idea is that a belief is rational or justified when it is formed and retained in a way that *generally* leads to true (or known) beliefs in a certain range of circumstances (e.g. in 'normal' circumstances) even if the token belief is not itself true (or known).

At least initially, it seems that epistemic rules provide a natural way of spelling out the proposed tie between rationality or justification on the one hand and the aim or goal of belief on the other. The idea is that a belief is rational or justified when it is formed in accordance with the correct epistemic rules, and when one forms one's beliefs in accordance with the correct epistemic rules, then one is managing one's beliefs in a way that is appropriately tied to the aim or goal of belief (see Wedgwood 2002).

Those attracted to a very *direct* tie between justification and the aim of belief might therefore be attracted to an epistemic rule like the following:

Knowing: Believe that p just in case you know that p

Those attracted to a more *indirect* tie might take a principle like **Perception** to be a good proposal for a correct epistemic rule.

According to Lasonen-Aarnio (2014a:333-334), proponents of what she calls the *rule-driven picture of epistemic cognition* are not merely thinking of epistemic rules as rules that state the conditions under which a belief is rational or appropriately aimed at truth. What is crucial to the *rule-driven picture* is that the correct epistemic rules are rules that can *offer guidance* in attaining the aim of belief (i.e. in attaining truth or knowledge).

---

exclusively on Lasonen-Aarnio's (2014a) attempt to paint a broad picture of how these questions connect to what she calls the *rule-driven picture of epistemic cognition*.

[109] See, e.g., Sutton (2005).

Because we cannot simply believe truths at will, we need rules that we can follow and be guided by in our belief formation, and the correct epistemic rules are therefore supposed to be the rules that actual subjects genuinely follow and are guided by[110]. Epistemic rules are the tools that actual subjects use in their attempt to obtain true beliefs (or knowledge).

One might worry that epistemic rules like **Knowing** cannot provide the kind of guidance that proponents of the *rule-driven picture of epistemic cognition* are after, and for exactly that reason Lasonen-Aarnio (2014a) focuses on epistemic rules like **Perception** that seem to connect naturally with a more *indirect* understanding of the tie between justification and the aim of belief:

> ”Whatever the truth (or knowledge) connection amounts to, rules such as “Believe p just in case p is true” won’t serve as the kind of means for pursuing truth that we are looking for, as such rules don’t offer enough guidance as to what exactly it is that one should do in order to believe the truth. What we are looking for is something by means of which true belief and knowledge can be obtained.” (2014a:333)

We can now spell out the tension between the *rule-driven picture of epistemic cognition* and the attempt to avoid the problems for **Perception** by providing an infinite or incredibly complicated epistemic rule that has build-in provisos for any kind of higher-order evidence. Just as **Knowing** might seem too uninformative to provide any guidance, the infinitely long restatement of **Perception** seems too complicated to be a rule that actual subjects could follow and be guided by in their belief-formation. As Lasonen-Aarnio (2014a) writes, such a rule

---

[110] Boghossian (2008) provides a detailed discussion of what he calls the *rule-following picture of rational belief* with special attention to a number of difficulties he takes to arise when we try to get clear on what it means to follow epistemic rules. He does, however, claim to find the rule-following picture ”…as natural and as compelling as the next person…” (2008:473).

> "…just don't seem like the kind of rule that can offer genuine guidance. It differs from all of the candidate rules that epistemologists have formulated and taken actual subjects to be guided by." (Lasonen-Aarnio 2014a:333-334)

It is therefore not clear that such a rule is

> "…at all fit to play the kind of guidance-role that proponents of a rule-driven picture commonly require (correct) epistemic rules to play." (Lasonen-Aarnio 2014a:335)

If we insist that epistemic rules have build-in provisos to deal with every kind of higher-order evidence, we therefore run the risk of undermining the *rule-driven picture of epistemic cognition* where the correct epistemic rules are rules that actual subjects manage to follow in their attempt at attaining true beliefs.

Now, one could try to accommodate higher-order evidence within the *rule-driven picture of epistemic cognition* by proposing a set of epistemic rules that specifically concern how to revise our beliefs in the face of higher-order evidence. Such rules would occasionally conflict with rules such as **Perception**, but in such circumstances the recommendations of one rule would have greater weight and would therefore overrule the recommendations of the other. If, for example, one has evidence that one is prone to misjudge how things seem, then the recommendations of **Perception** might be overruled by the recommendations of a rule concerning how to revise one's belief in the face of evidence that the belief is flawed.

According to Lasonen-Aarnio (2014a), this attempt to accommodate higher-order evidence is not ultimately compatible with the *rule-driven picture of epistemic cognition*.

To get the above kind of overruling to happen, we need a weighting or ordering of the various epistemic rules, where the ordering can be understood as a kind of meta-rule recommending which rules to follow in any given set of circumstances. The problem, however, is that the meta-rule will have to be infinitely long or at least incredibly complicated in order for it to be able to provide recommendations for which rules to follow in all possible circumstances (including every kind of circumstance involving misleading higher-order evidence). But then the question arises of how we manage to follow the recommendations of the meta-rule? Just as the attempt to formulate epistemic rules that accommodate all kinds of higher-order evidence leads to rules that cannot provide any guidance, the meta-rule will be too complicated to provide any guidance.

According to Lasonen-Aarnio (2014a), the attempt to accommodate defeat by higher-order evidence therefore pushes us towards a kind of epistemic particularism, because we are forced into a view on which "…there are no useful, informative generalisations about what the epistemically rational ways of managing one's beliefs are" (2014a:334), and actual subjects therefore have no such generalizations to guide them in their belief formation.

Since Lasonen-Aarnio (2014a:333-335) takes the *rule-driven picture of epistemic cognition* to provide a good framework for understanding how it is possible for us to manage our beliefs in ways that generally lead us to believe truths, she thinks we should be hesitant to give up this picture.

Instead, we should accept that there are circumstances in which strong misleading higher-order evidence does not have defeating force, because none of the correct epistemic rules have provisos for relevant kinds of higher-order evidence. It may, for example, be that it is perfectly rational for me to follow the recommendations of **Perception** although I

have misleading evidence that I am prone to misjudge how things seem. Perhaps there just are no rules telling me to stop following **Perception** in circumstances where I have evidence that I am prone to misjudge how things seem.

Of course, when I have evidence that I am prone to misjudge how things seem I have evidence that my attempts to follow **Perception** are likely to be unsuccessful. This, however, does not mean that I am less than perfectly rational if my attempts to follow **Perception** succeed. If, on the other hand, I take such higher-order evidence into account when it is misleading and stop forming beliefs on the basis of how things seem to me, then I am violating the correct epistemic rules by failing to follow the recommendations of **Perception**.

If Lasonen-Aarnio's (2014a) argument is correct, then a question arises about what kinds of higher-order evidence the correct epistemic rules have provisos to handle? Are we often in circumstances where following the correct epistemic rules involves disregarding (misleading) higher-order evidence or do the correct epistemic rules have provisos concerning higher-order evidence of the kinds we typically encounter?

Lasonen-Aarnio (2014a) does not try to settle this question. Her discussion does, however, suggest that she considers it at least an open question whether the correct epistemic rules have provisos for the kinds of misleading higher-order evidence we receive when we encounter peer disagreement, when we acquire evidence that we are hypoxic, or when we acquire evidence that we have been drugged. As such, she seems to think that cases like B, C, and D are potential candidates for cases in which we violate the correct epistemic rules by trying to take the relevant kinds of misleading higher-order evidence into account.

7. The usefulness of higher-order evidence. Below I argue that if none of the correct epistemic rules have provisos for higher-order evidence of the kind we receive in cases like B, C, and D, then the *rule-driven picture of epistemic cognition* does not look like a plausible picture of how we manage our beliefs. A desire to preserve the *rule-driven picture of epistemic cognition* therefore cannot motivate the idea that we violate the correct epistemic rules by revising our beliefs in cases like B, C, and D.

Before presenting my argument for this conclusion, I should note that there are independent reasons to question the plausibility of the *rule-driven picture of epistemic cognition*. According to Lasonen-Aarnio (2014a), it is central to the *rule-driven picture* that we follow and are guided by epistemic rules in our belief formation. Guidance here means something relatively strong, since she seems to think that rules such as **Knowing** do not provide the kind of guidance that proponents of the *rule-driven picture* are after.

This claim about how we manage our beliefs looks like a substantial psychological claim, and it is not at all obvious why we should accept it. Whether we are *guided* by epistemic rules in our belief forming practices looks like an empirical claim that epistemologists are not particularly well placed to answer. And it is even more difficult to see what could justify claims about which rules provide sufficient guidance and which rules do not.

If defeat by higher-order evidence is in tension with the *rule-driven picture of epistemic cognition* and the rules that epistemologists have taken to provide the requisite guidance, it is therefore questionable that our motivations for accepting the *rule-driven picture* are strong enough to choose the *rule-driven picture* over an epistemology that can accommodate defeat.

While I think the concern mentioned above presents a challenge for Lasonen-Aarnio's (2014a) argument, my main goal in this section is to argue that the commitments of the *rule-driven picture* as characterized by Lasonen-Aarnio (2014a) fits badly with the idea that subjects violate epistemic rules in cases like B, C, and D.

I do not take a stance on whether this means that we should give up the *rule-driven picture* or whether there are plausible candidates for epistemic rules that are capable of offering guidance and that are not violated in cases like B, C, and D. My point here is merely that cases of peer disagreement, cases of misleading evidence that one is hypoxic, or cases of misleading evidence that one has been drugged cannot be the kinds of cases in which we violate epistemic rules if Lasonen-Aarnio's (2014a) characterization of the *rule-driven picture* is correct.

The argument starts from the following observation. It should be uncontroversial that we often use higher-order evidence to manage our beliefs. When we receive strong evidence that our beliefs are the result of a flawed process we tend to give up those beliefs. In general, this is not a bad practice, since higher-order evidence on many occasions isn't misleading. Using such evidence to manage our beliefs therefore helps us minimize the number of false beliefs we have. In our attempt to acquire true beliefs and avoiding false ones, higher-order evidence therefore seems to play an important role (see Christensen (2010a:213) for a similar observation).

Evidence that we have been drugged, evidence that we suffer from hypoxia, or the evidence that peer disagreement can provide us with respect to the correctness of our beliefs is not always misleading. People are occasionally drugged and do occasionally suffer from hypoxia, and this occasionally makes them form beliefs that are irrational or even false. Similarly, we occasionally disagree with our peers because we have made an

error in reasoning. The higher-order evidence we acquire by being told that we have been drugged, by being told that we are hypoxic or by realizing that a peer disagrees with us can therefore sometimes help us to avoid flawed or false beliefs when the evidence in question is not misleading. In fact, it is quite plausible that we sometimes do make use of higher-order evidence of exactly these kinds to avoid flawed or false beliefs[111].

Initially, it might not seem particularly complicated for Lasonen-Aarnio (2014a) to account for how we make use of these kinds of higher-order evidence. While she suggests that the correct epistemic rules might not have provisos concerning these kinds of higher-order evidence, she could simply say that the relevant kinds of higher-order evidence helps us figure out whether our beliefs are in fact formed in accordance with the correct epistemic rules.

If I have been drugged or if I am hypoxic, then I tend to form beliefs in ways that are in conflict with the correct epistemic rules, and when I receive evidence that I have been drugged or evidence that I am hypoxic, I therefore receive evidence that my beliefs are not formed in accordance with the correct epistemic rules.

Typically, I will respond to this evidence by giving up the relevant beliefs, and when the evidence in question is not misleading I will thereby have used the relevant higher-order evidence to get rid of flawed or false beliefs. Of course, even strong higher-order evidence will sometimes be misleading, and in those circumstances the relevant higher-order evidence will lead me to give up beliefs that are in fact formed in accordance with the correct epistemic rules. Misleading higher-order evidence can therefore

[111] As the growing literature on the epistemology of disagreement reveals, there are difficult questions about how and when encountering peer disagreement affects the epistemic status of our beliefs. Nonetheless, it looks somewhat implausible to claim that we never can and do make use of the higher-order evidence that peer disagreement provides us with (see, e.g., Christensen & Lackey (2013) for some recent contributions to the disagreement debate).

sometimes make me violate the correct epistemic rules, because it leads me to revise my beliefs in circumstances where the correct epistemic rules say that I shouldn't. Or at least that is what Lasonen-Aarnio (2014a) might say.

The problem for Lasonen-Aarnio (2014a) is that this explanation of how we make use of higher-order evidence does not seem plausible if we accept her characterization of the *rule-driven picture of epistemic cognition*.

To spell out the problem, consider the following line of thought. According to the *rule-driven picture*, we rely on the guidance of epistemic rules to help us manage our beliefs. If the correct epistemic rules have no provisos concerning how we should revise our beliefs in the face of higher-order evidence of a certain kind, then we have nothing to guide us in how to make use of the relevant kind of higher-order evidence. Proponents of the *rule-driven picture* therefore cannot explain how we make use of certain kinds of higher-order evidence in managing our beliefs if the correct epistemic rules do not have provisos concerning higher-order evidence of the relevant kinds.

This means that if none of the correct epistemic rules have provisos concerning the kinds of higher-order evidence we acquire by being told that we have been drugged, by being told that we are hypoxic or by realizing that a peer disagrees with us, then proponents of the *rule-driven picture* cannot explain how we make use of the relevant kinds of higher-order evidence. But as noted above, we do sometimes make use of these kinds of higher-order evidence to avoid flawed or false beliefs. Consequently, the proponents of the *rule-driven picture* should insist that the correct epistemic rules include rules that do in fact have provisos for these kinds of higher-order evidence.

The idea that epistemic rules should offer guidance therefore cannot motivate the conclusion that we violate the correct epistemic rules by revising our beliefs in the face of

the higher-order evidence we acquire by being told that we have been drugged, by being told that we are hypoxic or by realizing that a peer disagrees with us. In fact, if we insist that none of the correct epistemic rules have provisos for these kinds of higher-order evidence, then the plausibility of the *rule-driven picture* is under significant pressure[112,113].

8. Conclusion. Christensen's (2010a) and Lasonen-Aarnio's (2014a) arguments do not provide us with a reason for thinking that rational ideals or rules of rational belief formation are violated in the kinds of cases that Christensen (2010a) and Lasonen-Aarnio (2014a) suggest as candidates for cases in which such violations occur.

---

[112] This point seems to generalize to a variety of cases in which someone might suggest that we violate epistemic rules by revising our beliefs in the face of misleading higher-order evidence. If the kind of higher-order evidence we acquire in such cases is higher-order evidence that we sometimes make use of to avoid flawed or false beliefs when the evidence in question is not misleading, then proponents of the *rule-driven picture* should hold that some of the correct epistemic rules have build-in provisos for these kinds of higher-order evidence. Otherwise, it will be rather difficult for proponents of the *rule-driven picture* to explain how we manage to make use of the relevant kinds of higher-order evidence when it is not misleading.

[113] There are at least two responses to my line of argument worth mentioning. The first response is to simply claim that we do not require epistemic rules to guide us when it comes to making use of higher-order evidence. The problem, however, is that this looks like a somewhat ad hoc proposal to make for proponents of the *rule-driven picture*. Why think that we require extensive guidance from rather specific epistemic rules in order to manage our beliefs more generally, but require no guidance in order to make use of higher-order evidence?

The second response is to claim that the correct epistemic rules have no provisos concerning *misleading* higher-order evidence, but do have provisos concerning the kinds of higher-order evidence we receive when it isn't misleading. It is therefore not a problem for proponents of the *rule-driven picture* to explain how we make use of higher-order evidence that isn't misleading while at the same time claiming that we violate epistemic rules by revising our beliefs in the face of higher-order evidence that is. One way of motivating this asymmetry could be by appealing to the idea that the *misleading* higher-order evidence is importantly different from higher-order evidence that isn't misleading. Assume, for instance, that someone tells me that I have been drugged. If the informant is telling the truth, I get to know that I have been drugged, and if the informant is lying, then I merely get to know that the informant is telling me that I have been drugged. If we assume E=K, the evidence I receive in the case where the informant is telling the truth will therefore be different from the case where he isn't. The idea could be that this difference in the evidence I receive is crucial to whether I violate epistemic rules by revising my beliefs in the face of the relevant evidence. The problem with this idea is that often the higher-order evidence we receive is exactly the same in the case where it is misleading and in the case where it isn't. Suppose someone tells me that I have been given a drug that causes 99% of all people to make errors in reasoning. Suppose I have been given such a drug and therefore come to know that I have. The higher-order evidence I have hereby received seems to be the same irrespective of whether or not I am among the unaffected 1%. If I violate epistemic rules by revising my belief in the face of this kind of higher-order evidence when it is misleading, this is presumably because the correct epistemic rules have no provisos for such evidence. But then it becomes difficult for proponents of the *rule-driven picture* to explain how I make use of such evidence when it is not misleading.

CHAPTER 5

Consuming Knowledge Claims Across Contexts

1. Introduction. Williamson (2005:100-101) and others (e.g. Hawthorne (2004:109-111), Rysiew (2012:286-291), Schaffer (2004:84-85)) have argued that contextualist theories of the semantics of 'know' have a special problem of accounting for our practices of 'consuming' knowledge attributions and denials made in other contexts. In what follows, I shall provide two interpretations of how to understand this problem. The first interpretation takes contextualism to have a special problem of accounting for how we are able to acquire epistemically useful information from knowledge attributions made in other contexts. The second interpretation takes contextualism to have a special problem of accounting for how we are able to understand and report the semantic content of knowledge attributions made in other contexts. I respond to the first interpretation by arguing (a) that the defeasibility of knowledge makes it difficult for everyone to acquire epistemically useful information from knowledge claims made in other contexts, and (b) that there is no *special* problem for contextualism when it comes to acquiring epistemically useful information from knowledge claims made in other contexts. I respond to the second interpretation by pointing to a number of resources available to the contextualist that become salient when we consider the behavior of other, more uncontroversially context-sensitive terms.

2. Contextualism and the outline of a problem. I begin this section with a schematic formulation of contextualism. I then present some quotations from Williamson (2005) which suggest the two alternative interpretations of the supposed problem for contextualism.

According to epistemic contextualism (e.g., DeRose (2009:2-3)), the truth-conditions (and semantic content[114]) of knowledge attributing and denying sentences such as ”S knows that p”/”S doesn’t know that p” can vary from one context of utterance to another. What vary between contexts are the epistemic standards that a subject S must meet with respect to a proposition p in order for ”S knows that p” to be true (where ”S doesn’t know that p” is true if S fails to meet the contextually determined standards).

While there are various ways of spelling out the contextualist framework, contextualists typically claim that what is at stake in a context and/or what possibilities of error are made salient in a context play a role in fixing the epistemic standards that S must meet in order for ”S knows that p” to be true as uttered in that context. If, in a context C, not much depends on whether p is true and/or no serious error-possibilities have been made salient in C, the standards that S must satisfy for ”S knows that p” to be true as uttered in C will be comparatively low. If, on the other hand, the truth of p is of utmost importance and/or serious possibilities of error have been made salient in C, the standards S must satisfy for ”S knows that p” to be true as uttered in C will be comparatively high.

Although schematic, the above outline of the contextualist framework will suffice for the purposes of presenting a problem raised for contextualism in Williamson (2005:100-101). What Williamson points to are difficulties that contextualists are supposed to face when they attempt to account for ”...the preservation of information in memory and its transmission by testimony” (2005:100). The starting point for the articulation of the problem is the following observation concerning context-sensitive sentences:

[114] This paper focuses exclusively on indexical contextualism as opposed to MacFarlane’s (2009) non-indexical alternative.

> ”Suppose that in a context *c* you acquire an item of information, expressed by ‘P’ as uttered in *c*. You could store the sentence ‘P’ in your memory. But when the time comes to retrieve the information, for guiding action or passing on to someone else, you will be in a new context *c**; even if ‘P’ expresses a truth as uttered in *c*, it need not express a truth as uttered in *c**.” (2005:100)

Simply storing the context-sensitive sentence ‘P’ in memory therefore does not seem sufficient to allow for the successful retrieval and transmission of the information expressed by ‘P’ in the initial context of utterance. While there are ways of working around the problem of retrieving and transmitting the information expressed by ‘P’ in *c*, Williamson (2005) is sceptical about how far these strategies can take us:

> ”If in *c** you remember that you acquired ‘P’ in *c*, and what *c* was like, you might be able to construct another sentence ‘P*’ that as uttered in *c** expresses the very truth that ‘P’ expresses as uttered in *c*. But that procedure imposes a heavy burden on memory. Most people cannot remember where they got much of the information on which they rely.” (2005:100)

In other words, the extent to which people would be required to keep track of various facts about *c* in order to allow for the successful retrieval and transmission of information acquired by utterances of ‘P’ in *c* is (implausibly) demanding and in any case seems empirically unfounded. In the case of ‘know’ it therefore seems that

> ”...the more extensively ‘know’ varies in reference with context, the worse the problems of using it to preserve and transmit information will be.” (2005:101)

These difficulties with explaining the possibility of retrieving and transmitting information is a problem for a contextualist treatment of ‘know’, because

> "...'know' does not seem to be designed like 'here' and 'now' primarily for immediate consumption; we need to preserve and transmit information about who has what kind of knowledge, or who knew what when."[115] (2005:101)

In what follows I shall provide two different interpretations of the slogan that 'know' is not *designed primarily for immediate consumption*. Each of these interpretations will give rise to a different objection to contextualism. In succeeding sections, I consider and respond to these objections in turn.

3. The first interpretation of immediate consumption. On the first interpretation, the slogan that 'know' is not *designed primarily for immediate consumption* is understood as the idea that knowledge attributions are sources of epistemically useful information that we can access and make use of not only in the context of attribution, but in other contexts as well. In particular, an utterance of the form "S knows that p" is evidence that S is in a certain epistemic position with respect to the proposition p, and the access to this evidence does not seem restricted to the initial context of utterance.

This leads to an objection to contextualism that can be stated as follows. As a matter of fact, we retrieve and transmit epistemically useful information from knowledge attributions made in other contexts all the time. If contextualism is true, however, this practice would require that we keep track of what is at stake, what error-possibilities are salient and/or what epistemic standards are operative in the various knowledge attributing

[115] In a somewhat similar spirit, Hawthorne (2004:109-111) argues that, if contextualism is true, we appear semantically blind to the shiftiness in the semantic value of 'know' in ways that make it "...very unclear, on the one hand, how we could manage to preserve information in memory, and, on the other, how we could avoid finding ourselves with hundreds of false beliefs on account of failure to update in line with shifting standards." (2004:111).

contexts. But engaging in that kind of *epistemic scorekeeping*[116] is implausibly demanding. Moreover, the hypothesis that we do engage in such scorekeeping in any case seems unfounded if we consider how much people actually remember about the sources of their information. Contextualism therefore appears to have the consequence that we are often unable to retrieve epistemically useful information from utterances such as ”S knows that p”. This inability, however, seems to be in conflict with our ordinary practice of retrieving and transmitting information from knowledge attributions made in other contexts.

Rysiew[117] (2012:286-291) seems to be getting at a similar point when he claims that ‘non-traditional views’ (such as contextualism, subject-sensitive invariantism, and relativism) struggle to explain:

> ”...how (...) [a knowledge report] can serve as a piece of common coin, usable by and useful to any number of people or groups not part of the immediate situation from which it emerges.” (2012:286).

And further:

> ”...how knowledge reports come to be useful to and fit for ‘consumption’ by the inquiring public at large.” (2012:286)

In an argument against contextualism, Schaffer (2004:84-85) uses as a premise the similar idea that:

> “Indexicality precludes ‘knows’ from scoring the overall progress of inquiry, because indexicals cannot keep a consistent score across contexts…” (2004:85)

---

[116] The expression ‘epistemic scorekeeping’ is borrowed from Rysiew (2012).
[117] See also Henderson (2009) for a similar challenge to contextualism.

In the following sections I will provide two responses to this kind of objection to contextualism. In section (**4**) I focus on the consequences that the defeasibility of knowledge has for our need for epistemic scorekeeping. I argue that anyone who accepts that we can lose knowledge by receiving misleading evidence will have to accept that we are sometimes unable to keep track of 'who knew what when' and we are therefore sometimes unable to retrieve epistemically useful information from knowledge claims made in other contexts. More importantly, our ability to keep track of 'who knew what when' and to retrieve epistemically useful information from knowledge claims across contexts is significantly improved if we engage in certain kinds of epistemic scorekeeping.

These observations suggest that the worry raised for contextualism is at risk of overstating how easily we retrieve epistemically useful information from knowledge claims made in other contexts or is mistaken about the level of epistemic scorekeeping we actually engage in.

In section (**5**) I argue that there is no difference in the level of epistemic scorekeeping that contextualist and rival semantics are committed to. What semantic theory is correct for 'know' makes no difference to the kind or level of epistemic scorekeeping we need to engage in to retrieve epistemically useful information from knowledge attributions and denials made in other contexts.

4. Defeasibility and epistemic scorekeeping. In this section I focus on the consequences that the defeasibility of knowledge has for our need for epistemic scorekeeping.

Although my observations in the following are not restricted to any particular semantics for 'know', it will simplify my discussion to assume a moderate insensitive invariantist (MII) semantics for 'know'. If insensitive invariantism is true, then changes in

what is at stake and/or what error-possibilities are salient in the context of utterance do not affect the semantic content or the truth-value of knowledge attributing and denying sentences. In all contexts of utterance, the semantic content of a knowledge claim of the form "S knows that p"/"S doesn't know that p" is simply that S knows or doesn't know that p (and whether S knows is not affected by what is at stake for S or what error-possibilities are salient to S). Furthermore, the standards are low enough for many of our ordinary knowledge claims to be true (as denied by skeptical invariantists).

If we have a sentence of the form "S knows that p" in our 'belief box' (Hawthorne 2004:110) or store the sentence in our memory (Williamson 2005:100), then if MII is the correct semantics for 'know' and if we have a time index on the sentence, we can always retrieve the information semantically encoded in the sentence, namely the information that S knows that p (or knew that p at time t).

While it is true that retrieving the semantic content of knowledge attributions across conversational contexts is relatively easy if insensitive invariantism is true, the question is how useful or epistemically beneficial this really is.

Of course, knowledge claims are sometimes false, in which case relying on the information semantically encoded will not be very beneficial. Let us, however, for present purposes focus exclusively on knowledge claims that are true and known to be true by those making them.

Take first a case involving memory. Suppose I read in *The Guardian* that Man U won an important match against Chelsea. Man U did in fact win the match and it therefore seems that I come to know that Man U won the match. In fact, it seems plausible that I know that I know that Man U won the match.

I need to keep track of 'who knew what when', so I store the sentence "Emil knows that Man U won the match" (WON) in my memory. To save memory capacity, I don't store any information about the evidence or source of information on the basis of which I came to know that Man U won. Later that day, I read in *The Times* that Man U lost the match. I consider whether to trust *The Times*, and find WON stored in my memory. Assuming insensitive invariantism is true and assuming that WON is time indexed, I can easily retrieve the semantic content of WON, namely that I knew (at time t) that Man U won the match. This is so, even though I have no information about the evidence or source of information on which I originally based my conclusion that Man U won. While I am able to retrieve the semantic content of WON, it is not clear how useful it is in my current situation. Assuming that knowledge is defeasible, I don't know anymore that Man U won the match, and having WON stored in my memory does not allow me to know that I knew earlier today that Man U won the match. As such, I have lost track of 'who knew what when'.

We can construct a parallel case involving testimony rather than memory. Assume that I ask "who won yesterdays match". Harvey is sitting at the other end of the room reading *The Guardian*. He reads on the front page that Man U won. Man U did in fact win and Harvey therefore comes to know that they did. He says "I know. Man U won the match". Just as Harvey is uttering these words, I read on the front page of *The Times* that Man U lost. Although Harvey is speaking the truth (and plausibly even knows that he is), and although I am able to retrieve the semantic content of his utterance, I cannot come to know

that he knows[118]. As such, I am unable to retrieve the epistemically useful information from his utterance (i.e. that he knows) and am therefore barred from keeping track of 'who knew what when'.

Keeping track of 'who knew what when' is therefore not perfectly easy even if insensitive invariantism is true. What is particularly interesting, however, is that engaging in certain kinds of epistemic scorekeeping significantly improves our ability to keep track of 'who knew what when' and allows us to retrieve epistemically useful information from knowledge claims that would otherwise be of limited use.

To see why, consider the following modifications of the memory case and the testimony case from above. In the modified memory case I have come to know that Man U won by watching the match on TV rather than by reading the result on the front page of *The Guardian*. In the modified testimony case Harvey has similarly come to know that Man U won by watching the match on TV rather than by reading the match report on the front page of *The Guardian*.

If I haven't stored information about the evidence or source of information on the basis of which I have come to the conclusion that Man U won, then having WON stored in my memory will not be of much use when I read in *The Times* that Man U lost. The problem is that if I don't know whether WON is stored in my memory because I read the

[118] There is an interesting question about how this kind of observation interacts with a motivation sometimes given for accepting the knowledge norm of assertion. Fricker (2007) writes: "The basic force of asserting that P to an audience is that the asserter vouches for the truth of P, she offers to her audience her word that P. Correlative with this (conventionally constituted and mutually known) import of the speech act of assertion is the fact that it is governed by the norm: one should assert P only if one knows that P." (2007:104). In this and other places Fricker (2007) seems to suggest that the knowledge norm of assertion in part is motivated by how it combines with fact that testimony is a source of knowledge. A proper view of the role of defeaters in communication may at least complicate our picture of the connection between the norm of assertion and the circumstances under which we can acquire knowledge by testimony.

result on the front page of *The Guardian* or because I saw the match on TV, then plausibly I cannot rely on WON to rule out the (false) match report in *The Times*.

If, on the other hand, I know that I stored WON in my memory because I saw the match on TV, then I am plausibly able to use the epistemic information encoded in WON (i.e. that I knew at time t that Man U won) to rule out the correctness of the match report in *The Times*. The epistemic scorekeeping involved in remembering the evidence or source of information on the basis of which I came to know that Man U won the match therefore allows me to keep track of 'who knew what when' in a situation like this (and thereby allows me to retrieve epistemically useful information from WON that I would otherwise not have been able to make use of).

The situation is similar in the testimony case. If I know that Harvey has watched the match on TV, then I can make use of the epistemic information encoded in his self-attribution of knowledge to rule out the correctness of the match report in *The Times*. Otherwise, I can't. Even though epistemic scorekeeping may not be necessary to retrieve the semantic content of knowledge claims made in other contexts, epistemic scorekeeping therefore significantly improves my ability to keep track of 'who knew what when' and allows me to retrieve epistemically useful information from the relevant knowledge claims[119].

[119] How much epistemic scorekeeping improves our ability to keep track of 'who knew what when' depends to some extent on how often we have misleading evidence and how often epistemic scorekeeping makes a difference as to whether that misleading evidence has defeating force. While these are empirical questions, it is worth noting that misleading evidence can take many different forms and are not restricted to false testimony of the kind considered here. Misleading evidence might, for example, be evidence about the likelihood of certain events, evidence of the reliability of certain sources of information, evidence about one's own ability to evaluate the evidence etc. It would therefore not be surprising if epistemic scorekeeping is often helpful in keeping track of 'who knew what when' even if insensitive invariantism is true.

What is interesting to note here is that the kind of scorekeeping that improves our ability to make use of knowledge claims made in other contexts if insensitive invariantism is true would also improve our ability to make use of knowledge claims made in other contexts if contextualism is true.

Suppose I have the sentence ”Harvey knows that Man U won the match” stored in my memory and suppose I know that Harvey’s information about the match is from the front page of *The Guardian*. Then even if contextualism is true I have the information needed to determine whether Harvey’s epistemic position with respect to the outcome of the Man U match is good enough for me to assert ”Harvey knows that Man U won” in my current conversational context.

My knowledge of Harvey’s source of information about the outcome of the match even provides me with some evidence about the standards operative in the context in which knowledge was initially attributed to Harvey. It seems, for example, that the standards cannot have been that high if having read the match report on the front page of *The Guardian* is sufficient to satisfy them.

The kind of scorekeeping that helps us to make use of knowledge claims made in other contexts if insensitive invariantism is true therefore also helps us if contextualism is true.

While Hawthorne (2004), Rysiew (2013), Schaffer (2004), and Williamson (2005) all focus on *knowledge that*, other authors have primarily focused on *knowledge whether* in their discussion of why we need to keep track of ‘who knew what when’. Craig (1990) has famously proposed that a central (or the primary) function of the term ‘know’ is to flag reliable informants, where the idea is *roughly* that if I know that a subject S knows *whether*

p, then I have someone to go to if I want to know whether p myself[120]. We therefore need to keep track of who knows *whether*, because we need to keep track of potential sources of knowledge.

Just as in the case of *knowledge that*, the defeasibility of knowledge has the consequence that keeping track of whether a subject knows *whether* p is not perfectly easy even if insensitive invariantism is true. Imagine as before that by reading the front page of *The Guardian* Harvey has come to know that Man U won yesterday's match. Just as I read that Man U lost on the front page of *The Times*, Harvey says "I know that Man U won". Even though Harvey is speaking the truth and knows that he is, I cannot come to know that he knows *whether* Man U won by relying on his say-so (just as I cannot come to know that he knows *that* they won).

Furthermore, epistemic scorekeeping improves my ability to keep track of who knows *whether* p just as it improves my ability to keep track of who knows *that* p. If Harvey had come to know that Man U won by watching the match on TV and I had known that his beliefs about the match resulted from watching it on TV, then Harvey's utterance of "I know that Man U won" would have put me in a position to know that Harvey knows *whether* Man U won (just as it would have put me in a position to know that Harvey knows *that* Man U won).

As we have seen above, the defeasibility of knowledge has the consequence that it is not perfectly easy to keep track of 'who knew what when' even if insensitive invariantism is

[120] To be more precise, Craig's (1990) idea is that we use 'know' to flag reliable informants, and that we need to keep track of 'who knows whether' exactly because we need to keep track of who the reliable informants are. For present purposes, I will focus only on the related and Craig-inspired idea that we need to keep track of 'who knows whether', because we need to keep track of potential sources of knowledge.

true. An interesting further consequence of the defeasibility of knowledge is that even when we manage to keep track of 'who knew what when' by keeping track of who knows *whether* p, that knowledge can sometimes be rather useless. As noted above, Craig's (1990) idea was roughly that if I know that someone knows whether p, then I have someone to go to if I want to know whether p myself. The defeasibility of knowledge puts pressure on that idea. To see how, consider the following line of reasoning.

Suppose I read in *The Times* that Man U lost and suppose I have no further information about the match. On the basis of reading *The Times* I form the belief that Man U lost, but in fact *The Times* is mistaken and my belief is therefore false. I have no way of knowing that *The Times* is mistaken and I therefore take myself to know that Man U lost. Suppose Harvey knows that Man U won the match, and that he has come to know by watching the match on TV.

Suppose I overhear a conversation between Sally and Ann. Ann asks Sally whether Man U won last nights match. Sally responds, "Harvey just told me, but I have already forgotten. Ask him, he saw the match on TV and knows whether Man U won". I trust Sally and store the sentence "Harvey knows whether Man U won" (WHETHER) in my memory, but to save memory capacity I don't store any information about how Harvey has come to know.

In this kind of scenario, it is rather plausible that I come to know the truth of WHETHER (i.e. I come to know on the basis of Sally's say-so that Harvey knows whether Man U won). Sally knows that Harvey knows whether Man U won, I have no particular reason to distrust her in the current circumstances and I therefore come to know that Harvey knows whether Man U won.

Suppose I subsequently overhear Ann's conversation with Harvey. Ann asks Harvey who won yesterday's match, to which Harvey responds "Man U won". Since I haven't stored any information about Harvey's evidence or about how he has come to believe that Man U won, my knowledge that Harvey knows whether Man U won is plausibly destroyed, and I will not be in a position to know that Man U did in fact win. Despite having WHETHER stored in my memory and despite being fully capable of retrieving the semantic content of WHETHER, I am no longer able to retrieve epistemically useful information from WHETHER.

Notice, however, that if I had stored information about how Harvey has come to believe that Man U won (i.e. by watching the match on TV), then I would still be able to retrieve epistemically useful information from WHETHER and would be able to realize that my belief that Man U lost is mistaken. Epistemic scorekeeping therefore improves my ability to make use of my knowledge that Harvey knows whether Man U won.

As should be clear by now, it is not perfectly easy to make use of knowledge reports made in other contexts even if insensitive invariantism is true, but the usefulness of such reports is significantly improved if we engage in certain kinds of epistemic scorekeeping.

The objection raised for contextualism is therefore at risk of overstating how easily we retrieve epistemically useful information from knowledge claims made in other context or is mistaken about the level of epistemic scorekeeping we actually engage in.

5. Why there is no *special* problem for contextualism. Despite the above considerations pertaining to the defeasibility of knowledge, some might think that the scorekeeping

required to make use of knowledge claims across contexts is much more demanding if contextualism is true.

In this section I argue what semantic theory is correct for 'know' makes no difference to the epistemic scorekeeping that we need to engage in to retrieve information about the epistemic position of the subject about whom knowledge is attributed or denied.

Information about a subject's epistemic position with respect to a given proposition (e.g. that the subject knows the relevant proposition) is the kind of *epistemically useful information* that we can hope to retrieve from knowledge attributions and denials made about that subject. What I will argue in this section is that it does not require additional epistemic scorekeeping to retrieve this kind of *epistemically useful information* from knowledge attributions and denials if contextualism rather than insensitive invariantism is true.

What is true, however, is that a certain empirical claim that contextualists often make in defense of their view has consequences for the epistemic scorekeeping required to retrieve epistemically useful information for knowledge reports made in other contexts. Anyone who accepts this empirical claim will therefore have to accept the need for epistemic scorekeeping irrespective of what semantics they take to be correct for 'know'.

The relevant empirical claim is the claim that our ordinary practices of attributing and denying knowledge are sensitive to contextual factors such as variations in what is at stake and/or what possibilities of error are salient in the context of attribution. That is, people's willingness to attribute knowledge appears to decrease as the stakes are raised and/or additional error-possibilities are introduced, and conversely their willingness to deny that a given subject knows appears to increase as the stakes go up and/or further error-

possibilities are introduced. Standard low- and high-stakes cases, such as *Bank-* (DeRose 1992:913) and *Airport-cases* (Cohen 1999:58), are supposed exemplify this sensitivity in our ordinary practices of attributing and denying knowledge. The context-variability exemplified by these cases is not a consequence of accepting any particular semantics for 'know'. Neither does this context-variability directly entail the truth of any particular semantics for 'know'. Rather, the context-variability in question (insofar as it occurs) is a datum that any theory of the semantics of knowledge reports should be able to account for, whether by seeing it as revealing something about the *semantics* or by explaining it away as a consequence of, e.g., the *pragmatics* of knowledge attributions and denials.

The mainstream tendency in the debate over contextualism has been for both sides to accept that our willingness to attribute and deny knowledge to some extent is sensitive to contextual factors such as stakes and/or salient error-possibilities. Since the objection to contextualism currently under consideration does not have this empirical claim as its target, I will assume its truth for present purposes.

We can now see that it is the alleged shiftiness or context-variability in our ordinary practices - and not any particular semantics for 'know' - that is responsible for our need for epistemic scorekeeping. This is made clear by the following considerations. Assume that a subject S's epistemic position with respect to a proposition p is such that in a context where the stakes are low and no possibilities of error are salient ordinary speakers will typically be willing to attribute knowledge that p to S. Now, keep fixed S's epistemic position with respect to p. What the shiftiness or context-variability in our ordinary practice tells us is that ordinary speakers' willingness to attribute knowledge that p to S will decrease if the stakes are raised and/or additional error-possibilities are introduced. Conversely, ordinary speakers will be more inclined to deny knowledge that p

to S as the stakes are raised and/or additional error-possibilities are introduced. Now, what should we conclude about S's epistemic position with respect to p in the light of people's willingness to attribute or deny knowledge that p to S on some particular occasion? That, it seems, will depend on what is at stake and/or what error-possibilites are salient in the context of the attribution. As a consequence, reliably retrieving information about the epistemic position of the subject of a knowledge attribution or denial made in another context requires that we have some information about what is at stake and/or what error-possibilities are salient in that context (or alternatively some information about the subject's sources of information with respect to the target proposition).

Non-contextualist semantics such as, e.g., insensitive invariantism have no advantage here. The following considerations will help to illustrate why. Assume that some version of moderate insensitive invariantism (MII) is true. As noted in the last section, MII holds that the epistemic standards a subject must satisfy (or the epistemic position S must be in) for "S knows that p" to be true do not vary from context to context. Furthermore, the standards are low enough for many of our ordinary knowledge claims to be true.

Consider now a knowledge denial of the form "S doesn't know that p" made in a situation where the stakes are unusually high. For example, John might say to Mary "Smith doesn't know that the flight has a layover in Chicago" in a situation where it is very important that John and Mary get to Chicago. Now, maybe Smith does in fact satisfy the MII-conditions for "Smith knows that the flight has a layover in Chicago" to be true, and John's denial that Smith knows the relevant proposition is just the result of the unusually high stakes. If, however, we attend only to what MII takes to be the content of John's knowledge denial, we may on the basis of John's utterance mistakenly conclude that Smith

does not satisfy the MII-conditions for ”Smith knows that the flight has a layover in Chicago” to be true. As a consequence, we may in another context where the stakes are low miss out on the opportunity to truly attribute knowledge to Smith, which might have been avoided if we had attended to the unusually high stakes in the context of John’s denial of knowledge to Smith.

The opposite problem arises if, in contexts where the stakes are unusually low, ordinary speakers are sometimes willing to attribute knowledge, even though the MII-conditions are not satisfied. Consider for example Alice, who in relaxed settings may be willing to attribute knowledge to a pupil who correctly answers a difficult question in class or to a friend who correctly answers a question as part of a quiz. Strictly speaking both the pupil and the friend may fail to satisfy the MII-conditions required for Alice’s attributions of knowledge to be true.

Now, if we attend only to what MII takes to be the content of the relevant attributions, we may, on the basis of Alice’s utterances mistakenly conclude that her pupil and her friend satisfy the MII-conditions for these utterances to be true. As a consequence, we may end up falsely attributing knowledge to the pupil and the friend in other contexts, which might have been avoided if we had attended to the very low stakes in the original contexts of attribution.

In a number of cases, then, attending only to what the moderate insensitive invariantist (MII) takes to be the content of the relevant knowledge attributions and denials (rather than the stakes and/or salient error-possibilities) can lead to errors.

On the other hand, it seems that ordinary speakers may avoid these unfortunate potential consequences if they make efforts to attend to a variety of contextual factors.

These factors are precisely the ones that the contextualist claims go into fixing the semantic content of ”S (doesn’t) know(s) that p”[121].

The proponent of insensitive invariantism may suggest that, while attending to (what she takes to be) the content of ”S (doesn’t) know(s) that p” may have unfortunate consequences in some of these high- and low-stakes cases, such extensive contextual variations are actually quite rare. Attending to what the insensitive invariantist takes to be the semantic content of ”S (doesn’t) know(s) that p” is therefore generally a quite effective heuristic for getting at the epistemically useful information expressed by knowledge attributions and denial.

But even if we grant the empirical claim about contextual variations in stakes and/or salient error-possibilities underlying this proposal, this does not seem to provide the insensitive invariantist with an advantage over the contextualist when it comes to explaining how people reliably retrieve epistemically useful information from knowledge attributions made in other contexts.

On a contextualist construal, the moderate invariantist’s suggestion merely amounts to the claim that in cases where one has no information about the stakes and/or error-possibilities salient in the context of attribution or denial, it is generally a useful heuristic to assume that the epistemic standards operative in the context of attribution or denial are the same as those operative in one’s own context (unless one has reason to think that one is in a context with abnormally low or high standards).

---

[121] It is, of course, an empirical question exactly what the consequences are of implementing any particular strategy of this kind. Maybe carefully attending to what error-possibilities are actually salient goes along with a more general disposition to be overly attentive error-possibilities in a way that turns any situation into a high-standards context. If so, encouraging people to do so may not be very good advice.

Consequently, the first suggested objection to contextualism fails. As argued in section (**4**), the defeasibility of knowledge makes it difficult for everyone to acquire epistemically useful information from knowledge claims made in other contexts. Furthermore, everyone who accepts that there is a context-variability in our willingness to attribute and deny knowledge will have to accept that people have to engage in certain kinds of epistemic scorekeeping. Contextualists do not face any special pressure in this respect: anyone who accepts the relevant data must agree that we need to engage in such scorekeeping to retrieve epistemically useful information from knowledge claims made in other contexts.

I now turn to a second interpretation of the slogan that 'know' is not *designed primarily for immediate consumption*. This second interpretation gives rise to a second putative objection to contextualism.

6. The second interpretation of immediate consumption. On the second interpretation, the slogan that 'know' is not *designed primarily for immediate consumption* is understood as pertaining specifically to the *semantic* content expressed by a knowledge attribution.

While retrieving epistemically useful information from knowledge attributions or denials is (as I have argued) no more difficult if contextualism is true than if, e.g., insensitive invariantism is true, it seems that when it comes to understanding the *semantic* content of what is expressed by a knowledge attribution or denial the latter of the two views has an advantage. According to insensitive invariantism, the semantic content expressed by "S (doesn't) know(s) that p"[122] does not vary from one context of utterance to another. As a result, if insensitive invariantism is true, we are always in a position to at

[122] Assuming that 'p' is not context-sensitive.

least understand *what was said* by a particular attribution or denial of knowledge. A further question is then whether a subject's willingness to attribute or deny knowledge on any particular occasion actually reflects the content of what was said.

If contextualism is true, however, things are more complicated. In situations where we have limited information about the stakes and/or the error-possibilities salient in the context of a knowledge attribution or denial, it will presumably be difficult for us to understand the semantic content expressed by the relevant knowledge claim, since these factors go into determining what that content is. And if we cannot understand the semantic content of an attribution or denial of knowledge, it will be difficult for us to share beliefs and, more generally, thought contents pertaining to 'who knew what when' in the way we often take ourselves to do.

Having limited information about the relevant contextual factors seems also to make it difficult for a speaker to talk about and re-express the exact semantic content of the initial knowledge attribution. It seems that we cannot simply restate the initial knowledge attribution by uttering the sentence "S knows that p", since our current context may differ from the initial context of attribution in ways that affect the semantic content expressed by the sentence in question. Additionally, it seems that we cannot make what Cappelen & Lepore (2005) have called an *inter-contextual disquotational indirect report* such as "T said that S knows that p", since, once again, the present context may differ from the context of T's utterance in ways that affect the semantic content expressed by 'know'. As an illustration of the difficulties of disquoting sentences with context-sensitive terms, consider someone who gives a disquotational indirect report of Sarah's utterance of "I am hungry" by uttering "Sarah says that I am hungry". This report obviously fails to get at the

semantic content of Sarah's initial utterance (assuming Sarah isn't making the report herself). Disquotation can go terribly wrong when context-sensitive terms are involved.

Even in cases where we are well-informed about the initial context of attribution, the contextualist picture seems to make re-expressing the semantic content expressed by the initial knowledge attribution rather difficult. Of course, it may be possible to construct a sentence 'P' that in a different context expresses the same semantic content as the initial attribution of knowledge did.[123] But doing so can be both complicated and potentially problematic. Whereas the semantic content of Sarah's utterance of the sentence "I am hungry" can easily be re-expressed by uttering the sentence "Sarah is hungry" (assuming the time is kept fixed), it is less clear how to construct a new sentence expressing the semantic content of "S knows that p" that doesn't involve an extensive specification of stakes, error-possibilities, S's evidence etc.

The second objection to contextualism focuses on these problems with understanding and re-expressing the semantic content of knowledge attributions made in other contexts.

On this interpretation, the slogan that 'know' is not "...designed like 'here' and 'now' primarily for immediate consumption..." (Williamson (2005:101)) is taken to express the idea that the semantic content of knowledge attributions does not seem to have the properties of being difficult to understand and re-express in the way that contextualism seems committed to saying that it does.

[123] This possibility is also noted by Williamson (2005:100)

My response to this objection has two parts. The first part concerns cross-contextual understanding. The second part focuses on linguistic devices for re-expressing the semantic content expressed by knowledge attributions made in other contexts.

Externalist theories of mental content individuation provide the key to the first part of my response. I will argue that those attracted to externalist theories of mental content individuation should also be attracted to the idea that an utterance of a context-sensitive sentence can result in thoughts with the exact content semantically expressed by that utterance, even when the thinker has incomplete information about the context of utterance.

The idea is to suggest that even though contextualism has as a consequence that we cannot always fully grasp the propositional content semantically expressed by uses of 'know' in other contexts, that need not imply that people cannot share beliefs or thoughts with the exact propositional contents expressed by the relevant uses of 'know'. Contextualism can still account for the datum that people can understand and think about the exact contents of knowledge attributions made in other contexts, even if they do not completely grasp the relevant contents.

To get this proposal off the ground we need a short introduction to externalist theories of mental content individuation. According to externalists, there are situations in which a subject can appropriately be described as possessing concepts and as entertaining mental states the content of which the subject incompletely grasps. A social externalist example (e.g., Burge 1979,1982) can serve to illustrate the point.

Consider a subject, Mary, who has a number of beliefs expressible by use of the term 'jade'.[124] For example, she has beliefs expressible by sentences such as "hard pieces of rock consisting of the mineral jadeite are a kind of jade", "jade typically has a green colour", "sapphires, emeralds and rubies are not kinds of jade". Mary does not, however, have beliefs expressible by the sentence "nephrite is a kind of jade" or anything along those lines. As it turns out, the term 'jade' as used in Mary's linguistic community correctly applies to all and only hard pieces of rock consisting of either the mineral nephrite or the mineral jadeite. The question now arises of what concept Mary's use of 'jade' is appropriately seen as expressing, and, by extension, what mental states her use of sentences containing 'jade' should be seen as expressing. Should we construe her uses of 'jade' as expressing the community concept JADE that has both nephrite and jadeite in its extension, or should we construe her as using 'jade' to express a non-standard concept, call it JADE*, that does not include nephrite in its extension? According to social externalists, we should construe Mary as possessing and expressing the community concept JADE by her use of 'jade'. Although she herself is unable to specify the correct application of 'jade', her being part of a linguistic community in which 'jade' applies to all and only nephrite and jadeite gives her the concept JADE and mental states featuring this concept (almost[125]) for free.

Had Mary been in a linguistic community in which, unbeknownst to her, 'jade' applied only to the mineral jadeite, then her use of 'jade' would have appropriately been

---

[124] It may be worth noticing that 'jade' is *not* a natural kind term, as made clear by LaPorte (2004:94-100) and Hacking (2007).

[125] There do seem to be certain minimal requirements of competence with a word that one must satisfy in order to count as possessing the community concept expressed by the relevant word. Had Mary been strongly convinced that 'jade' applies to all and only dark shades of red, then we might have been more reluctant to describe her as possessing and expressing the concept JADE by her use of 'jade'.

seen as expressing the concept JADE*, and not JADE. That is, according to social externalists:

> "…we can conceive of counterfactual situations in which all the physical and non-intentional psychological states of a thinker are the same as in an actual case, but in which the thinker's intentional states vary because of a difference in some feature of his or her linguistic environment." (Crane 1991:16).

Social externalists hold that the individuation of a subject's mental state is in part *parasitic* or *dependent* on the subject's linguistic community.

My suggestion is that contextualists about 'know', too, can endorse a model analogous to the model endorsed by social externalists. I will first present the idea with respect to an uncontroversially context-sensitive term, and then show how the idea extends naturally to the case of 'know'.

Consider the following case involving the indexical 'here'. You call me from your holiday in Florence and utter the sentence "it is 37 degrees here". If I know that you are in Florence, then presumably your use of a context-sensitive term like 'here' will not cause problems for my understanding the semantic content of your utterance and for my forming a belief with that very content. Contrast this with a situation in which I know that you are on holiday, but have no idea where you are. Your utterance "it is 37 degrees here" still expresses the same proposition, but since I don't know where you are, I cannot fully grasp the proposition uttered.

Interestingly, however, this does not prevent me from successfully referring to and talking about your holiday location. The term 'there' serves as a linguistic vehicle allowing me to do so. In a conversation with a common friend about your holiday I can therefore

say things like ”it is 37 degrees there”, whereby I seem to succesfully talk about the temperature in Florence although I have no idea where you are on holiday.

The question now is what thoughts I am appropriately described as entertaining on the basis of our conversation and that I subsequently express by use of the sentence ”it is 37 degrees there”. My proposal is that the belief I form on the basis of our conversation and that I subsequently express by ”it is 37 degrees there” has exactly the same content as your initial utterance of ”it is 37 degrees here” expressed. As with Mary in the case of ‘jade’, although there is a sense in which I am unable to fully grasp the content of your utterance of the sentence ”it is 37 degrees here”, I should nonetheless be described as entertaining and expressing that very same content.

As in the case of ‘jade’, we can imagine a counterfactual case where you, unbeknownst to me, are on holiday in Florida rather than Florence. In this case the contents of the beliefs I form on the basis of our conversation and express by my utterance of ”it is 37 degrees there” will be different from what I would express if you were in Florence. My use of ‘there’ is in some sense *parasitic* on your use of ‘here’: what I express by ”it is 37 degrees there” depends in part on the context of your initial utterance. Just as the mental states Mary expresses by her use of ‘jade’ are dependent on her linguistic community, at least one of my mental states depends on the context of your initial utterance.

The idea generalizes to other context-sensitive terms as well. If ‘know’ is context-sensitive, the suggestion is that although we cannot always fully grasp the propositional content semantically expressed by uses of ‘know’ in other contexts, a kind of *parasitic* understanding might nonetheless allow us to share thought contents with the subjects that initially attributed knowledge. As in the case of ’here’, the contents of these thoughts would depend in part on the initial contexts of attribution for their individuation.

A natural worry with this suggestion concerning 'know' is that we don't have a term that (like 'there' in the case of 'here') allows us to express the supposedly incompletely grasped contents acquired on the basis of knowledge attributions made in other contexts. 'Know' itself seems ill-suited to express these contents, since (as noted above) it may express something different in the current context than in the initial context of attribution. Moreover, since we are currently assuming that we incompletely grasp the contents of the relevant attributions, we lack the understanding to get at the contents in a different way. Even if we did have a firm grasp of the relevant contents, constructing new sentences not containing 'know' to express them might not be particularly easy.

If, however, there is no way for us to re-express the contents semantically expressed by the initial knowledge attributions, then the suggested content externalist account will commit the contextualist to a wide range of mental states that people are in principle unable to express. This commitment is at best unattractive.

This line of thought leads us to what is perhaps the most pressing part of the objection articulated against contextualism at the start of this section, namely, that we need some story of how it is possible to *linguistically* get at or re-express the semantic content expressed by knowledge attributions made in other contexts. On its own, this objection constitutes an important worry for a contextualist treatment of 'know'. In addition, the plausibility of the above externalist-style suggestion also depends on the availability of some story of how to express beliefs formed in response to knowledge attributions made in other contexts. What can contextualists say in response to this worry?

One natural way for people to talk about utterances made in other contexts is by *disquotational indirect reports* (also sometimes called 'say that' reports). If, e.g., T has

uttered the sentence ”there are coffee beans in the kitchen cupboard” a natural way to report T’s utterance is by use of the sentence ”T said that there are coffee beans in the kitchen cupboard”. Reports of this form are also familiar in the case of ‘know’. Someone may, for example, say: ”T said that John knows which Oxford college is the oldest”. The problem, however, is that if contextualism is true, reports of this kind may not succeed at getting at the semantic content expressed by the knowledge attribution which is being reported. Making successful disquotational indirect reports is not as straightforward when context-sensitive terms are involved. In the previous section this was made vivid by considering the attempted indirect report of Sarah’s utterance ”I am hungry” by the utterance of ”Sarah says that I am hungry”. Cappelen & Lepore (2005:89) have gone as far as to suggest that the possibility of successful disquotational indirect reports of utterances containing an expression *e* is evidence that *e* is not context-sensitive. If that heuristic is correct, the prospects of providing a contextualist story of how to linguistically get at the semantic content expressed by knowledge attributions made in other contexts seem meager at best.

But Cappelen & Lepore’s heuristic is misguided. Their scepticism about the possibility of successful ‘says that’ reports of utterances containing context-sensitive terms stems from an excessively narrow focus on indexicals such as ‘I’ and ‘here’ (as in the case of ”I am hungry”). As Hawthorne (2006) has pointed out, there are a number of context-sensitive terms that are ”...much more amenable to disquotational reporting than ‘I’...” (2006:445). Hawthorne’s (2006) treatment of ‘nearby’ can serve as an example:

> ”Suppose Ernie is in New York City and I am in Birmingham. Ernie says ‘A nearby restaurant has good Vietnamese food’. I can report this by saying ‘Ernie said that a nearby restaurant has good Vietnamese food’, even though I am far away from him.” (Hawthorne 2006:444)

Successful disquotational indirect reports of utterances containing ‘nearby’ are sometimes possible even in cases where ‘nearby’ doesn’t refer to anything close to the subject making the report. Moreover, the success of such disquotational reports does not depend on the reporter knowing anything about the location of the speaker whose utterance is being reported. Just as ‘there’ allows me to refer to your holiday location even if I don’t know where you are, ‘nearby’ allows me to talk about locations close to Ernie even if I don’t know where he is.

The fact that ‘nearby’ is more amenable to disquotational reporting than ‘I’ is explained by differences in how the semantic values of ‘I’ and ‘nearby’ are determined by the context of utterance. Whereas the semantic value of ‘I’ is fixed in an entirely mechanical and automatic way, namely as the speaker, the semantic value of ‘nearby’ is determined in part by the speaker’s intentions and purposes, and perhaps sometimes by which location is salient in the context of utterance.[126] As these factors will vary from context to context, ‘nearby’ will sometimes pick out a locale close to the speaker, sometimes a locale close to the subject whose utterance is being reported and sometimes even a third locale.

[126] Leslie (2007) proposes a more detailed story of how the semantic value of context-sensitive expression like ‘nearby’ and ‘is ready’ is fixed. I have reservations about certain aspects of her proposal, but since the details of such an account are irrelevant for my argument I will not discuss them here.

While paradigmatic indexicals like 'here' and 'now' behave similarly to 'I'[127], there are a number of context-sensitive terms such as, for example, 'ready', 'local', and 'left' that exhibit a behavior similar to that of 'nearby' (see Hawthorne (2006) and Lesley (2007)). In other words, a number of context-sensitive terms allow for successful disquotational reporting.

If contextualism is true, should we expect 'know' to behave like 'I' and 'here' or like 'nearby' and 'local' when it comes to disquotation? In responding to this question, it is important to realize that the contextualist claim that the semantic value of 'know' is fixed by the context of attribution does not entail that it is always, say, what is at stake for the *attributor* that fixes the semantic value of 'know'. As DeRose (2009) points out:

> "...there is nothing in contextualism to prevent a speaker's context from selecting epistemic standards appropriate to the practical situation of the subject being talked about, even when the subject being discussed is no party to the speaker's conversation..." (2009:240)

Suppose, for example, that an attributor A is in a low-stakes situation and considering whether to attribute knowledge of the form "S knows that p" to a subject S who is in a high-stakes situation and contemplating whether to act on the proposition that p. In this situation, what is at stake for S may play a role in fixing the standards that S must meet for "S knows that p" to be true as uttered by A. Parallel to the case of 'nearby', the attributor's intentions and purposes, and perhaps the salience of S's practical situation, here seem to play a role in fixing the semantic value of 'know'. Since intentions, purposes and which (if

[127] Of course, 'here' and 'now' allow for slightly more latitude than 'I' when it comes to determining the exact semantic value in a given context. 'Here' can refer both to my room, to Oxford, to England etc.

any) practical situation is salient will vary from one context to another, a contextualist treatment of 'know' plausibly allows for a flexibility somewhat similar to that exhibited by 'nearby' and 'local'.

This flexibility provides the contextualist with resources to explain the possibility of re-expressing the semantic content initially expressed by knowledge attributions made in other contexts. If a speaker's use of 'know' can sometimes select epistemic standards appropriate to the situation of the subject of a knowledge attribution, it seems extremely plausible that a speaker's use of 'know' in a disquotational report can similarly select standards appropriate to the context of the attribution being disquoted. For anyone to claim that contextualism is committed to 'know' being different from 'nearby' and 'local' in this particular respect requires an argument.

If contextualism is true, we should therefore still expect there to be a number of situations in which 'say that' reports succeed in re-expressing the semantic content of the knowledge attributions reported. This goes some way towards responding to the objection to contextualism currently under consideration. Below, however, I want to consider a worry for my proposed solution to the problem of linguistically getting at the semantic content of knowledge attributions made in other contexts.

The worry is that while contextualism about 'know' may be compatible with successful 'say that' reports in a variety of situations, there is a problematic class of cases in which it is far from clear that 'say that' reports succeed if contextualism is true. As we saw above, there is some plausibility to the thought that a speaker's context can select the epistemic standards appropriate to the situation of a third person when the practical situation of that person is salient. It is much less clear, however, that something similar is possible when

the practical situation of the speaker is the salient one. When an urgent practical decision is salient in the speaker's own immediate context, there is some inclination to think that, e.g., what is at stake for the speaker will play a significant role in fixing the semantic value of 'know' if contextualism is true.

The problem, however, is that we often seem to use 'says that' reports exactly in circumstances where we are in fact facing such practical decisions, especially in circumstances where we are unsure of whether we can rely on a knowledge attribution made in another context. Because we are not confident enough to make a flat-out knowledge attribution, we may make a disquotational indirect report.

A couple of examples illustrate this point. Consider Fred and Sarah, who on a Friday are considering whether to go to the bank now[128] or whether to go on Saturday. After talking to Linda on the phone, Sarah says to Fred, "Linda claims to know that the bank is open on Saturday, although she probably hasn't considered that banks sometimes change their hours". Or, for a second example, consider John and Mary, who are in the L.A. airport contemplating taking a certain flight to New York. John and Mary need to go Chicago, and after talking to another passenger, Smith, John says to Mary, "that passenger, Smith, says that he knows that the New York flight has a layover in Chicago."

In both these cases, a practical decision is salient in the speaker's own immediate context, and hence one might think that what is at stake in the speaker's own context will play a role in fixing the semantic value of 'know'. Often, however, the knowledge attributions we purport to be reporting are made in a context where the stakes and other supposedly standard affecting factors are not the same as in our own current context. In our

[128] This example is modelled on DeRose's (1992:913) famous bank-cases, whereas the following example is modelled on Cohen's (1999:58) airport-case.

examples, the factors usually thought to determine the semantic value of 'know' differ between Linda's (or Smith's) initial contexts of attribution Sarah's (or John's) contexts of disquotation. As a consequence, one might worry that Sarah's and John's disquotational reports fail to get at the semantic content of the knowledge attributions they are reporting. If these reports fail, it seems that ordinary speakers are insensitive to their failure: contextualism will have the unfortunate consequence that ordinary speakers are blind to the unsuccessfulness of a number of 'say that' reports containing 'know'. This, then, looks like a problem for a contextualist treatment of 'know'.

Now to my response. I will not discuss here to what extent ordinary 'say-that' reports involving 'know' actually succeed or fail in cases where the reporter is concerned with a practical decision or an error-possibility salient in her own immediate context. Instead, I want to point out that we often make 'say-that' reports containing more uncontroversially context-sensitive terms, such as 'nearby', 'local', 'tall', and 'rich', in situations where our immediate concern is whether something counts as nearby or local to us, or as tall or rich given the standards operative in our own current context. And yet, just as in the case of 'know', we often seem to use 'say that' reports containing these context-sensitive terms exactly in circumstances where we are too unsure about the circumstances of the initial utterance to just flat-out restate it.

The term 'local' can serve as an example. If I am in Williamsburg talking to my friends about where to go for dinner, I may make an indirect report of the following form "I remember Ernie talking about a very good local vietnamese restaurant, but I can't recall whether it was in Williamsburg". Consider a similar case involving 'rich'. My friends and I are talking about new rich billionaires, and in this context I may say, "Last week Scott said that Alison is rich, but I don't know whether she is a billionaire".

Such reports are common in ordinary speech, and the utterances we purport to be reporting are sometimes made in situations very different from our own. Now, insofar as a contextualist about 'know' is committed to the failure of 'say that' reports involving 'know' in cases like those mentioned earlier, a contextualist about 'local', 'nearby', or 'rich' seems equally committed to the failure of 'say that' reports in cases like the ones just described. But surely such failures in our ordinary practice of 'say that' reports involving 'local' and 'rich' do not pose an objection to a contextualist treatment of these terms. So similar facts about failures in our ordinary practice of 'say that' reports involving 'know' should not be seen as an objection to a contextual treatment of 'know'.

If, on the other hand, the above 'say that' reports involving 'local' and 'rich' *succeed* in re-expressing the content of the reported utterances, a contextualist about 'know' may hold that 'say that' reports involving 'know' succeed in the same way. On this view, indirect disquotational reporting does not pose the contextualist any special problem at all.

The central focus in this section has been on 'say that' reports (e.g. "T says that S knows that p"). It is worth noting, however, that a number of the above considerations carry over to the case of straightforward restatements of the initial knowledge claims (e.g. "S knows that p"). There is therefore reason to think that contextualism is consistent with a view on which speakers in contexts where the stakes are substantially different can sometimes express the same content by uttering "S knows that p", even if no 'say that' report is involved. If, for example, the practical situation in which a knowledge claim was initially made is salient in my current conversational context, then the practical situation of that context (as opposed to my present context) may fix the semantic value of 'know' as uttered

by me. The flexibility of 'know' thereby allows me to re-express the content of the initial knowledge claim by my utterance of "S knows that p".

It appears, then, that the contextualist has a number of resources available to respond to an objection that takes contextualism to have a special problem of accounting for cross-contextual understanding and re-expression of the semantic content of knowledge attributions.

First, the contextualist about 'know' can use semantic externalism to explain how a speaker may entertain and use 'know' to express a content which he or she may not completely grasp.

Second, the contextualist about 'know' can use observations about the behavior of uncontroversially context-sensitive terms like 'nearby', 'local', and 'rich' to show how a number of context-sensitive terms allow for much more flexibility than 'I' and 'here' when it comes to disquotational reporting. 'Nearby', 'local', 'rich', and a number of other context-sensitive terms can be used in 'say that' reports to express the same content they expressed in the initial context of utterance. If 'know' is understood by analogy to some of these terms, then contextualism can explain how it is at least sometimes possible to re-express the semantic content of knowledge attributions made in other contexts. There may be cases in which such re-expression is not possible, but these cases should not be seen as posing a problem for contextualist treatments of 'know', since such cases will arise for terms like 'local' and 'rich' without this posing a problem for a contextualist treatment of these terms.

7. Conclusion. I conclude, then, that on the two proposed interpretations, the slogan that 'know' is not *designed primarily for immediate consumption* does not pose a serious problem for a contextualist treatment of 'know'.

Bibliography

Alston, W. 1988. 'An Internalist Externalism'. *Synthese*, Vol. 74, No. 3, pp. 265-283.

Bergmann, M. 2006. *Justification Without Awareness: A Defense of Epistemic Externalism*. Oxford University Press.

Boghossian, P. 2008. 'Epistemic Rules'. *Journal of Philosophy*, Vol. 105, No. 9, pp. 472-500.

Bonjour, L. 1980. 'Externalist Theories of Empirical Knowledge'. *Midwest Studies in Philosophy*, Vol. V, pp. 53-73.

——— 1985. *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press.

Burge, T. 1979. 'Individualism and the Mental'. *Midwest Studies in Philosophy*, Vol. 4, No. 1, pp. 73-121.

——— 1982. 'Two Thought Experiments Reviewed'. *Notre Dame Journal of Formal Logic*, Vol. 23, pp. 284-293.

Cappelen, H. & Lepore, E. 2005. *Insensitive Semantics*. Oxford: Blackwell.

Chisholm, R. M. 1982. *The Foundations of Knowing*. Minneapolis: University of Minnesota Press.

Christensen, D. 2007. 'Does Murphy's Law Apply in Epistemology? Self-Doubt and Rational Ideals,' *Oxford Studies in Epistemology 2*, pp. 3–31.

——— 2010a. 'Higher-Order Evidence'. *Philosophy and Phenomenological Research*, Vol. LXXXI, No. 1, pp. 185-215.

——— 2010b. 'Rational Reflection'. *Philosophical Perspectives*, 24, pp. 121-140.

——— 2013. 'Epistemic Modesty Defended'. In Christensen & Lackey's (eds.), *The Epistemology of Disagreement: New Essays*. Oxford University Press.

Christensen, D. & Lackey, J. (eds). 2013. *The Epistemology of Disagreement: New Essays*. Oxford University Press.

Cohen, S. 1988. 'How to be a Fallibilist'. *Philosophical Perspectives*, 2, pp. 91-123.

——— 1999. 'Contextualism, Skepticism, and the Structure of Reasons'. *Noûs*, Vol. 33, Supplement: *Philosophical Perspectives*, 13, pp. 57-89.

——— 2002. 'Basic Knowledge and the Problem of Easy Knowledge'. *Philosophy and Phenomenological Research*, Vol. 65, No. 2, pp. 309-329.

———2010. 'Luminosity, Reliability, and the Sorites'. *Philosophy and Phenomenological Research*, Vol. LXXXI, No. 3, pp. 718–30.

Cohen, S. & Comesaña, J. 2013. 'Williamson on Gettier Cases and Epistemic Logic'. *Inquiry*, Vol. 56, No. 1, pp. 15-29.

Comesana, J. 2006. 'A Well-Founded Solution to the Generality Problem'. *Philosophical Studies*, Vol. 129, pp. 27-47.

Conee, E. 1992. 'The Truth Connection'. *Philosophy and Phenomenological Research*, Vol. 52, No. 3, pp. 657–669.

Conee, E. & Feldman, R. 1998. 'The Generality Problem for Reliabilism'. *Philosophical Studies*, Vol. 89, pp. 1-29

Craig, E. 1990. *Knowledge and the State of Nature: An Essay in Conceptual Synthesis*. Oxford University Press.

Crane, T. 1991. 'All the Difference in the World'. *The Philosophical Quarterly*, Vol. 41, No. 162, pp. 1-25.

DeRose, K. 1992. 'Contextualism and Knowledge Attributions'. *Philosophy and Phenomenological Research*, Vol. 52, No. 4, pp. 913-929.

——— 2009. *The Case for Contextualism*, Vol. I. Oxford University Press.

Dokic, J. & Égré, P. 2009. 'Margins for error and the transparency of knowledge'. *Synthese*, Vol. 166, pp. 1-20.

Dougherty, T. & Rysiew, P. 2009. 'Fallibilism, Epistemic Possibility, and Concessive Knowledge Attributions'. *Philosophy and Phenomenological Research*, Vol. LXXVIII, No. 1, pp. 123-132.

Dretske, F. 2000. 'Epistemic Operators'. In Dretske, *Perception, Knowledge and Belief*. Cambridge University Press.

——— 2005. 'The Case against Closure'. In Steup & Sosa (eds.), *Contemporary Debates in Epistemology*. Blackwell Publishing.

Elga, A. 2013. 'The puzzle of the unmarked clock and the new rational reflection principle'. *Philosophical Studies*, Vol. 164, pp. 127-139.

——— unpublished. 'Lucky to be rational'. Available on author's webpage (Last seen: 08-09-2014): https://www.princeton.edu/~adame/papers/bellingham-lucky.pdf.

Fricker, E. 2007. 'Audi on Testimony'. In Timmons, Greco, & Mele (eds.), *Rationality and the Good: Critical Essays on the Ethics and Epistemology of Robert Audi*. Oxford University Press.

Gendler, T. & Hawthorne, J. 2005. 'The Real Guide to Fake Barns: A Catalogue of Gifts for Your Epistemic Enemies'. *Philosophical Studies*, Vol. 124, pp. 331-352.

Goldman, A. 1976. 'Discrimination and Perceptual Knowledge'. *Journal of Philosophy*, Vol. 73, No. 20, pp. 771-791.

——— 1979. 'What Is Justified Belief?'. In G. Pappas (ed.), *Justification and Knowledge*, Dordrecht: Reidel. Reprinted in A. Goldman, *Liaisons: Philosophy Meets the Cognitive and Social Sciences*, MIT Press (1992)

——— 1986. *Epistemology and Cognition*. Harvard University Press.

Goodman, J. 2013. 'Inexact Knowledge without Improbable Knowing'. *Inquiry*, Vol. 56, No. 1, pp. 30-53.

Hacking, I. 2007. 'The Contingencies of Ambiguity'. *Analysis*, Vol. 67, No. 296, pp. 269-277.

Harman, G. 1973. *Thought*. Princeton University Press.

Hawthorne, J. 2004. *Knowledge and Lotteries*. Oxford University Press.

——— 2005. 'The Case for Closure'. In Steup & Sosa (eds.), *Contemporary Debates in Epistemology*. Blackwell Publishing.

——— 2006. 'Testing for Context-Dependence'. *Philosophy and Phenomenological Research*, Vol. 73, No. 2, pp. 443-450.

Hawthorne, J. & Lasonen-Aarnio, M. 2009. 'Knowledge and Objective Chance'. In Greenough & Pritchard's (eds.), *Williamson on Knowledge*. Oxford University Press.

Hawthorne, J. & Srinivasan, A. 2013. 'Disagreement Without Transparency: Some Bleak Thoughts'. In Christensen & Lackey's (eds.), *The Epistemology of Disagreement: New Essays*. Oxford University Press.

Henderson, D. 2009. 'Motivated Contextualism'. *Philosophical Studies*, Vol. 142, No. 1, pp. 119-131.

Jeffrey, R. 1983. *The Logic of Decision*, 2nd edition. Chicago: University of Chicago Press.

Klein, P. 1976. 'Knowledge, Causality, and Defeasibility'. *Journal of Philosophy*, Vol. 73, No. 20, pp. 792-812.

——— 1981. *Certainty: A Refutation of Skepticism*. University of Minnesota Press.

Kvanvig, J. 2006. 'Closure Principles'. *Philosophical Compass*, No. 1/3, pp. 256-267.

Lackey, J. 2008. *Learning From Words: Testimony as a Source of Knowledge*. Oxford University Press.

LaPorte, J. 2004. *Natural Kinds and Conceptual Change*. Cambridge Studies in Philosophy and Biology. Cambridge University Press.

Lasonen-Aarnio, M. 2010. 'Unreasonable Knowledge'. *Philosophical Perspectives*, Vol. 24, Issue 1, pp. 1-21.

——— 2014a. 'Higher-Order Evidence and the Limits of Defeat'. *Philosophy and Phenomenological Research*, Vol. LXXXVIII, No. 2, pp. 314-345.

——— 2014b. 'The Dogmatism Puzzle'. *Australasian Journal of Philosophy*, Vol. 92, No. 3, pp. 417-432.

Leslie, S-J. 2007. 'Moderate Sensitive Semantics'. In Preyer & Peter (eds.) *Context-Sensitivity and Semantic Minimalism*. Oxford University Press.

Littlejohn, C. 2011. 'Concessive Knowledge Attributions and Fallibilism'. *Philosophy and Phenomenological Research*, Vol. LXXXIII, No. 3, pp. 603-619.

——— 2012. *Justification and the Truth-Connection*. Cambridge University Press.

MacFarlane, J. 2009. 'Nonindexical Contextualism'. *Synthese*, Vol. 166, No. 2, pp. 231-250.

Nozick, R. 1981. *Philosophical Explanations*. Oxford University Press.

Pollock, J. 1986. *Contemporary Theories of Knowledge*. Savage, MD: Rowman and Littlefield.

Pritchard, D. 2005. *Epistemic Luck*. Oxford University Press.

Rysiew, P. 2012. 'Epistemic Scorekeeping'. In Brown & Gerken (eds.) *Knowledge Ascriptions*. Oxford University Press.

Schaffer, J. 2004. 'From Contextualism to Contrastivism'. *Philosophical Studies*, Vol. 119, pp.73-103.

Schecter, J. 2013. 'Rational self-doubt and the failure of closure'. *Philosophical Studies*, Vol. 163, pp. 429-452.

Sharon, A. & Spectre, L. 2008. 'Mr Magoo's Mistake'. *Philosophical Studies*, Vol. 139, No. 2, pp. 289-206.

——— 2010. 'Dogmatism repuzzled'. *Philosophical Studies*, Vol. 148, pp. 307-321.

Sosa, E. 1999. 'How to Defeat Opposition to Moore'. Philosophical Perspectives, 13, pp. 141-153.

——— 2000. 'Skepticism and Contextualism'. *Philosophical Issues*, Vol. 10, pp. 1-18.

——— 2007. *A Virtue Epistemology: Apt Belief and Reflective Knowledge*. Oxford University Press.

Srinivasan, A. 2013. 'Are We Luminous?'. *Philosophy and Phenomenological Research*, Online first.

Stine, G. C. 1976. 'Skepticism, Relevant Alternatives, and Deductive Closure'. *Philosophical Studies*, Vol. 29, pp. 249-61.

Sutton, J. 2005. 'Stick to What You Know'. *Noûs*, Vol. 39, No. 3, pp. 359-396.

Vogel, J. 1990. 'Are there Counterexamples to the Closure Principle?'. In M. Roth & G. Ross (eds.), *Doubting: Contemporary Perspectives on Skepticism*. Dordrecht: Kluwer.

——— 2010. 'Luminosity and Indiscriminability'. *Philosophical Perspectives*, 24, 1, pp. 547–572.

Weatherson, B. 2004. 'Luminous Margins'. *Australasian Journal of Philosophy*, Vol. 82, pp. 373–383.

Wedgwood, R. 2002. 'The Aim of Belief'. *Noûs*, 36, pp. 267–297

Williamson, T. 1992. 'Inexact Knowledge'. *Mind*, Vol. 101, No. 402, pp. 217-242.

——— 2000. *Knowledge and Its Limits*. Oxford University Press.

——— 2005. 'Knowledge, Context, and the Agent's Point of View'. In Preyer & Peter (eds.) *Contextualism in Philosophy*. Oxford University Press.

——— 2009a. 'Probability and Danger'. *The Amherst Lecture in Philosophy 4*, pp. 1-35. <http://www.amherstlecture.org/williamson2009/>.

——— 2009b. 'Reply to John Hawthorne and Maria Lasonen-Aarnio'. In Greenough & Pritchard's (eds.), *Williamson on Knowledge*. Oxford University Press.

——— 2011. 'Improbable Knowing'. In Dougherty (ed.), *Evidentialism and its Discontents*. Oxford University Press.

——— forthcoming. 'Very Improbable Knowing'. *Erkenntnis*. Available on author's webpage (Last seen: 08-09-2014): http://www.philosophy.ox.ac.uk/__data/assets/pdf_file/0015/19302/veryimprobable.pdf