"Internal Justification and Anti-Luck Epistemology: A Compatibilist Proposal"

A thesis submitted in requirement of the Degree of Doctor of Philosophy

Charles Anthony Neil

Philosophy Department
University College London

January 2020

Abstract

The thesis examines a version of awareness internalism about epistemic justification according to which justified belief requires reliably formed belief in conjunction with second-order awareness. It supports this version of internalism by appealing to a conception of knowledge as a variety of non-lucky true belief; different types of epistemic luck are delineated and a modal safety-based response to epistemic luck is adopted. An epistemically justified belief, according to this version of awareness internalism, is more than a blameless belief; reliabilist insights about the nature of justification are used to explain this verdict. The thesis identifies two sceptical arguments against using anti-luck considerations to motivate awareness internalism but finds the arguments surmountable. The resultant view is a form of awareness internalism motivated by anti-luck considerations, which incorporates internal and external conditions for justification to respond to sceptical objections and explain a broad class of epistemic intuitions. Thus, it will illustrate how both sides of the contemporary internalist and externalist debate are partially correct.

Nevertheless, the anti-sceptical import of awareness internalism is qualified. By considering whether awareness internalism can respond to regress scepticism, the thesis identifies a meta-epistemological anxiety about whether we do, *in fact*, have justified beliefs; awareness internalism does not ameliorate this concern. Similarly, the thesis identifies counterexamples to the awareness condition. The thesis contends that a non-dogmatic response to counterexamples, one which accommodates the broadest range of intuitions, allows that knowledge does not always require a belief to satisfy the awareness condition for justification. It is contended that counterexamples to awareness internalism may be regarded as instances of normatively impoverished knowledge, but knowledge, nonetheless.

Impact Statement

The thesis advances the hypothesis that a construal of awareness internalism and its relationship to knowledge as a species of non-lucky true belief represents a compelling and novel conception of justification. This is placed in the context of a detailed critical review of the relevant literature. The hypothesis developed is intended to make a significant contribution to the understanding and advancement of this area of modal epistemology and part of the thesis has already been published as a research paper (with 1,200 downloads as of December 2019). A further two papers will shortly be submitted for publication (based on chapters 2 and 3).

The thesis may also have an educational value in its presentation of philosophical arguments and use of the current literature; in particular, ideas and examples within the thesis may inform research-based teaching curricula for students studying epistemology. It provides examples, both hypothetical and empirical, that are intended to elicit pre-theoretical intuitions, which do not presuppose specialist academic knowledge. Inclusion of these examples into an undergraduate syllabus could facilitate group discussion, academic debate and a collaborative approach to learning.

Acknowledgements

Table of Contents

Chapter 1

Introduction and Hypothesis

The thesis asks the following question:

> **Q**. Why, and how, might one endorse an internalist view of justification which also holds that a belief's justification depends, partly, on externalist conditions? Could such a view accommodate objections?

I will start with an informal statement of internalism and externalism about justification. Let internalism denote the view that a belief's justification depends on factors necessarily internal to the believer. By contrast, let externalism denote the view that a belief's justification does not depend on factors that are necessarily internal to the believer.

This chapter examines the research question (**Q**) and explains the research hypothesis.

However, first I will offer a brief characterisation of the position advanced in the thesis and how the thesis will proceed. The thesis defends a version of internalism called "awareness internalism", and a specific version of awareness internalism. The specific version of awareness internalism holds that a belief's justification depends on "externalist" facts (such as reliability) and the believer's (internal) awareness that such facts hold for one's belief. On this view, justified belief is more than a merely blameless belief, and reliabilist insights are used to explain how. The thesis appeals to anti-luck considerations to motivate awareness internalism. The resultant notion of awareness internalism is one consistent with knowledge as a species of non-lucky true belief.

The thesis identifies two sceptical arguments against using anti-luck intuitions to motivate awareness internalism. Chapter 2 considers the notion that internalist anti-luck conditions are unsatisfiable and commit us to scepticism. Chapter 3 considers the contention that an anti-luck motivation for awareness internalism can only be preserved by endorsing an awareness requirement that results in scepticism. It is argued that we can, respectively, accommodate both objections by re-evaluating the connection between second-order awareness and epistemic luck, and by making justification dependent on a reliability condition.

However, the thesis identifies counterexamples to awareness internalism. The counterexamples are cases where either awareness does not seem necessary for justified belief (in Chapter 2, §4 & Chapter 6), or where reliability does not seem necessary for justification (in Chapter 4). The thesis seeks a non-dogmatic response to counterexamples. To this end, I examine and support a response to counterexamples which maintains that knowledge, for a class of cases, does not depend on second-order awareness. The resultant view is that awareness is consistent with a qualified form of externalism about knowledge.

The account may be characterised as "compatibilist" in the following respects. First, at the level of intuitions, I draw on a range of epistemic intuitions, explaining how an account of awareness internalism that incorporates externalist condition allows us to meet a broader class of epistemic intuition. More specifically, it allows us to understand how epistemic intuitions that may otherwise be regarded as supporting opposing conceptions of justification are consistent with awareness internalism. Second, at the theoretical level, the thesis appeals to anti-luck considerations to support (awareness) internalism; rather than anti-luck intuitions posing a problem for internalism, I argue that internalist anti-luck intuitions are compatible with a sound motivation for internalism, providing sceptical objections to awareness internalism are surmountable.

I will devote this chapter to an examination of the research question, hypothesis and relevant literature. As indicated, Chapters 2 and 3 discuss sceptical objections to awareness internalism. Chapter 4 considers thought experiments purporting to show that reliability is not a necessary condition for justified belief. Chapter 5 provides a more detailed account of anti-luck conditions. Ch. 6 considers a further objection to awareness internalism. Finally, I summarise which conclusions may reasonably be drawn (see §4 of this chapter for an exact overview of structure).

This introductory chapter is structured as follows:

§ 1. I isolate the senses of "justification" and "knowledge" which concern the research question and offer a definition of internalism and externalism (about both justification and knowledge). I consider motivations for the respective views.

§ 2. I outline the research hypothesis.

§3.I define the proposed internal condition for justification and distinguish the condition from other internal and external conditions.

§ 4. I explain how the thesis proceeds in the remaining chapters.

§ 1. Terminology

I wish to start by considering the terms "knowledge" and "justification". The thesis question concerns knowledge and justification of empirical propositions, propositions of the form "the book is red" or "Bob is justified in believing that his headache is caused by dehydration". In the case of empirical propositions, I shall assume that it is logically possible that the propositions are false. For example, it is possible that the book is white but illuminated by red light. In the broadest sense, we are inclined to think of the propositions under consideration as contingently true. That is, they could be false.

The research question concerns the belief or knowledge a subject has in relation to a proposition. If one believes that a given proposition is true, then we may inquire about the status of that belief; to ask, for example, whether that belief is justified or whether that belief constitutes knowledge. For instance, a subject may believe but also know that the book is red. Such knowledge is propositional knowledge. Of course, not all knowledge attributions take this linguistic form. For example, one may know how to ride a bicycle – this is often termed "procedural knowledge". The thesis question will not be concerned with procedural knowledge or knowledge-how. Instead, the question of the thesis concerns knowledge of empirical propositions.

There are, also, multiple senses of justification. For example, for a subject "S" and a proposition "p", it may seem doubtful that the attribution "S is justified in believing that p" is equivalent to the attribution "S's belief that p is justified". Mylan Engel (1992b) regards the former attribution as an attribution of *personal justification*, and the latter an attribution of *doxastic justification* (p.136). Engel observes that the referents in both attributions have different domains of evaluation. Specifically, in the first attribution [S is justified in believing that p] the domain of evaluation is the subject S, and in the latter attribution the domain of evaluation is the belief (p. 135).[1] If, as Engel contends, a *belief* may be justified independently of whether the *believer* is justified in holding that belief, then, as Engel observes, these two attributions will not be extensionally equivalent.[2]

Consider also a further type of justification, namely "p is justified for S". This attribution may be true in circumstances in which S has not assented to the belief that p. For example, p may be justified for S because *believing* p would be consistent with

---

[1] The claim personal and doxastic justification are equivalent notions is termed the "equivalency thesis" (Engel, 1992b, p. 135).

[2] Ibid., p. 136.

S's evidence, even if S has not yet assented to the belief that p.[3] Let us call this *propositional justification.* It is prima facie plausible that propositional and doxastic justification are inequivalent. For example, p may be justified for S if there is evidence for p of which S is unaware, such that if S were to believe p then S's belief that p would be consistent with evidence for p: if a justified doxastic attitude is one consistent with one's evidence (Conee and Feldman, 1985), then S would have a justified belief that p.

The research question concerns doxastic justification rather than personal or propositional justification (where doxastic justification is understood in Engel's way - as concerning the justification of beliefs).[4] One reason for this focus is that if knowledge requires belief, then there arises the question of whether one's *beliefs* need to be justified to constitute knowledge.

Furthermore, the thesis question is concerned with the *normative* conception of justification. According to the normative conception, when a belief is justified, then that is because it meets some standard for justification, such that it merits approval, and it is a belief one *ought* to hold (Audi, 2001.) Analogously, an action may be morally justified or unjustified to the extent that one ought (or ought not to) perform that action. Similarly, a belief may be justified or unjustified to the extent that one ought (or ought not to) to hold that belief. Beliefs which are justified in the normative sense are beliefs which we *ought* to hold. Conversely, unjustified beliefs are ones we *ought not* to hold.[5][6] The standard for justification may be construed as the *norm* for belief, such that one's belief is justified to the extent that it satisfies the norm for belief.

Holding a justified belief (in the normative sense, i.e. connoting a belief one *ought* to hold) amounts, I suspect, to more than having a merely *blameless* belief. Epistemologists have often argued that being blameworthy is enough to defeat one's justification - even if one's belief is true and is formed using a reliable method of belief formation.[7] However, despite the possibility that blamelessness is necessary for

[3] Conee and Feldman (1985) affirm that a doxastic attitude D is justified for S just in case D fits S's evidence (p.15). Such a notion may allow a belief B to be justified for S in circumstances where S has not yet assented to B.

[4] Another use of "doxastic justification" is for a belief *based* on justification. In this thesis, I adopt Engel's use of the term, that is, as a statement concerning whether a belief is justified.

[5] See also Kim, (1988, p. 383) for a similar characterisation.

[6] We may use terms other than "justification" to express the normative conception of justification: beliefs may be rational, reasonable, or warranted; and conversely, irrational, unreasonable, or unwarranted. Furthermore, as Alston (2005) notes, there are uses of "justification" which depart, etymologically, from reasonable belief. For instance, a technical use of justification is *as the difference between knowledge and true belief* (what Plantinga (1993a) calls warrant). The difference between true belief and knowledge is not what I intend by the normative conception of justification. I want to allow that *other* non-justificatory conditions may be required for knowledge in addition to truth and belief.

[7] See Goldman's (1988) distinction between strong and weak distinction. Goldman characterizes weak justification as merely blameless believing, and strong justification as blameless but reliably formed belief. Chapter 4 will examine Goldman's distinction.

justification, having a blameless belief does not seem sufficient for justification. For example, imagine the following case: FLAT-EARTHER: a flat-earther believes the earth is flat and is unaware of any evidence to the contrary because she lives in a community unaware of modern science. However, she is not a conspiracy theorist. Instead, she is told by all members of her community, whom she trusts, that the world is flat. Her belief seems blameless, but not epistemically justified, because it is not – strictly speaking – a belief which she ought to hold.

The thesis question concerns epistemic justification rather than prudential justification. I shall assume that it may be prudent to have beliefs which are not epistemically justified. For example, an AMBITIOUS POLITICIAN might be prudentially justified in believing that she will win the leadership contest. Imagine her belief instils her with the confidence to run a more effective campaign. Her belief seems a prudent belief to hold. Nevertheless, her belief is not, necessarily, epistemically justified if she has little chance of winning. Her belief does not seem reasonable or justified in the normative sense.

Finally, different kinds of beliefs can be distinguished. Firstly, one may be consciously considering a proposition "p" one believes. In which case, one's belief that p will be "occurrent". However, many of our beliefs are not, presumably, before our consciousness at every point in time – human beings can only consciously attend to a limited number of propositions at any given moment in time. However, intuitively, we hold many beliefs about propositions even when we are not consciously considering the propositions. The total number of beliefs we have, at any given time, plausibly outstrips the number of beliefs we can have before our consciousness at any given moment. It is doubtful, or at least not intuitive, that we should lose our original beliefs in the course of consigning them to memory.

The beliefs stored in memory may be characterised as "non-occurrent". The distinction between occurrent/non-occurrent beliefs will be returned to in §3. Importantly, many non-occurrent beliefs are intuitively justified or cases of knowledge – for otherwise it would transpire that we justifiably believe or know much less than we ordinarily think we do.

In summary, the research question is restricted to the following: propositional knowledge rather than procedural knowledge; epistemic justification rather than prudential justification; the justification of beliefs, rather than the justification of persons.

§ 1.1. Internalism and Externalism

In this section, I characterise internalism/externalism about justification. I then characterise internalism/externalism about knowledge. (§3 defines a specific internal awareness condition for justification).

There is no univocal definition of internalism or externalism (either about justification or knowledge). Different definitions often yield opposing verdicts for which propositions one can be said to know or justifiably believe. Similarly, there is not some single motivation for either an internalist or externalist account of justification; different accounts have arisen in response to a diverse range of epistemological problems. [8] The first task, therefore, is to stipulate internalism/externalism in a way which clarifies the condition for justification supported in the thesis. Doing so will provide a foundation for the more specific condition in §3.

First, there is a stipulation of internalism and externalism which holds that (doxastic) justification depends solely on states internal to the believer, such as phenomenal or mental states. This may be expressed by the claim that justification *supervenes* on the phenomenal states of a believer (Audi, 2001) or on a believer's mental states (Conee and Feldman, 2001). If internalism is defined in this way, then an account would be "externalist" if it made justification depend, even partially, on facts other than internal states, such as facts about the reliability of one's belief-forming procedures. [9]

However, stipulating internalism and externalism in the above way may obscure a possibility: namely, that one's *internal* awareness of paradigmatically "externalist" facts (e.g. facts about the reliability of the belief or the normality of the environment in which one forms the belief) may be relevant to a belief's justification. For instance, an account of justification which makes reliability a *necessary* condition for justified belief, but which further holds that one must be *aware* that one's belief is reliably formed, would not satisfy the criteria for internalism, if internalism maintains that justification is *exclusively* determined by one's mental or phenomenal states. [10]

---

[8] Conee and Feldman (2001) observe that any definition is stipulative. The stipulation they offer differs from the stipulation I offer here. Chapter 1 (§ 3) considers their stipulation.

[9] Distinguishing internalism/externalism in this way, as Conee and Feldman (2001,p.3), observe is akin to how the internal/external distinction has been used in the Philosophy of Mind. For example, an externalist view of belief content permits factors external to one's mind to determine the content of one's beliefs.

[10] See Conee and Feldman (2001). They observe that "the epistemic internalist is principally opposed to the existence of any justification determining role for plainly external factors such as the general accuracy of the mechanism that produces a given belief or the belief's environment" (p.3), concluding that "mentalism bears this out". By contrast, the version of internalism I support makes external factors necessary for justified belief. As Conee and Feldman observe, any definition of internalism is "to some extent stipulative" (p2). I am opting

Nevertheless, there is still a difference between accounts of justification which hold that a believer's *awareness* of externalist factors (such as the reliability of one's belief-forming method) is necessary and accounts of justification (such as Goldman's, 1979) which do not require awareness of externalist factors. That difference may have important implications for which beliefs are justified. I therefore seek a stipulation of internalism and externalism to explain that contrast.

Let us therefore stipulate that a theory is "externalist" if it holds that the justification of subject S's belief B depends on conditions which may obtain irrespective of whether S is aware that they obtain. For example, the fact that B is formed using a reliable method of belief-formation could, on an externalist view, be necessary and sufficient for B to count as justified (Goldman, 1979). Furthermore, one may in principle use a reliable method of belief formation, even if one is unaware that one's method is reliable. We can imagine, for instance, reflectively unsophisticated believers forming reliably true beliefs even though they are unaware that their beliefs are reliably formed.[11]

The version of awareness internalism supported in this thesis maintains that factors such as reliability, or facts about one's environment, are at least *necessary* justification. The version of awareness internalism I will support is consistent with the following principle:

Awareness internalism (about justification): Subject S has an epistemically justified belief B only if B has some property Φ and S is aware that B has Φ.[12]

The above principle imposes a *second-order awareness* requirement on justified belief (specifically, that S is aware that B has Φ). Awareness may come in different forms. A second-order belief (specifically, a belief that B has Φ) is one mode of second-order awareness.

Notably, awareness internalism is currently neutral on what "Φ" denotes: Φ may be the condition that one's method of belief formation is reliable, or that one has evidence for a belief, or that one's belief is formed in a normal environment, or indeed

for a different stipulation, since mentalism would not readily allow me to contrast between externalism and the version of awareness internalism under consideration.

[11] This formulation allows non-doxastic modes of apprehension to constitute awareness. Chapter 3 will discuss a non-doxastic mode of awareness.

[12] This is similar to BonJour (1985, p. 31), see also Bergmann (2006. p. 15), though both assert a bi-conditional. I wish to avoid a bi-conditional to allow that normative defeaters may defeat one's justification (Chapter 3).

a conjunction or disjunction of conditions. However, according to the version of awareness internalism I will support, doxastic justification requires that one is further aware that one's belief is reliable (§3 provides a more exact presentation).

By contrast, to yield an externalist condition, the second-order awareness requirement may be abandoned. The following principles expresses an externalist conception of epistemic justification:

> **Externalism (about justification)**: Subject S has an epistemically justified belief B just in case B has Φ (for a Φ that may hold for B in circumstances where S is unaware that B has Φ).

Externalism is now the negation of awareness internalism. Specifically, externalism negates the second-order awareness requirement for justified belief. An example of an externalist account is Goldman's (1979) reliability theory. Goldman holds that one's belief is justified just in case it results from a reliable cognitive belief-forming procedure. On Goldman's account, one needn't be aware that one's belief results from a reliable procedure for that belief to be justified.

The internalist view, it seems, will impose reflective demands on having a justified belief in some proposition p, over and above believing that p. Indeed, depending on the character of the second-order awareness, internalism predicts that reflectively unsophisticated subjects, such as animals, which lack the requisite second-order awareness, may fail to have internally justified beliefs. Conversely, externalism is consistent with reflectively unsophisticated subjects holding justified beliefs; because externalism abandons the second-order awareness requirement for justified belief.

Internalism and externalism about justification can be distinguished from internalism and externalism about knowledge (the latter I will term *epistemic internalism* and *epistemic externalism*). Epistemic internalism maintains that one's belief constitutes knowledge only if it is internally justified.[13] Conversely, epistemic externalism negates the requirement that knowledge depends on an internally justified belief; epistemic externalism claims that knowledge does not necessarily require internal justification. More schematically:

---

[13] Let "internally justified" be shorthand for having a belief which meets the internalist standard justification". Similarly, when I characterise a subject as having internal justification" for a proposition, assume I am referring to the subject having a belief which meets the internalist standard for justified belief.

Epistemic internalism: S's knowledge that p depends on an internally justified belief that p.[14]

Epistemic externalism: S's knowledge that p does not depend on an internally justified belief that p.

Epistemic externalism, however, does not entail that knowledge depends on a belief justified by externalist standards for justification. For instance, epistemic externalism is, in principle, consistent with internalism about justification: this might happen if knowledge depends on externalist conditions (such as reliability) and that justification depends on internalist conditions such as second-order awareness, but that knowledge does not depend on justification (e.g. Audi, 2001). Epistemic externalism is committed to justificatory externalism only to the extent that knowledge depends on justification. Similarly, epistemic internalism does not entail that having an *internally justified true belief* is sufficient for knowledge. Epistemic internalism is, in principle, compatible with the possibility that further, non-justificatory conditions are required for propositional knowledge, in addition to truly believing a proposition (§3 discusses candidate conditions).

§ 1.2. The Justificatory Condition for Knowledge

The current definition of epistemic externalism is neutral on whether knowledge requires justification. I will therefore consider how a justification condition for knowledge bears upon our research question.

Let us represent the claim that knowledge requires justification as follows:

$[K \rightarrow J]$: S knows that p only if S has a justified belief that p.

Let us understand the above principle, $[K \rightarrow J]$, as a universal claim about *all* kinds of proposition knowledge. As indicated at the beginning of this chapter, the thesis ultimately takes a different view. However, let us imagine - for argument's sake - that $[K \rightarrow J]$ is true.

---

[14] Note, this does not imply that an internally justified true belief is sufficient for knowledge. §3 considers non-justificatory conditions, other than truth and belief.

If $[K \rightarrow J]$ is true *and* if an internalist account of justification is defensible, then epistemic externalism is compromised. For both epistemic externalism and the $[K \rightarrow J]$ principle are correct only to the extent that an externalist account of justification is correct. It seems, therefore, that only if $[K \rightarrow J]$ is rejected could epistemic externalism be endorsed alongside justificatory internalism. A difficulty with rejecting $[K \rightarrow J]$, however, is that $[K \rightarrow J]$ is a prima facie attractive principle. One reason for finding $[K \rightarrow J]$ intuitive, I suggest, is that justification is a normative notion concerning what one ought to believe. A worry is that it would seem revisionist of our intuitions to claim that there are propositions which we *know* but which we are unjustified in believing, and which we *ought not to believe*. Having an unjustified belief seems, on first inspection, a way of lacking knowledge.

§ 1.3. Epistemic Externalism and Justificatory Internalism: initial motivations

The research hypothesis (in § 2) indicates the view supported in the thesis: namely, that justification requires second-order awareness of externalist states of affairs (specifically, the fact that one's belief is reliably formed), and that anti-luck considerations motivate awareness internalism. Before unpacking the research hypothesis, I will outline some initial motivations for externalism and for awareness internalism.

Take, first, the case of human beings. We believe, and take ourselves to know, many empirical propositions. Moreover, we can reflect on the credentials of our beliefs about propositions; for example, we may believe that our beliefs are true, justified, or formed via a reliable method of belief formation. We may scrutinize beliefs, revise them, use them in explanations and use them to plan for the future.

However, many of our empirical beliefs are not formed on the basis of consciously attending to propositions. We formed beliefs automatically, often, on the basis of sensory perception. If an awareness, requirement for justification were to demand that of every belief, we must be consciously aware of what makes it justified, then it may seem to be imposing on justification an unduly demanding condition for justified belief. By contrast, providing that our beliefs are in fact reliably formed, then those beliefs may be understood as epistemically justified according to an externalist standard. Externalism allows that a wide class of intuitively justified beliefs satisfy externalist requirements such as Goldman's. By contrast, if awareness internalism imposes stringent demands for each justified belief token, it will be doubtful that a wide

class of intuitively justified beliefs will satisfy internalist requirements. The resultant intuition will be that an awareness requirement is not necessary for justified belief.

A further motivation for externalism is supplied by the fact that justification, as the thesis will indicate, plausibly requires more than blamelessly believing a proposition. Chapters 3 and 4 will further motivate this view. Moreover, for a wide class of cases, we can diagnose the failure to have a justified belief in terms of a failure to have a reliably formed belief. Externalist accounts such as Goldman's (1988) amenably explain this verdict, by making blamelessness a *necessary* condition for justification while also requiring that justified beliefs are reliably formed. By contrast, awareness internalism, at least when it is not supplemented with a reliability condition, may not straightforwardly explain this verdict.

A further motivation for externalism is its response scepticism (Armstrong, 1973). In particular, depending on the nature of the awareness involved, awareness internalism risks a sceptical regress (BonJour, 1985, Bergmann) For instance, if the mode of awareness is doxastic (requiring that one has a second-order belief that one's first-order belief is justified), and if that second-order belief in turn has to be justified in order for one's first-order belief to be justified, then a regress threatens. In particular, the second-order belief will only be justified by an internalist standard if one *further justifiably believes* that one's second-order belief – that third-order belief will also have to be justified, and so on, ad infinitum. Chapter 3 will provide an exact presentation of the regress problem for awareness internalism. However, initially, we can sense how that regress need not arise on an externalist treatment of justification; because externalist theories such as Goldman's require only that one's first-order belief is reliably formed (as opposed to requiring a further second-order belief which needs to be justified by internalist standards). Chapter 3 considers the regress problem and how it arises as a challenge to the version of awareness internalism outlined in §3.

A further consideration concerns reflectively unsophisticated subjects, for example, animals, infants, small children; these may lack the requisite second-order awareness to be internally justified. [15] Such subjects may respond in sophisticated ways to external stimuli while having, at most, only a rudimentary awareness of their beliefs. If non-sophisticated subjects can truly be said to know, then these cases pose an explanatory challenge for internalism: how to explain an intuition that they *know* if they do not satisfy the internalist (awareness) condition for justification? Chapter 6 considers reflectively unsophisticated subjects as a potential counterexample to awareness internalism. Notice, however, than externalist account of justification can

---

[15] Chapter 6 examines potential objections to this view.

explain how reflectively unsophisticated subjects *may* possess knowledge. This is because if it is appropriate to attribute beliefs to reflectively unsophisticated subjects, and if it is further appropriate to attribute knowledge, then externalism allows reflectively unsophisticated creatures to hold justified beliefs. For instance, providing that their belief-forming procedures *are* reliable in Goldman's sense, then they may have externally justified beliefs. An externalist treatment of justification offers a way of explaining an intuition that they know: because by satisfying necessary and sufficient conditions for justified belief, they satisfy an externalist justificatory condition for knowledge.

Consider, now, why one might find awareness internalism about justification attractive. Recall that awareness internalism rejects the *sufficiency* of facts such as reliability for justified belief, requiring that one must furthermore be *aware* of the factors which contribute to a belief's justification. Laurence BonJour (1985) provides a counterexample to externalist accounts of justification.[16]

> NORMAN: "Norman, under certain conditions which usually obtain, is a completely reliable clairvoyant with respect to certain kinds of subject matter. He possesses no evidence or reasons of any kind for or against the general possibility of such a cognitive power or for or against the thesis that he possesses it. One day Norman comes to believe that the president is in New York City, though he has no evidence either for or against this belief. In fact the belief is true and results from his clairvoyant power under circumstances in which it is completely reliable." (BonJour, 1985, Structure of Empirical Knowledge, p.41).

BonJour does not regard Norman as holding a justified belief. Nevertheless, BonJour imagines that NORMAN satisfies necessary and sufficient conditions for having an externally justified belief. The case of NORMAN elicits the intuition that externalist conditions, such as reliability, are *insufficient* for justified belief. An analogous case that will be examined in Chapter 2 is the example of chicken-sexers, who can reliably distinguish between the sex of hatchling chickens by inspecting them; their method is olfactory. Chapter 2 will examine the case of an unenlightened chicken-sexer (Pritchard, 2005), who, we are to imagine, is unaware that her method of belief formation is olfactory. I will argue that such cases highlight how externalism is vulnerable to a harmful species of epistemic, and that this insight may be used to

---

[16] BonJour further regards Norman as knowledge failure. In later work, (1987) pp. 297-314, BonJour examines a similar case, directing it at externalist treatments of knowledge.

support awareness internalism. There seems, at least, a reflective failing on behalf of Norman and the unenlightened chicken-sexer, which may explain the intuition they neither know nor justifiably believe.[17]

§ 2. Research Hypothesis

At this point, I have provided preliminary definitions of internalism and externalism about justification and knowledge. I have considered some general motivations for the respective views and the explanatory challenges facing each view. Now, to reconsider the research question:

> **"**Why, and how, might one endorse an internalist view of justification which also holds that a belief's justification depends, partly, on externalist conditions? Could such a view accommodate objections?

Awareness internalism is contentious. Goldman, for example, took his externalist view to explicate "ordinary standards" for justification.[18] However, I shall argue that anti-luck epistemology motivates awareness internalism and that without awareness, our knowledge is vulnerable to a harmful species of luck, which either undermines or degrades knowledge.[19]

Anti-luck epistemology is concerned, first, with the intuition that some species of epistemic luck are incompatible with knowledge, and second, with the postulation of necessary conditions for knowledge which exclude epistemically lucky beliefs from the extension of knowledge. The familiar way in which a true belief may be epistemically lucky is illustrated by an adaption of Bertrand Russell's (1948) stopped-clock example. Suppose a believer truly believes that the time is 12.00, but unbeknownst to him, the clock stopped exactly 24 hours beforehand. In this case, it seems intuitive that his belief is luckily true in a way which excludes or degrades knowledge. Similarly, I will suggest that cases such as NORMAN and the unenlightened chicken-sexer illustrate an important intuition about epistemic luck. Minimally, it seems at least a matter of luck from *Norman's perspective* that his belief is true (Bergmann, 2006; BonJour, 1985). The thesis will appeal to our anti-luck intuitions to motivate awareness internalism (about justification) as a response to cases such as NORMAN.

---

[17] See also Fumerton, 1995, (p. 116).
[18] Goldman, 1979, p. 1
[19] *Degraded* knowledge allows for the possibility that knowledge may survive epistemic luck, but that the kind of knowledge which is epistemically lucky is not of the best kind (see Chapter 2, §4), in the way that Hetherington (2001) argues.

The **research hypothesis** is that a construal of awareness internalism and its relationship to knowledge as a species of non-lucky true belief represents a persuasive conception of justification. An epistemically justified belief, according to this version of awareness internalism, is more than a blameless belief; reliabilist insights about the nature of justification are used to explain this verdict. The thesis identifies two sceptical arguments against using anti-luck considerations to motivate awareness internalism but finds the arguments surmountable. The resultant view is a form of awareness internalism motivated by anti-luck considerations, which incorporates internal and external conditions for justification to respond to sceptical objections and explain a broad class of epistemic intuitions.

The account appeals to the necessity of external conditions for justification but argues that a further (internal) awareness requirement is needed for justified belief. Counterexamples will be met by adopting a pluralistic approach to defining knowledge, according to which there is more than one type of knowledge. Following Hetherington (2001), minimal knowledge will be distinguished from non-minimal knowledge. I will explain how awareness internalism is consistent with a qualified form of externalism, one which holds that *certain kinds of knowledge* do not require justification.

The first potential value of the hypothesis is that, if defensible, it will allow us to retain some paradigmatically externalist claims about the *necessary* conditions required for justified belief, while also meeting an internalist demand for second-order awareness. Thus, it will capture a sense in which both sides of the contemporary internalist and externalist debate are partially correct. By contrast, there are internalist views of justification which make externalist conditions such as reliability unnecessary for justified belief (Audi, 2001).[20] The account I support departs from accounts such as BonJour's (1985) because it does not equate justification with the fulfilment of epistemic duty. Furthermore, the account makes justification require more than a merely blameless belief in a proposition. Crucially, the claim that justified belief amounts to more than merely blameless believing has been captured by externalist accounts (e.g. Goldman 1988), and it will be revealing to see how this view of justification can be reconciled with an account of justification that makes awareness necessary for justified belief.

Secondly, while there are many defences of internalism about justification in the literature, they have not been applied in a thoroughgoing way to the version of awareness internalism supported in this thesis. When versions of awareness internalism have appealed to anti-luck considerations in motivating internalism (e.g.

[20] Chapter 4 considers this in detail.

BonJour, 1985), they do so only implicitly – or at least, not explicitly enough, and the appeal to anti-luck considerations has been taken as problematic for the view that second-order awareness is necessary for justification (Pritchard, 2005).[21] For example, it has been claimed awareness internalism – at least when motivated by anti-luck considerations – has implausible sceptical implications which are avoidable only by abandoning the anti-luck motivation for internalism (Bergmann, 2006).[22] I hope the version of awareness internalism, described in the next section, will allow us to meet these objections.

§ 3. Awareness Internalism, Types of Internalism, and Epistemic Externalism

I will now specify the awareness condition for justification supported in subsequent chapters:

> **Awareness internalism (+externalist condition):** Subject S has a justified belief B only if 1) B has some property Φ AND S is aware that B has Φ (for a Φ that may hold for B in circumstances where S is unaware that B has Φ).

I will assume that condition Φ, in the right-hand side of the material conditional, will need to be specified in non-epistemic terms if the principle is to yield a non-circular account of necessary conditions for justified belief (see Goldman, 1979, pp. 1-2, on this point). For instance, specifying that Φ holds for whenever S has a *good reason to believe*, or is *reasonable in believing*, is vulnerable to circularity. This is because *being reasonable in believing* is essentially a normative-epistemic notion. Specifically, beliefs that are reasonable are beliefs one ought to hold, which is precisely the normative conception of justification we are trying to explicate on the left-hand side of the conditional. Specifying Φ in normative-epistemic terms would not necessarily allow the right and left-hand side of the conditional to be distinct. Awareness internalism is neutral on whether Φ is an externalist condition or an internalist condition, such as a

---

[21] Duncan Pritchard (2005). Chapter 2 discusses Pritchard's argument.
[22] Michael Bergmann (2006). Chapter 3 examines Bergmann's argument.

mental or phenomenal state. ₂₃ However, Chapter 3 will support the following condition:

Internal (justification) condition: S has a justified belief B *only if* S uses a reliable method of belief formation in arriving at B and S believes she has used a reliable method of belief formation in arriving at B.

I wish to make the following observations about the internal condition.

Firstly, let us assume that relevant methods are broad methods such as testimony and deduction. A broad method may be applied in different instances, and in counterfactual circumstances (Chapter 5 provides a more exact presentation of methods).

Secondly, the internal condition makes awareness of reliability a *necessary* condition for justified belief rather than a sufficient condition for justified belief. The condition does not imply that [reliability & awareness] is sufficient for justification. To support [reliability & awareness] as sufficient for justification, one would need to demonstrate that there exist no cases in which [reliability & awareness] is satisfied but where one is justified. Doing so would be difficult, and the claim about sufficiency is – I suspect - stronger than that needed to explain the cases of belief-formation in this thesis.

Thirdly, doxastic awareness requirements need not require occurrent second-order beliefs. If having a second-order belief is a way of having second-order awareness, then recall from §1.1 that beliefs may be non-occurrent (i.e. not being consciously held at any given time) but may instead reside in memory; such beliefs are accessible but are not being accessed (see Audi, 2001). Nevertheless one is plausibly aware of the propositional content of non-occurrent beliefs, insofar as one could recall the non-occurrent beliefs from memory. This fact is not merely a coincidence as the total number of propositions one is aware of (at any given point in time) outstrips the number of propositions before our consciousness at any given point in time. By parity of reason, one's second-order awareness need not be restricted to occurrent second-order beliefs. For example, of the many first-order beliefs we are non-occurrently aware of, we may also non-occurrently believe that our first-order beliefs are true, reliably formed, justified and so on. We are not, after all, always consciously considering the justificatory grounds for our beliefs. Thus, if *non-occurrently believing* is a way of being

---

₂₃ Notwithstanding the possibility that mental states are understood in externalist terms, where one can be in a mental state M without being aware that one is in M (e.g. Williamson, 2000).

aware, then one may have *a non-occurrent* (second-order) belief that one's *occurrent* (first-order) belief that p is reliable, and *a non-occurrent* (second-order) belief that one's *non-occurrent* (first-order) belief is reliable.

Furthermore, we needn't assume that for every justified first-order belief B, one must have, at some point or other, occurrently believed that B is a belief token resulting from a reliable method of belief formation. We form true perceptual beliefs which may, intuitively, be justified even though we have not – at some stage or other – formed a second-order belief about the belief. However, when asked why we believe that p, we may cite that we perceived it or heard it and that we believe that our sensory faculties are functioning normally and that we used a familiar method of belief formation, one which generally yields true beliefs under normal conditions. In such cases, we may be disposed to occurrently believe that our belief token is reliably formed. If non-occurrently believing is a way of being aware and if one can be aware because one is disposed to occurrently believe, then the internal condition for justified belief may not seem especially demanding. However, it may still exclude – from the extension of justification - those believers who have never had occurrent second-order beliefs about the reliability of their belief-forming methods and who are not disposed to having such second-order beliefs.

Fourthly, the condition does not demand that second-order beliefs must also be epistemically justified. Chapter 3 will explain how a requirement that one's second-order beliefs are epistemically justified generates a skeptical regress. The chapter will propose a non-epistemic awareness principle, which merely requires a blameless second-order belief. The chapter how the sceptical challenge can be surmounted by endorsing a non-epistemic principle.

Fifthly, the awareness condition allows for justified false beliefs. This is because a method of belief formation M may be a reliable type of method if it produces a high proportion of true belief tokens. The proportion of true beliefs M yields need not be 100% for M to be reliable.[24] Hence, the internal condition is compatible with having false justified beliefs.[25]

---

[24] Caveats: first, Chapter 5 will explain, the narrowest individuation of methods allows only ever instance of a method M at *t*, where M is 100% reliable if it produces a true belief at t or 100% unreliable if it produces a false belief at t (see Goldman, 1979), in which case there will not be scope for having a justified false belief at t. Second, a modal construal of the reliability condition for justified would put pressure on the possibility of justified false beliefs. For instance, suppose that a belief is reliably formed at t just in case at the closest possible world in which not-p obtains, S does not believe that p. If S falsely believes that p in the actual world, then the closest world in which not-p obtains just is the actual world; in which case a false belief would be predicted as unreliable. In this thesis, I do not opt for the narrowest possible individuation of methods, nor do I argue that having a justified belief requires that one's belief is modally reliable.

[25] Chapter 5 considers methods.

Finally, the awareness condition is not a bi-conditional. This is for the reason that even occurrent awareness of all the externalist facts necessary for justification, may be insufficient for justified belief. The condition allows that one's first order justification may be defeated in cases where one holds a *blameworthy* second-order belief.

In the remainder of this chapter, I outline internalist views in the literature which differ to the awareness condition I have presented in this section. I then outline specific externalist conditions for knowledge which are, in principle, compatible with the possibility that knowledge also requires internal justification. Section 4 explains how the remainder of the thesis is structured.

§ 3.1. Other Internalist Views

Awareness internalism, as formulated in §3, is distinct from supervenience internalism. Supervenience internalism is the view that any two worlds which are internally identical are necessarily identical with respects to justification. To examine this further, we can imagine a sceptical world in which one is a disembodied brain-in-a-vat stimulated to have the experiences we have in the actual world, but in which one's empirical beliefs are false, and formed using an unreliable method of belief formation. Such a world would *look the same from the inside* as the actual world (i.e. we would think we were having the same experiences which we are having in the actual world); the world would ex-hypothesi be *phenomenally* the same as the actual world. If beliefs formed in the brain-in-a-vat world are *intuitively justified*, then supervenience internalism may explain that verdict.

Importantly, awareness internalism is inequivalent to supervenience internalism, for reasons explained in Chapter 4. Approximately, the accounts are inequivalent because it is compatible with awareness internalism that two worlds W and W* are internally the same (i.e. the same in their phenomenal facts) and yet different with respects to justification because one is reliable in W but unreliable in W. The awareness internalism of §3 allows phenomenal duplicates to differ in justification.

Another internalist view is Conee and Feldman's (2001). They argue that justification supervenes on the internal, but they describe the supervenience-base as comprising of one's entire occurrent or dispositional mental states, a view they characterise as "mentalism". According to mentalism, whatever contributes to justification is determined by occurrent and dispositional mental factors (ibid, p. 3). Mentalism holds that differences in justification cannot result from contingent, non-mental differences such as differences in reliability (ibid, p.3). Conee and Feldman express this with the strong supervenience principle: if any two individuals are exactly

alike mentally, then they are exactly alike concerning justification (ibid, p.3). Mentalism, however, differs from the awareness condition presented in §3. For awareness internalism allows that non-mental differences may confer justificatory differences. More specifically, the reliability requirement of the awareness condition allows differences in reliability to confer differences in justification between any two subjects S and S*, even if S and S* have the same mental states.

§ 3.2. Externalist Accounts of Justification

One pre-eminent externalist view, Goldman's (1979) account, makes a belief's justification depend only of whether one uses a reliable belief-forming process, and in particular whether a belief token is the result of a reliable process type. Goldman argues that reliable belief-forming processes are both necessary and sufficient for justified belief[26] According to externalist reliabilism, we may have second-order awareness that our beliefs result from reliable processes, but for the externalist reliabilist this is a contingent fact; awareness is not - strictly speaking - necessary for justified belief.

Goldman contends that since processes need not be perfectly reliable, one can have justified false beliefs. Consequently, Goldman's condition is similar to the awareness condition of §3 because both conditions make reliability necessary for justified belief and both are consistent with the possibility of justified false beliefs. The significant difference between Goldman's condition and the awareness condition of §3 is that Goldman's condition does not require second-order awareness.

Other externalist accounts of justification, however, are inconsistent with having justified false beliefs. For example, Williamson claims that knowledge is necessary and sufficient for justified belief. On Williamson's account (2000), one has a justified belief that p only to the extent that one knows that p. Williamson argues that knowledge is the norm for belief; when a belief conforms to the knowledge norm, then it is justified, and when it fails to conform to the knowledge norm, then it is unjustified. Williamson, however, does not analyse knowledge in terms of justification but, instead, analyses justification in terms of knowledge, where knowledge is regarded as a primitive and non-analysable mental state. Williamson's account of justification is externalist because the mental state of knowing is externalist. Accordingly, one can be in a mental state M without knowing or believing that one is in M (this is what he

---

[26] An austere version of Goldman's (1979) principle is "If S's believing that p at t results from a reliable cognitive belief-forming process (or set of processes), then S's belief in p at t is justified" (p. 13). Goldman refines the principle, to distinguish between belief-dependent and belief-independent cognitive belief-forming processes.

characterises as *luminosity failure*).[27] If knowledge is an externalist mental state, one can know that p without knowing or believing that one knows that p. By extension, since knowing that p is necessary and sufficient for a justified belief that p, one can be justified in believing that p without believing that one is justified in believing that p.

§ 3.3. Externalist Conditions for Knowledge

There are many conditions for knowledge, sometimes formulated as necessary and sufficient conditions for knowledge, which are not explicitly justificatory conditions. Many of these conditions are externalist (i.e. they may obtain independently of whether the believer believes they obtain for her belief)

Consider, for instance, D.M. Armstrong's (1973) treatment of knowledge as requiring a "law-like" relationship between one's belief and the truth of one's belief,". Armstrong writes, for some person A, "A's non-inferential belief that p is non-inferential knowledge if and only if: 1) p is the case 2) there is some specification of A (A(x)) such that, if any person is so specified, then, if they further believe that p, then p is the case." (p. 168). Armstrong presents the account as an "externalist" theory of "non-inferential knowledge.[28] Furthermore, on Armstrong's view, the law-like connections can "in principle" be investigated by the natural sciences. (ibid, p. 168), and crucially the law-like connection can "hold independently of us who may record its existence" (ibid, p. 168).

The law-like relation invoked by Armstrong need not be a relation which a believer is *aware* obtains for her belief. Furthermore, Armstrong is explicit that he is not offering an account of the reasons or justification for our beliefs but is instead offering an account of non-inferential knowledge.[29]

Similarly, subjunctive conditions for knowledge (conditions of the form *if p were false, then S would not believe that p* [~p→ ~Bp]) are not manifestly justificatory conditions. However, they are epistemically externalist conditions, to the extent that

---

[27] Williamson defines luminosity as follows: a condition C is luminous if and only if for every case a, if in a C obtains, then in a one is in a position to know that C obtains (2000, p. 95). Williamson characterizes a luminous condition as that which "always shines brightly enough to make its presence visible" (Ibid).

[28] Indeed, Armstrong is (to the best of my knowledge) the first person to use the term "externalist" to explicitly characterise theories of knowledge in the published literature (see p. 157). Notwithstanding, the crucial point is that Armstrong construes law-like relationships as "externalist" (ibid, p. 157), see also Fumerton (1995) on this idea.

[29] Armstrong (1973) regards the law-like relationship between the belief and the truth (or what he characterises as the state of affairs), as required for non-inferential knowledge, as similar how a reliable thermometer responds to the temperature. Armstrong describes such a relationship as involving a true belief which is unsupported by reasons (p.166). Armstrong's account avoids the regress problem to be discussed in Chapter 3(See, Armstrong,1973. p. 157).

they may obtain (for a belief) independently of whether the believer *believes* they obtain.[30] Subjunctive conditions need not offer analyses of knowledge; subjunctives have been supplemented with further conditions for knowledge (e.g. Nozick, 1981), which Chapter 5 will examine. Alternatively, subjunctive conditions may be specified as necessary conditions for knowledge without being necessary and sufficient conditions for knowledge (see Williamson, 2000). Subjunctive conditions may, in principle, be allied to either an internalist or an externalist account of justification.

The externalist conditions for knowledge which I will predominantly focus on in other chapters in the thesis are subjunctive conditions. Subjunctive conditions have been deployed in response to the problem of epistemic luck, to explain how specific types of epistemic luck undermine knowledge. Recall from earlier our adaption of Bertrand Russell's (1948) STOPPED-CLOCK example, where one cannot intuitively know that the time is 12.00 by having a true belief in the proposition if, unbeknownst to you, the clock stopped exactly 24 hours beforehand. How to respond to such cases? One response, which I will favour, is to argue that knowledge requires that one would not easily falsely believe that p, which is expressed with the subjunctive conditional *if S were to believe that p, then p would be true* ($Bp \rightarrow p$). Sosa calls this condition "safety" (Sosa, 1999). The truth-value of the safety subjunctive has since been translated modally, such that ($Bp \rightarrow p$) holds just in case there do not exist any *close possible worlds* – possible worlds similar to the actual world - in which one falsely believes that p.[31] Applying safety to STOPPED-CLOCK, there is a close world in which one falsely believe that the time is 12.00, hence one's belief in this proposition in the actual world is unsafe.

In summary, §3 has isolated the view of justification which will be supported, distinguishing it from externalism about justification and from externalism about knowledge. This will have provided a clearer understanding of the research hypothesis.

§ 4. Thesis Structure

The research hypothesis is that a construal of awareness internalism and its relationship to knowledge as a species of non-lucky true belief represents a compelling conception of justification. The conception appeals to the necessity of external conditions for justification, but parts company with externalism by arguing that a further (internal) awareness requirement is needed for epistemically justified belief.

---

[30] See, for example, Nozick (1981) who (for subject S and proposition p) does not make second-order awareness a necessary condition for sensitivity ($\sim_s p \rightarrow \sim B_s p$) to obtain. Chapter 5 examines this condition.

[31] Chapter 2, and esp. Chapter 5, will further define this notion.

The conception of justification is one that uses reliabilist insights to explain how justified belief amounts to more than merely blamelessly believing a proposition. Anti-luck considerations will be used to motivate awareness internalism. It will further be argued that two sceptical objections can be met, and that a non-dogmatic response to counterexamples may be adopted by avoiding the claim that knowledge must always depend on justified belief.

The remainder of the thesis is structured as follows:

Chapter 2 examines BonJour's NORMAN counterexample to externalist accounts of justification and knowledge. I identify difficulties for BonJour's account. I then motivate the intuition that NORMAN is a case of knowledge failure by appealing to anti-luck considerations. I consider the objection that the relevant way in which Norman is epistemically lucky (namely, that he is reflectively lucky) is benign. I explain how a sceptical arguments against awareness internalism depends on a particularist methodology; I explain the particularist methodology and I find support for the very methodology that poses a problem for awareness internalism. I then respond to the sceptical objection by arguing that the kind of luck at stake, when one fails to be internally justified, is not merely of the reflective variety but is instead the same type of luck at stake in the STOPPED-CLOCK example. Finally, I provide counterexamples to the proposal.

 I respond to the counterexamples by explaining how some kinds of knowledge and justification may survive in the absence of internal justification. I argue that by placing awareness internalism in a non-absolutist framework for knowledge attributions (following Hetherington, 2001), there is scope to offer a non-dogmatic response to counterexamples. This response involves a qualified externalism about knowledge, where a normatively unsatisfactory form of knowledge is consistent with not satisfying an awareness condition. I support this view by considering our comparative judgments.

Chapter 3 defines regress scepticism and considers the contention that awareness internalism generates regress scepticism (Bergmann, 2006; and BonJour, 1985). The chapter explains how sceptical consequences of awareness internalism have motivated externalist treatments of justification and knowledge (Bergmann, 2006) and non-foundationalist views about the structure of justification (BonJour, 1985).[32] The chapter explains and responds to regress scepticism. In response. I argue that the awareness requirement can avoid regress scepticism while maintaining an anti-luck motivation, providing one abandons the requirement that second-order doxastic

---

[32] Chapter 3 discusses this matter.

awareness must consist of a further *justified* second-order belief. However, I indicate a form of scepticism (specifically, the meta-epistemological scepticism which Fumerton (1995) identifies) awareness internalism and externalist accounts of justification are vulnerable to, explaining how awareness internalism may be viewed in light of this.

Chapter 4 considers an objection to awareness internalism. The objection considered is that justification does not require reliability. To understand this objection, I explain Cohen's (1984) demon problem, which implies that justification does not require reliability. I consider supervenience internalism as an alternative to awareness internalism to accommodate the demon problem. I argue that the demon problem is an inconclusive basis to abandon the reliability condition for justified belief, providing that we understand beliefs in the demon world as excusable but unjustified (Williamson, 2007). However, I argue the demon problem can only be adequately met by distinguishing between excuses and exemptions. It is contended that merely describing beliefs in the demon world as epistemically blameless would fail to explain our comparative normative intuitions, leaving awareness internalism vulnerable to a species of modal luck.

Chapter 5 analyses anti-luck conditions. I consider the anti-luck conditions deployed in earlier chapters, namely the condition that safety is needed to accommodate cases of epistemic luck. I consider the possibility that safety is the incorrect anti-luck requirement. In response, I provide further reasons for favouring safety ($Bp{\rightarrow}p$) over its logical contrapositive, sensitivity ($\sim p{\rightarrow}\sim Bp$). The chapter argues that safety rather than sensitivity can be supported as the correct requirement for propositional knowledge.

Chapter 6 considers how awareness internalism may be reconciled with a wide class of animals possessing knowledge. I consider the claim that it is correct to think of a wide class of subjects knowing and that this arises as a challenge to awareness internalism. In particular, the claim that animals have knowledge is contentious, but I explore the implications for awareness internalism if the claim is correct. I examine a pluralistic response to the problem consistent with the conclusions of Chapter 2, which allows for more than one kind of knowledge and I explain how this view would be consistent with awareness internalism. I use this as a basis for explaining a type of animal knowledge that may persist in the absence of satisfying an awareness condition for justified belief.

Finally, I summarise how the thesis has addressed the research question and which conclusions can reasonably be drawn.

Chapter 2

Internalism and Reflective Luck

Chapter 1 defined "epistemic externalism" as the view that knowledge does not require internal justification. By contrast, "epistemic internalism" was defined as the view that knowledge requires internal justification. This chapter outlines an intuitive objection to epistemic externalism, specifically that there are counterexamples to the claim that externalist conditions, such as reliability, are sufficient for justification. I motivate the claim that counterexamples to epistemic externalism highlight that externalism is vulnerable to a harmful species of epistemic luck. I distinguish between veritic luck (Engel, 1992a) and reflective luck (Pritchard, 2005). I explain how standard presentations have treated the counterexamples to externalism as *merely* instances of reflective luck rather than veritic luck. I present a sceptical challenge to internalism, which arises if reflective luck is ineliminable.

I respond to the sceptical objection by arguing that the cases marshalled in support of internalism are cases of veritic luck rather than merely cases of reflective luck. I exploit comparative intuitions to argue that differences in internal justification confer differences in veritic luck. And I argue that understanding counterexamples as *merely* instances of reflective luck cannot unify our intuitions about a class of cases. Thus, I argue that by understanding counterexamples as instances of veritic luck we can identify an anti-luck motivation for awareness internalism, one which does not entail global scepticism.

Finally, I consider and respond to counterexamples and objections to this proposal.

The chapter is structured as follows:

§ 1. I examine Laurence BonJour's (1985) counterexample to externalist accounts of justification and knowledge.

§ 2. I motivate the claim that BonJour has highlighted a harmful type of epistemic luck. However, I explain how the current literature treats counterexamples as cases of reflective luck rather than veritic luck. I explain how the safety condition has been used to support the standard view. I explain how the standard view poses a sceptical

challenge to awareness internalism, because reflective luck is ineliminable and making its elimination a necessary condition for knowledge would result in global scepticism.

§ 3. I respond to the sceptical challenge by arguing that BonJour's counterexample may be interpreted as a case of veritic luck.

§ 4. I consider and respond to several objections to the arguments of § 3. I ultimately recommend that a non-absolutist approach to defining knowledge (following Hetherington, 2001) allows us to accommodate a class of opposing intuitions about the counterexamples to externalism.

§ 1. The Intuitive Objection

I wish to start by discussing BonJour's (1985) counterexamples to epistemic and justificatory externalism (see Chapter 1, §2, for that distinction). BonJour intends his counterexamples as a challenge to externalist accounts of justification and knowledge generally. [33] His focus in 1985 is chiefly on justification, but on the proviso that knowledge requires justification, his cases are also counterexamples to epistemic externalism. To examine BonJour's account, I suggest taking the following view as an example of paradigmatic externalism:

> **Externalist reliability condition**: S has a justified belief that p just in case S's belief that p is formed via reliable belief-forming process or method.

Let us also recall the awareness requirement for justification (Chapter 1, § 3):

> **Awareness internalism:** Subject S has a justified belief B only if: 1) B has some property Φ AND S is aware that B has Φ (for a Φ that may hold for B in circumstances where S is unaware that B has Φ)

Imagine that a reliability condition is necessary for having a justified belief (per Chapter 1, § 3), which will be further supported in Chapter 3 and Chapter 4. An externalist view would be that having a reliably true belief is necessary and sufficient. By contrast,

---

[33] BonJour later (in 1987) applies the same objection to externalist accounts of knowledge, specifically to Robert Nozick's sensitivity analysis of knowledge (BonJour, 1987, pp. 297-314). The counterexamples in BonJour (1987) share the same logic as his earlier counterexamples in (1985).

awareness internalism insists on a further second-order awareness of a belief's reliability.

In the 4 examples discussed in this section, we should imagine that there is a subject "S" who is a reliable clairvoyant and who frequently has reliably true beliefs about the location of the President of the United States. In each case, S satisfies a necessary and sufficient condition for having an externally justified belief. We can further imagine that S satisfies an externalist condition for justified belief but fails to satisfy the awareness condition for justified belief. For brevity, I outline the structure of Cases 1-3, since BonJour argues externalism can accommodate Cases 1-3:

CASE 1: S has a reliably true clairvoyant belief that the President is in New York City but has a reason to think that her clairvoyant belief is false (SEK, p. 38)

CASE 2: S has a reliably true clairvoyant belief that the President is in New York City but has a reason to think that she does not possess the power of clairvoyance (SEK, p. 39)

CASE 3: S has a reliably true clairvoyant belief that the President is in New York City but has a reason to think that clairvoyance is impossible (SEK, p. 40)

We can envisage that Cases 1-3 satisfy Goldman's (1979) necessary and sufficient condition for justification because the subjects can be understood as having a reliably true belief. However, there is an intuition that the cases described are neither examples of justified belief nor cases of knowledge.

BonJour (1985) concedes that externalist reliabilism can accommodate Cases 1-3 by modifying the externalist view. Specifically, externalism can insist that the believer must not have any positive reason for questioning the reliability of the method or of the particular belief (SEK, p41). For BonJour, this requirement offsets the "subjective irrationality" associated with having (externalist) epistemic justification for one's belief that p while having a good reason to think that p is false, or that the method used to believe that p is unreliable (SEK, p. 41). In other words, externalism may allow for one's justification to be *defeated* by awareness of countervailing evidence. The modification is consistent with externalism because such a requirement does not demand that we must be in possession of a positive reason to think that our beliefs are *true* in order for

them to be justified. The requirement is negative, demanding only that the agent doesn't have any reasons against the belief.

To understand this further, Goldman (1988) proposed a distinction between weak justification and strong justification. According to that distinction, having a reliably true belief is not enough for having an epistemically justified belief. On Goldman's view, unreliable beliefs which are blameless are "weakly justified". By contrast, "strong justification" (which Goldman equates with *epistemic* justification) requires reliable belief formation *plus* the absence of blame.[34] On Goldman's view, weak justification is a necessary but insufficient condition for being strongly justified. Since Cases 1-3 are examples of blameworthy believing (e.g. wilfully ignoring countervailing evidence), Goldman's 1988 account of justification predicts that Cases 1-3 are not epistemically justified. In particular, since Cases 1-3 are not weakly justified, they are *a fortiori* not epistemically justified.

BonJour concedes that modifying the proposal to allow that awareness of evidence against the truth of the belief or the reliability of the method leads to a more plausible externalist account (*SEK*, p. 41). However, BonJour introduces a fourth case, which is intended to place the burden of proof on the externalist:

> NORMAN**:** "Norman, under certain conditions which usually obtain, is a completely reliable clairvoyant with respect to certain kinds of subject matter. He possesses no evidence or reasons of any kind for or against the general possibility of such a cognitive power or for or against the thesis that he possesses it. One day Norman comes to believe that the president is in New York City, though he has no evidence either for or against this belief. In fact the belief is true and results from his clairvoyant power under circumstances in which it is completely reliable." (SEK, p41).

Unlike Cases 1-3, Norman satisfies the modified externalist requirement because Norman is unaware of any grounds which count against either the reliability of the method of belief formation used to arrive at the belief or the truth of the belief.

Despite satisfying an externalist reliability requirement, BonJour regards Norman's beliefs as unjustified because it is *subjective irrational*. In BonJour's description, Norman lacks any reasons for thinking the proposition is true or for thinking that he does, in fact, possess the faculty of clairvoyance (p. 420), which

---

[34] A belief is weakly justified if it is ill-formed but blameless and non-culpable (1988, p. 53). By contrast, a belief is strongly justified just in case it is blameless, *and* it results from a reliable method of belief formation (ibid., p. 52).

ostensibly makes them subjectively irrational. Furthermore, BonJour claims that Norman violates his epistemic duty concerning the "cognitive goal of truth", writing that part of one's epistemic duty is to "reflect critically upon one's beliefs (p. 42), concluding that such reflection "precludes believing things which one has, to one's knowledge, no reliable means of epistemic access" (p. 42).

One necessary condition for awareness internalism (Ch1, § 3) is violated in NORMAN. This is because Norman lacks second-order awareness that his first-order belief (about the president's whereabouts) is reliably formed. For BonJour, in order to be justified one must at least *believe* with justification that a belief has a property which makes it likely to be true (p. 32). We can see that Norman lacks any such second-order awareness, which would preclude him from being able to reflect on the status of his first-order beliefs critically. BonJour's statements (1985, Ch2, p. 31) are consistent with this. For example, BonJour states that S has a non-inferentially justified belief B only if i) B has a feature X which makes it likely that B is true and ii) S believes that B has X.[35] On this view, the justification of B depends "on at least one other empirical belief".[36]

Externalist reliabilism, however, makes it possible that NORMAN is a case of knowledge because it predicts that NORMAN is a case of justified belief. By contrast, Norman's lack of second-order awareness means that he has neither an internally justified belief nor – given a justificatory condition on knowledge – knowledge of the proposition.

BonJour argues that Norman lacks knowledge and that he lacks knowledge because he lacks a justified belief. In the remainder of this section, I will examine BonJour's explanation of why Norman fails to have a justified belief. The section will outline concerns about BonJour explanation of NORMAN, before advancing – in Section 2 – an alternative interpretation of NORMAN.

§ 1.1. Laurence BonJour's Account

The starting point of BonJour's theory of justification is a distinction between epistemic justification and other kinds of justification, such as pragmatic or prudential justification. The mark of *epistemic* justification, BonJour states, is that it *aims* at truth.

---

[35] See BonJour (SEK, 85), pp. 31. BonJour regards the epistemic justification of one's first-order belief as dependent on a second-order belief.
Specifically, for a belief B to be justified, B must have some property Φ, where beliefs which have Φ are "highly likely to be true" - BonJour's claim on p. 32 is that B can be justified only if one "believes with justification the premises from which it follows that the belief is likely to be true". Chapter 3 will consider that condition further.
[36] Ibid., the requirement, as BonJour argues, leads to a regress (considered in Chapter 3)

BonJour further thinks that the aim of belief is truth (SEK, p. 8).[37] According to BonJour, one's beliefs are justified only to the extent that they are *directed* at this cognitive goal, where those directed at this goal are the ones we have "good reason" to think are true. [38] On BonJour's view, Norman is *epistemically irresponsible* because he violates his epistemic duty. BonJour states:

> "part of one's epistemic duty is to reflect critically on one's beliefs, and such critical reflection precludes believing things to which one has, to one's knowledge, no reliable means of epistemic access" (ibid, p. 42).

In what follows, I will consider the above claim that justified belief requires the fulfilment of an epistemic duty. I will then consider BonJour's statement that Norman is subjectively irrational, raising concerns about both claims.

If Norman is epistemically unjustified because he violates an epistemic duty, then – intuitively - Norman only violates an epistemic duty to the extent that he is free to suspended judgment about, or disbelieve, the proposition. Analogously, it seems intuitive that one *violates* a duty in committing an action only to the extent that one was free not to have performed that action. The thesis that belief is under voluntary control is called "doxastic voluntarism". I will suggest that doxastic voluntarism is a problematic feature of BonJour's account. First, I will explain how doxastic voluntarism is conceptually problematic. Second, I will explain how even if doxastic voluntarism is defensible on conceptual grounds, then it still seems intuitive that many justified beliefs do not emanate from the will.

Doxastic voluntarism is the thesis that our beliefs are - either directly or indirectly - under voluntary control. Direct doxastic voluntarism holds, approximately, that for any p one is aware of, at time *t*, one can choose to believe that p at *t*. By contrast, indirect doxastic voluntarism is the claim that even if we cannot choose to believe that p at *t*, we can choose how we conduct ourselves with respect to the grounds which are causally efficacious in the formation of our beliefs. Indirect doxastic voluntarism allows us to bear epistemic responsibility for believing that p at *t* due to how we have conducted ourselves (epistemically) before *t*.

---

[37] BonJour writes that "what makes us cognitive beings at all is our capacity for belief, and the goal of our distinctively cognitive endeavors is truth" (SEK, p. 7). On his view, *the cognitive goal of having true beliefs* has implications for which beliefs one *ought* to accept (SEK, p8).
[38] Likewise, BonJour argues that we are *epistemically responsible* when we pursue this goal, and when we have a good reason to believe that p, and we are *epistemically irresponsible* when we "accept" a belief for which we lack accessible supporting reasons (SEK, p. 8).

Doxastic voluntarism is contentious. For example, Bernard Williams (1973) contends that the notion of choosing to believe that p is discordant with the phenomenology of belief. Williams observes that "if I could acquire a belief at will, I could acquire it whether it was true or not"; Williams, in the same paragraph, concludes that "if in full consciousness I could will to acquire a "belief" irrespective of its truth, it is unclear that...I could seriously think of it as a belief. i.e. as something purporting to represent reality" (1973, p. 148). William Alston, also, uses an anatomical analogy, stating that "choosing don't hook up with propositional attitude inaugurations, just as they don't hook up with the secretion of gastric juices" (Alston, 2005, p 21). In the same vein, Audi affirms we cannot *believe* at will in the same way that we can move bodily parts at will" (Audi, 2001, p. 29).

Examples bear this out. For instance, if asked to believe that the world is 5000 years old, the belief does not seem available if we do not think the proposition *true* (whatever state I am in when considering this proposition, it does not *feel* like a belief). Of course, whether direct doxastic voluntarism is true won't be settled by a few glib examples. Furthermore, Bon Jour concedes that we may retreat to an indirect form of direct voluntarism, where we "bracket" certain beliefs over an extended period of time (SEK, p. 46).[39]

However, there is a more pressing difficulty for understanding NORMAN as a case in which the believer violates an epistemic duty. The difficulty is that even if some of our beliefs are chosen, as doxastic voluntarism allows, it does not appear that all of our beliefs are under voluntary control. Moreover, many of the beliefs which are not obviously under our voluntary may be either justified or unjustified.

Consider beliefs which result from paranoid delusions. Such beliefs are not under voluntary control but instead arise due to facts about a believer's pathology. It does not seem that a paranoid believer violates an epistemic duty and, furthermore, the beliefs appear *blameless* as the believer is not free to believe otherwise. As Conee and Feldman (1985) have observed in discussing paranoid believers, the beliefs are intuitively unjustified even though they are blameless. That is, the believer *ought* not to hold such beliefs even if *they* are blameless.[40]

In the case of the many perceptual beliefs we form, it does not seem that we choose the beliefs; instead, we tend to form these beliefs automatically, as a result of

---

[39]BonJour, for example, meets the objection of doxastic voluntarism by stating that "over an extended period of time" beliefs can be bracketed, and that we can refuse to take some beliefs seriously, which includes a refusal to act upon or draw conclusions from a belief (SEK, p. 46). My interpretation of BonJour is that he is endorsing a version of indirect doxastic voluntarism.
[40] Williamson (2007) similarly observes that in condemning a belief as *irrational,* we are not implying that a believer can switch the belief "on and off like a light" (p. 2).

being attuned to our environment. Yet many, if not most, of our perceptual beliefs are intuitively justified.[41] BonJour's notion of fulfilling a duty is that when one has fulfilled a duty then one is *praiseworthy* in the belief. But the justified perceptual beliefs, which come to us automatically, seem neither praiseworthy nor blameworthy, even though they may be evaluated as either justified or unjustified.[42] Such cases suggest that belief norms are not deontological norms. Consequently, there are at least two ways of denying that justified belief requires doxastic voluntarism:

(**1**) Doxastic voluntarism may be denied on conceptual grounds, such that no beliefs are under direct (or indirect) voluntary control.

(**2**) There are a) intuitively justified belief which do not emanate from the will (e.g. perceptual beliefs) and b) intuitively unjustified beliefs which do not involve a dereliction of duty (because the believers are blameless).

Importantly, the second claim does not entail that there are not some beliefs under voluntary control; more modestly, the second claim affirms that belief norms are deontological norms. [43] It does so by affirming that many of our justified beliefs are not the product of volition and that some instances in which one has an unjustified belief do not involve a dereliction of duty because the believer is blameless. Notice that we need only the second – more modest – reason for resisting the idea that epistemic justification requires doxastic voluntarism and the fulfilment of epistemic duty. The second claim rejects *the requirement of doxastic voluntarism* for epistemic justification, rather than the conceptual possibility of beliefs being under voluntary control.

The requirement that we can choose beliefs, directly or indirectly, is therefore a problematic element of BonJour's explanation of Norman.

Subjective Irrationality

BonJour further claims that Norman's belief is *epistemically irresponsible* and *subjectively irrational*. Let us consider the connection between epistemic responsibility and subjective irrationality, which does not appear straightforward. Firstly, if epistemic responsibility requires that we must *choose* our beliefs, and thus *be responsible for them*, then the notion would appeal to doxastic voluntarism and would straddle the internalist explanation of Norman with a controversial claim about the

---

[41] In both cases it is also doubtful whether we should describe the subjects are violating and fulfilling an *epistemic duty*. Audi (2001, p.21) makes a similar point.
[42] See Also Engel (1992b, p. 140) on this point.
[43] It is open to one to still argue that there are *epistemic obligations* without doxastic voluntarism (See for example, Feldman, 2000)

nature of belief. It seems this is not, necessarily, the notion of responsibility which the internalist should *want* to draw our attention to in soliciting intuitions about Norman.

Consider the claim that Norman is *subjectively irrational*. There are many ways in which an agent may be subjectively irrational. For example, a believer might believe two propositions with mutually exclusive implications (i.e. have a doxastic defeater) or believe a proposition when he knows there is evidence against the proposition (i.e. having a normative defeater). In such cases one *does* seem epistemically blameworthy.

Envisage explaining the intuition about subjective irrationality in terms of Norman being blameworthy. As we have already seen, blameworthiness may explain BonJour's Cases 1-3. The reason for this is because the intuition in Cases 1-3 is that the believers are blameworthy because they ignore countervailing evidence.

The difficulty for BonJour's position is that the reasons for thinking Cases 1-3 are cases of subjectively irrational beliefs are not reasons for thinking Norman is subjectively irrational. This is because Cases 1-3 involve the agent being aware of grounds not to hold the belief, which motivates the claim that the beliefs are blameworthy and subjectively irrational. By contrast, Norman's situation is not analogous to the situation of the believers in Cases 1-3. Norman according to BonJour's description, possesses no evidence against the general possibility of clairvoyance or against the fact that he possesses clairvoyance. This is not to dispute BonJour's verdict on Norman; instead, it is the *explanation* of why Norman is unjustified which is in contention. Specifically, the reasons for thinking that Cases 1-3 are subjectively irrational (namely that they are blameworthy) do not appear to be reasons for thinking that Norman is subjectively irrational.

The concern is that the explanation of why Cases 1-3 are epistemically unjustified does not generalize to Norman. Next, I shall examine a way of understanding Norman, which is not necessarily vulnerable to that concern.

§ 2. A Case of Epistemic Luck

In addition to claiming that Norman is subjectively irrational, BonJour highlights a further intuition about Norman. Specifically, BonJour claims that it is an *accident* from Norman's perspective that his belief is true:

> "One reason why externalism may seem initially plausible is that if the external relation in question genuinely obtains, then Norman will in fact not go wrong in accepting the belief, and it is, *in a sense*, not an accident that this is so: it would not be an accident from the standpoint of our hypothetical observer who knows all the relevant facts and laws. But

how is this supposed to justify Norman's belief? From his subjective perspective, it *is* an accident that the belief is true" (BonJour, SEK, 1985, pp. 43-44).

Here BonJour has highlighted an anti-luck intuition. Specifically, in the absence of Norman having any kind of awareness that his belief is true and reliably formed, it seems a matter of luck that the belief is true.[44] BonJour implores us to judge the justifiability of Norman's belief from "Norman's own perspective rather than from one which is unavailable to him" (SEK, p. 44).

The implication is that a species of epistemic luck excludes NORMAN as a case of justified belief and knowledge. More recently, Bergmann (2006) also identifies an anti-luck intuition in the case of NORMAN. Bergmann calls this the "*Subject Perspective Objection*" to externalism*:*

> The Subject Perspective Objection "If the subject holding a belief isn't aware of what that belief has going for it, then she isn't aware of how its status is any different from a stray hunch or arbitrary conviction. From that we may conclude that from her perspective it is an accident that her belief is true. And that implies that it isn't justified" (Bergmann, 2006, p. 12).[45]

The Subject Perspective Objection is an intuition about epistemic luck, or accident as Bergmann characterises it.[46] Indeed, Bergman regards the Subject Perspective Objection as the "clearest and most compelling" motivation for endorsing awareness internalism (p. 12). Notice that when we frame the explanation in these terms, we do not need to state that the reason why BonJour fails to know is that he neglects his epistemic duty, or that he is subjectively irrational. Rather, the worry is that by *merely* satisfying epistemically externalist conditions, Norman's belief is epistemically lucky in a way which undermines his knowledge.

Given what BonJour writes and how Bergmann interprets NORMAN, we have the beginnings of an explanation of why externalist (necessary and sufficient) conditions

---

[44] Further evidence for a luck-based diagnosis is found in BonJour (2003, p. 27),.

[45] Bergmann does not defend the Subject Perspective Objection to externalism, and he contends that the only versions of internalism which can meet the Subject Perspective result in scepticism. Chapter 3 examines Bergmann's argument for this position.

[46] Luck and accident differ (see Pritchard, 2005). For example, if one wins the lottery, then that is lucky, but it is not an "accident" – as Pritchard highlights, accident fits better in contexts where we are describing un-intentional consequences. However, we may assume – for present purposes - that accident and luck are understood in the same way and that accident is being used here to describe a case of epistemic luck

for justification may be insufficient for justified belief and insufficient for satisfying a justificatory condition for knowledge. Specifically, one can satisfy an externalist requirement and yet have a belief which is epistemically lucky and which intuitively does not constitute justified belief or knowledge. Importantly, even though we may offer a different blame assessment of NORMAN compared to Cases 1-3, it seems true that all the cases are examples of epistemic luck. After all, if it is lucky *from Norman's perspective* that his belief is true, then it is equally a matter of luck from the subjects' perspectives in Cases 1-3 that their beliefs are true.

What is notable about Cases 1-3 is that they are unlike NORMAN in that the attribution of blame (and subjective irrationality) does reasonably apply, yet they are like NORMAN in that they are epistemically lucky. The cases of blame (Cases 1-3) exhibit the luck which NORMAN exhibits and which threatens his justification and knowledge. We have, then, the beginning of an explanation of why Norman fails to know. The explanation does not appeal to the claim that Norman is blameworthy or guilty of violating an epistemic duty.

If the problematic feature of NORMAN is that his beliefs are epistemically lucky, there would seem no straightforward way in which Goldman's distinction between weak justification and strong justification could accommodate NORMAN, where weak justification is merely blameless belief. If the problem is that NORMAN is lucky rather than blameworthy, then NORMAN can satisfy Goldman's conditions for epistemic justification by being blameless and reliable (and therefore strongly justified) and yet still be epistemically lucky.

§ 2.1. A Case of Veritic Luck?

I have appealed to the intuition that Norman has an epistemically lucky belief because he does not satisfy an awareness condition for justified belief. What is striking about NORMAN is how unaware he is that his belief is reliably formed. On first inspection, it seems that in the absence of awareness, the belief seems *luckily true,* at least from the subject's perspective.

But in what *way* is Norman epistemically lucky? Recall the STOPPED-CLOCK EXAMPLE. Suppose a believer truly believes that the time is 12.00, but unbeknownst to the him, the clock stopped exactly 24 hours beforehand. In this and similar cases, such as those Gettier (1963) brought to our attention, it seems intuitive that one's belief is luckily true in a way which excludes, or at least degrades, knowledge.

A standard diagnosis of epistemic luck is that it involves an environmental abnormality, or a type *bad luck* (e.g. the clock has stopped) which is cancelled out by

good luck (e.g. that the believer happens to be seeing the clock at exactly 12.00) which generates the intuition that his belief is luckily true (Zagzebski, 1994). Here, we may pinpoint some abnormality in the believer's environment, the *bad luck*, together with some other feature of the believer's environment which, fortuitously, allows her to have a true belief. On first examination, Norman's belief does not seem lucky *in the same sense* because there is nothing in Norman's *environment* which makes it immediately apparent that under similar circumstances to Norman's actual circumstances, Norman would falsely believe.

To see this further, let us appeal to the safety condition. Following Ernest Sosa (1999), one's belief that p is safe just in case one would not easily falsely believe that p (,p. 144). This is expressed by the subjunctive conditional that *if* one were to believe that p, *then* p would hold. (Bp□→p).₄₇ Safety theorists have translated this modally:

> Safety: S's belief that p is safe in the actual world just in case at all close
> possible worlds in which S believes that p, p is true.₄₈

The STOPPED-CLOCK example clearly illustrates how safety can explain cases of epistemic luck, as in this case it seems *easily the case* that one could have falsely believed that the time is 12.00. In the language of possible worlds, there is a possible close possible world in which one mistakenly believes that the time is 12.00. The luck at stake in STOPPED-CLOCK cases (and Gettier's seminal cases), which safety excludes for a wide class of cases, has been called *veritic epistemic luck* by Mylan Engel (Engel, 1992a). I will adopt Engel's terminology. Veritic luck can be presented as follows:

> **Veritic luck:** S has a veritically epistemically lucky belief in some
> proposition p at time t just in case, given S's evidential situation, at t it
> is a matter of luck that S's belief that p at t is true. (Engel, 1992a, p. 67)

₄₇Bp□→p is distinct from Nozick's (1981) sensitivity condition [~p□→~Bp].Neither safety nor sensitivity are equivalent because subjunctive conditionals do not contrapose (see Sosa, 1999,). Chapter 5 compares the two conditions.

₄₈ Safety has, also, been relativized to the method of belief formation (M) which one uses in the actual world (Sosa, 2007, p. 26 – who adopts "basis relative" safety). Method-relativized safety holds that close possible worlds to the actual world in which S believes that p via some method other than M are not relevant to whether S safely believes that p. More schematically: Safety (method relativized): S's belief that p (via method M) is safe in the actual world just in case at all close possible worlds in which S believes that p (via M), p is true. Method relativization has been used to avoids counterexamples. Note also that there is weaker understanding of the safety condition. See Chapter 5, § 4 of this thesis.

The presumption is that NORMAN is safe and not veritically lucky. In particular, he is using a reliable method of belief formation and he tends to have true beliefs about the President's whereabouts. Furthermore, NORMAN may be construed as having a belief which is *conceptually connected* to the truth-value of what he believes. There does not appear to be a severed conceptual connection between the belief and the truth-value of the belief in NORMAN. Instead, as Bergmann (2006) indicates, BonJour is directing us to a variety of luck distinct from veritic luck, a type of luck considered solely *from the subject's perspective*.

Textual evidence for the view that NORMAN is not veritically lucky is found in BonJour's own characterisation of the case, when he writes that if the external relation (clairvoyance being reliable) genuinely obtains, then from the standpoint of a hypothetical observer who is aware of the facts, it is not an accident that Norman's belief is true (ibid, p. 43). Furthermore, BonJour later wrote that - in cases analogous to Norman, where the belief is formed using a reliable method - we may "suppose" that the belief satisfies a modal condition for knowledge.[49] More Recently, Job de Grefte (2017) claims that Norman would "not easily falsely believe" and as such his belief is not vertically lucky, even though his beliefs are "reflectively lucky" (pp. 3824-3825). Given that the safety subjunctive just is the requirement that one would not easily have falsely believed that p, the material implication of Grefte's claim (on the modal account of safety) is that there are no close possible worlds in which Norman falsely believes that p.

We need to be mindful, however, of one challenge to the view that NORMAN is not veritically lucky. Notice that the notion of veritic luck Engel employs is that one's belief is not veritically lucky just in case, *given the evidence which one has*, it is not a matter of luck that the belief is true (p. 67). It is thus open to contend that since Norman has no evidence for his belief, it is a matter of luck that his belief is true.[50]

I suspect the above way of supporting the verdict that NORMAN is not veritically lucky would be indecisive because there may still be *another sense* in which Norman's belief is lucky. In particular, it may be argued that – unlike STOPPED-CLOCK - there *is* a conceptual connection between Norman's belief and the truth-value of what he believes and that this is because Norman is using a reliable method of belief formation. Relatedly, this connection may obtain irrespective of whether Norman has evidence (in some internalist sense of evidence, i.e. one which allows him to lack evidence because

---

[49] BonJour claim about sensitivity, in a case analogous to Norman, that we may "suppose" that the case satisfies the sensitivity condition (1987, p. 303-304). The same supposition may be made of safety.

[50] I am indebted to an anonymous reviewer from the Australasian Journal of Philosophy on this point.

he is not aware of any evidence) for his belief. Simply lacking accessible evidence need not entail that Norman is lucky in *the same way* as STOPPED-CLOCK is lucky, providing that the conceptual connection between the belief and the truth-value is not severed. With this in mind, we may revise the veritic luck principle as follows:

> **Veritic luck (non-evidential construal)**: S has a veritically epistemically lucky belief in some proposition p at time t just in case, at t it is a matter of luck that S's belief that p at t is true.

On the above condition, it does not follow that Norman is lucky in the same way as STOPPED-CLOCK is lucky. This is because, even if it is a matter of luck that Norman's belief is true *given the evidence he is aware of,* it is open to claim that given the evidence Norman is unaware, or given possible non-evidential facts about his situation (for example, that he uses a reliable method) it is not a matter of luck that his belief is true.

This leaves us with the view that Norman is not veritically lucky but that there is a *subjective* sense in which Norman is lucky (i.e. a sense of luck considered solely from the perspective of the believer). Perhaps the most perspicuous account of this *other* type of luck is offered by Duncan Pritchard (2005). Pritchard distinguishes *veritic luck* from *reflective luck*:

> **Reflective luck**: Given only what the agent is able to know by reflection alone, it is a matter of luck that her belief is true. (Pritchard, 2005, p. 175)

Pritchard presents *reflective luck* as distinct from *veritic luck*. For example, Pritchard (2005) discusses chicken-sexers, people who can reliably distinguish the sex of hatchling chickens by inspection, but without further beliefs about the method of belief formation they use to arrive at true belief tokens about the sex of hatchling chickens. Pritchard distinguishes between the unenlightened chicken-sexer who forms reliable beliefs but who mistakenly thinks her method of belief formation is visual rather than olfactory, and the enlightened counterpart who knows that she is using her olfactory senses. [51] Pritchard highlights how initially it seems that the enlightened sexer is neither veritically nor reflectively lucky, whereas the unenlightened sexer is not

---

[51] Pritchard (2005) (Chapter 6, esp. p. 176).

veritically lucky but *seems* reflectively lucky.[52] However, Pritchard understands both chicken sexer's as having safe beliefs which are not veritically lucky.

Pritchard highlights that on a modal account of luck one must endorse *a non-standard ordering of possible worlds* for reflective luck (2005, Chapter 6). On the standard ordering of possible worlds, if there are no close possible worlds in which one falsely believes that p, then one's belief that p is safe and free from veritic luck, where these worlds are ordered in terms of their objective similarity to the actual world.

To distinguish reflective luck from veritic luck, Pritchard orders worlds non-standardly where worlds are ordered in terms of their reflective similarity to the actual world.

> **Non-standard ordering of possible worlds**: A belief that p is free from reflective luck in the actual world a only if there are no close possible worlds to a (ordered in terms of what an agent is able to know via reflection alone) in which S falsely believes that p.[53]

The difference between veritic and reflective luck now hinges on how the relevant worlds are ordered, with our intuitions about veritic luck being tested against a standard ordering of worlds, and our intuitions about reflective luck being tested against a non-standard ordering of possible worlds.[54]

For Pritchard, an under-determination sceptical argument highlights that reflective luck is ultimately ineliminable (Pritchard, 2005, see, pp. 203-207). The underdetermination argument appeals to the claim that to have internalist justification for an empirical belief that p, we must have evidence for p which excludes sceptical hypotheses in which ~p obtains; for example, hypotheses in which one falsely believes that p because one is a disembodied brain-in-a-vat. The underdetermination argument is just that, given our reflectively accessible evidence, we cannot exclude the possibility that one is not a disembodied brain-in-a-vat [BIV] in which one falsely believes that p. Notice that both BIV and ~BIV scenarios seem *close* when worlds are ordered purely in

---

[52] It is important to emphasise that these are the initial intuitions Pritchard draws our attention to. Ultimately, Pritchard contends that both cases are reflectively lucky, and that it is the nature of knowledge attributions which disguises this fact (2005, pp. 204-208).

[53] Pritchard does not offer this exact statement but describes the view more informally (see, Chapter 6, p. 175).

[54] On the non-standard ordering, Pritchard identifies, beliefs would be ordered in terms of how the agent believes she is forming her belief in the actual world. By contrast, we are told that on the standard ordering used for the safety condition, possible worlds are ordered in terms of how the agent believed she formed her belief in the actual world (p. 175). Cases wherein one has no beliefs whatsoever about their method are, as Pritchard points out, instances in which just about any possible world can count as close.

terms of what an agent can know via reflection alone. Thus, for some putatively true empirical p, one's evidence does not allow one to tell or identify that a [BIV&~p] possibility obtains. If knowing via reflection alone is understood in this way – as requiring that one's reflectively accessible evidence for p is not *underdetermined*, then it *seems* that the evidence one would have for the proposition p in a ~BIV world does not allow one to tell that one is not in a [BIV&~p] world.[55]

The result is that all our empirical beliefs are reflectively lucky. To see why, consider a non-standard ordering of worlds, where close possible worlds to the actual world α are ordered solely in terms of what one can know via reflection alone in α. Notice that one cannot know via reflection alone that ~[BIV&~p]. On the non-standard ordering, [BIV&~p] worlds are no more distant from α than [~BIV&p] worlds are from α. Thus, on the non-standard ordering, one's belief that p is reflectively lucky in α even if it is not veritically lucky in α. If knowledge is incompatible with reflective luck, then it would transpire that we know much less than we ordinarily think we do, because the non-standard ordering predicts that putatively true beliefs are reflectively lucky. On first blush, the sceptical implications of an anti-reflective luck condition seem wide-reaching. Furthermore, if there is an anti-sceptical presumption that we *do* know in broad class of cases, then it seems that being reflectively lucky need not exclude knowledge.

By contrast, it seems that being unsafe and therefore not veritically lucky – as seen in the STOPPED-CLOCK case – *does* intuitively exclude, or at least degrade one's knowledge. Moreover, since the safety condition which targets veritic luck uses the standard ordering of possible worlds, radical sceptical worlds are not pulled into the domain of close possible worlds to α. This is because, on the standard ordering, possible worlds in which one is a BIV are *distant* and thus irrelevant to whether one's belief that p is safe and free from veritic luck in α. If possible worlds are ordered in terms of their objective similarity to α, we can see that BIV worlds are irrelevant to the safety of one's belief in α. By using the safety condition, veritic luck is eliminable even if reflective luck is ineliminable.

In summary, given an anti-sceptical presumption and the demands of an anti-reflective luck condition, we have grounds for thinking that reflective luck is *benign*, or at least not knowledge-killing. This is important for awareness internalism, because if NORMAN is merely *reflectively lucky* because he fails to have awareness of the fact he is using a reliable method (or of other non-epistemic facts) and if reflective luck is

---

[55] Pritchard appeals to the under-determination argument. (Pritchard, 2005). See p. 204.

understood as a type of epistemic luck compatible with knowledge, then that undercuts the argument for awareness internalism which appeals to our anti-luck intuitions.

Pritchard's view is that internalists are seeking a more robust, epistemically desirable knowledge. According to this view, even epistemic externalism must countenance that, in the absence reflectively accessible evidence which allows one to rule out that one is not a brain-in-a-vat, the knowledge one has is far from ideal. This does not, on Pritchard's view, imply that knowledge is incompatible with reflective luck.[56]

§ 2.2. The Reflective Luck Argument

The last section considered the idea that reflective luck is ineliminable and that making its elimination a necessary condition for knowledge would have sceptical implications.

Anti-sceptical considerations motivate the claim that reflective luck is benign (or at least not knowledge-killing). Accordingly, if all of our beliefs are reflectively lucky, then insisting on its elimination would mean that we have much less knowledge than we intuitively think we do. The argument I will examine is that eliminating reflective luck from one's knowledge would commit us to global scepticism. The argument asserts that the implication of global scepticism is so implausible that we have reason to abandon anti-reflective luck conditions for knowledge. I will call this the "Reflective Luck Argument", which consists of the following premises:

I.   Beliefs which are not internally justified (where second-order awareness is lacking) are reflectively lucky (E.g. NORMAN) on the non-standard ordering of possible worlds.

II. However, all of our empirical beliefs are reflectively lucky because no degree of awareness (among finite beings) is sufficient to exclude sceptical worlds from the domain of close worlds (when worlds are ordered non-standardly).

III. The requirement that beliefs must be free from reflective luck to be cases of knowledge is unsatisfiable for finite believers and would result in global scepticism.

IV. Many of our empirical beliefs are intuitively cases of knowledge.

V.   Global scepticism is false.

---

[56] See for example Pritchard's claim (p.207) that eliminating reflective luck is still "epistemically desirable", even if ultimately unattainable.

Therefore,

VI. Knowledge is compatible with reflective luck.

Premise **V** of The Reflective Luck Argument embodies the presumption that *attributor scepticism* is false, where attributor scepticism is the view that when we make epistemic ascriptions, of the form "S knows that p" or "S has a justified belief that p", those ascriptions are incorrect. The Reflective Luck Argument implies that we should reject anti-luck conditions that imply attributor scepticism

If the Reflective Luck Argument is sound, then Norman does not fail to know because his belief is reflectively lucky, lest we succumb to global scepticism. If it is *reflective luc*k which motivates awareness internalism, then at least one motivation for endorsing internalism is undercut (specifically, the motivation which appeals to our intuitions about luck). If reflective luck does not undermine Norman's knowledge, then it seems we have no reason for endorsing awareness internalism based on anti-luck intuitions alone.

§ 2.3. Responses: weaker safety and methodism

I will describe two ways in which the Reflective Luck Argument might be resisted. Both ways involve retaining the view that NORMAN is *merely* reflectively lucky while preserving the view that NORMAN is not veritically lucky. I will explain why both proposals are wanting.

The first response to the Reflective Luck Argument is to abandon the claim that the elimination of reflective luck requires truth-adherence at *all* close possible worlds.[57] I raise this because I have so far assumed that a belief is free from reflective luck just in case there do not exist *any* close worlds (when worlds are ordered non-standardly) in which the believer falsely believes the target proposition (i.e. that the belief is *strongly reflectively safe*). This condition generates scepticism because there appears no principled way of excluding sceptical worlds from the domain of close worlds. However, an alternative approach – consistent with a weaker safety principle – is to insist that a belief is free from reflective luck just in case at *most* close worlds (ordered non-standardly) the agent's belief is true.[58]

---

[57] Where reflective safety just means truth adherence at close worlds ordered in terms of what an agent can know via reflection alone.
[58] A suggestion of this kind is in the internet encyclopedia of philosophy, see. https:www.iep.utm.edu (article on epistemic luck).

One response to Norman is thus to support the weaker requirement that reflective safety requires that at *most* close worlds (ordered non-standardly) one's belief is true. For instance, it may be that while there are close sceptical worlds in which Norman falsely believes, these worlds are far less numerous than the close non-sceptical worlds in which Norman truly believes. An anti-reflective luck condition would now tolerate the existence of *some* close error worlds, providing that they are not the most dominant strain of close possible worlds. In this way, it may seem we can explain why Norman is reflectively lucky while avoiding attributor scepticism.

I suspect that retreating from *strong reflective safety* to *weak reflective safety* is an ad-hoc move. The difficulty being that we lack robust modal intuitions that the weaker (reflective) safety requirement is satisfied here. The case depends on our modal intuitions as to whether one's belief would be true at *most* close possible worlds (ordered non-standardly). Yet on the non-standard ordering, for every close possible world β in which Norman's belief is true, there is a close possible world β* which is phenomenologically the same as β but in which an evil demon ensures that Norman's belief is false in β*. The difficulty when we consider modal space is that there is not just *one* sceptical world but rather a plurality of sceptical worlds in which Norman's belief is false. At a minimum a retreat from strong to weak reflective safety does not look promising because it does not yield a robust modal intuition that the condition is satisfied.

A second response to the Reflective Luck Argument is to reverse the argument as follows: since reflective luck *is* incompatible with knowledge, then sceptical implications of an anti-reflective luck condition do not imply that knowledge is compatible with reflective luck.

The above second response would challenge the anti-sceptical presumption of the Reflective Luck Argument. Recall how the arguments holds that we have knowledge in a broad range of cases and that global scepticism is false, implying that a necessary condition for knowledge ought to make it possible that we know. By taking this stance, the Reflective Luck Argument is particularist in spirit, following Roderick Chisholm (1982, pp. 65-68). Chisholm does not discuss "particularism" in relation to epistemic luck, but instead characterises a particularist approach to criteria for justified belief. Chisholm describes a particularist approach to the conditions for justified belief as one which makes it an adequacy condition on any criterion *c* (for justified belief) that *c* explains particular cases of justified belief. Extending this thought to epistemic luck: a particularist approach to luck holds that an anti-luck condition, if it is plausibly a necessary condition for knowledge, ought not to clash with the intuition that we know in a broad range of cases (i.e. ought not to imply scepticism). If reflective luck is

ineliminable and if making its elimination a necessary condition for knowledge would result in scepticism, then particularism implies reflective luck is compatible with knowledge.

The converse approach, what Chisholm termed *methodism* (1982, p. 66), holds that whether a belief B is justified is a function of whether B satisfies the correct criterion *c* for justified belief. On that view, if B satisfies *c* then B is justified. Once again, applying methodism to anti-luck conditions, a methodist approach to luck holds that whether a belief B is a case of knowledge is a function of whether B satisfies the correct anti-luck condition(s). Notice that a methodist approach to reflective luck would hold that sceptical consequences are *not enough* to imply that reflective luck *is* compatible with knowledge, for if the correct anti-luck requirement does require that one's beliefs are free from reflective luck if they are to constitute knowledge, then radical sceptical consequences will not count against an anti-luck condition. Global scepticism becomes a price we must be willing to pay.

I have assumed a particularist approach to reflective luck because I suspect a methodist approach to reflective luck is unattractive. First, methodism would be revisionist of all of our ordinary intuitions about cases of empirical knowledge – global scepticism seems an *exorbitant* price to pay.[59] Relatedly, the problem for methodism is, as Chisholm observes about justification, its use of arbitrary criteria: if we do not take any particular propositions to be justified irrespective and independently of the criterion for justified beliefs, then it is unclear what reasons one would have for adopting a given criteria as opposed to some other criteria (p. 67). Analogously, if all beliefs are reflectively lucky, then we have no more reason to think that reflective luck is compatible with knowledge than we do for thinking that reflective luck is incompatible knowledge.

I do not discount a methodist approach, when defined as the possibility of one's methodological commitments shedding light on whether any given intuition is apt. For example, a commitment to a type of luck as harmful for knowledge might lead us to revise our intuitions about whether instances of belief are cases of knowledge. Suppose we recognize that a case of belief formation B is veritically lucky and we have a theoretical commitment to veritic luck as incompatible with knowledge. If so, then it is not implausible that regarding B as veritically lucky would make it intuitive that B is a

---

[59] Bergmann (2006) appeals to the distinction between particularism/methodism in discussing responses to the Subject Perspective Objection, also conceding that "the methodist approach is, on occasion, a good one". However he highlights the same point which I do in this section, namely that "the scepticism implied by the strong awareness requirement is global.... surely this implication is a good reason to reject the proposed criterion" (p. 23).

case of knowledge failure. It is possible that anti-luck intuitions may be informed by and adjusted in light of one's theoretical commitments.

Adjusting our epistemic intuitions can only work if the theory enjoys at least *some* degree of intuitive support, so that our theory and our intuitions at large can co-exist in a state of reflective equilibrium. In the case of reflective luck however, this is not the case: the theory appears to enjoy no support from our intuitions about knowledge, because global scepticism is radically revisionist; the theory does not get off the ground.

In summary, if reflective luck is ineliminable, we cannot reasonably argue that knowledge is incompatible with reflective luck. Therefore, if Norman and the unenlightened chicken-sexer are *merely* reflectively lucky, then this would be insufficient for knowledge failure and would not show that an awareness condition is required for justified belief and knowledge.

§ 3. Safety and Internal Justification: a proposal

§2 explained the standard interpretation of Norman as safe and not veritically lucky, even though his beliefs are reflectively lucky. I outlined the argument that reflective luck is ineliminable – and that an anti-sceptical presumption implies that reflective luck is compatible with knowledge. Finally, I considered two unsatisfactory responses to the Reflective Luck Argument, supporting a particularist approach to understanding whether a given type of luck is compatible with knowledge.

My response to the Reflective Luck Argument will be to advance the following proposal: that Norman is not merely reflectively lucky but is also veritically lucky because he fails to satisfy an internal awareness condition for justified belief. The argument will be structured as follows:

§3.1. I present grounds for doubting that Norman is *merely* reflectively lucky.

§3.2. I argue that by understanding BonJour's counterexample as *merely* an instance of reflective luck, we cannot fully account for our comparative intuitions about cases in which one is internally justified and cases in which one lacks internal justification. I consider whether this judgment can be captured by the safety condition which uses the standard ordering of possible worlds (and therefore does not generate global scepticism). I argue that there is scope to understand Norman as unsafe.

§4. I consider the objection that knowledge does not require second-order awareness. I explain how opposing intuitions may be reconciled with the view that knowledge does require internal justification.

Some caution is needed here, since BonJour presents NORMAN as a case of justification failure and as a case of knowledge failure. By contrast, those who appeal to veritic luck considerations do not need to claim that veritic luck is incompatible with *justified belief*. Similarly, the *safety requirement* is not (necessarily) a constraint on epistemic justification, but rather a constraint on knowledge (see Chapter 1, § 3.2). Therefore, in what follows, I will focus on eliciting the intuition that Norman is a case of veritic luck and knowledge failure. I will consider whether the failure to satisfy an internal awareness condition for justification explains this verdict. This will not be equivalent to the claim that having an internally justified true belief is *sufficient* for safety, nor will it be equivalent to the claim that *all kinds* of knowledge require internal justification.

§ 3.1. Doubting the Reflective Luck Diagnosis

The view described in §2 was that Norman is safe and not veritically lucky, but is nevertheless reflectively lucky. I suggest there are at least two reasons to question whether Norman is *merely* reflectively lucky rather than veritically lucky.

First, the claim that NORMAN is *merely* a case of reflective luck is based on the idea that we may conduct third-person epistemic evaluations from *someone else's perspective*; for example, by considering how things appear from Norman's perspective. BonJour is imploring us to consider the question of justification and knowledge *from the subject's perspective*, even though Norman's perspective is unavailable to us. However, at the same time as needing to consider the question of justification and knowledge from the subject's perspective, we are being asked to consider what *our third-person reactions* are to a case which we need to consider *from the perspective of the subject*. We are being asked to do the seemingly impossible – to step into someone else's epistemic shoes while staying in our own. There is no straightforward way of explaining how we go about that.

Second, the reflective luck diagnosis of NORMAN does not offer discriminating reactions for a broad class of cases. The difficulty is this: the claim that Norman is merely reflectively lucky makes NORMAN *trivially* reflectively lucky because, on the non-standard ordering of worlds, *all* our beliefs are reflectively lucky. For Pritchard, this provides us with reason to think that reflective luck is ineliminable and to therefore abandon the requirement that all types of knowledge are incompatible with reflective

luck, even though having a belief that is free from reflective luck would be epistemically desirable. While Pritchard's conclusion is, I suspect, correct if reflective luck is ineliminable, what this conclusion does not fully reveal is the *intuitive difference in epistemic luck of some kind* between NORMAN and putative cases of knowledge which are not subject to veritic luck.

To elaborate, recall here that veritic luck *is* eliminable because it employs a standard ordering of possible worlds and that many of our beliefs are safely held and not veritically lucky. Notice that for putative cases of knowledge such as HANDS (I have hands), such cases will not be veritically lucky even though they are reflectively lucky. This is because on the standard ordering of possible worlds, to the extent that HANDS is a case of knowledge, there are no close possible worlds in which one falsely believes this proposition. 60 However, the standard interpretation of NORMAN holds that Norman is also not veritically lucky. Thus, for both HANDS and NORMAN, we are left with the judgment that neither are veritically lucky but that both are reflectively lucky. The standard view predicts no difference in epistemic luck between NORMAN and HANDS, which runs contrary to our pre-theoretical intuition that there *is* a difference in luck between the two cases.

Internal/External Factors and Safety

I will raise general considerations which allow us to question the supposition that failure to have an awareness of one's reliability does not affect the safety profile of one's belief.

First, I propose in general terms to ask *what* makes a belief safe? It may be stated as part of a specification of an example of belief-formation that the belief *is* safe. However, it is not entirely satisfactory to *state* that a belief is safe, or to build into a specification of a case that a belief is safe (any more than it is satisfactory to state that a case of belief-formation is an example of knowledge, without listing non-epistemic features which would elicit the intuition that it is knowledge). The question is open: there must be some features of the believer's situation which make it *apt* to say that the belief has the modal property of safety. That is, some non-modal features of the situation which elicit the intuition, amongst a wide class of evaluators, that the belief is safe.

Given that safety is a modal condition, the relevant features of a case will be those which affect our intuitions about the ordering of close possible worlds. Pritchard (2005) in his discussion of chicken-sexers, indicates that facts about the reliability of

---

60 Chapter 5, §2, explains why this is so in detail.

one's belief-forming method, or that an agent has an *ability* concerning the truth (where one can possess or exercise that ability without being aware) makes plausible the claim that the beliefs are safe. ₆₁ On this view our modal intuitions about safety are not, in any essential way, fixed by further judgments about what facts the agent is *aware of*, over and above, believing the target proposition. They are instead fixed by externalist facts, such as the use of a reliable belief forming method, or the exercise of an ability which one does not (necessarily) know one has. We may think of this as the view that *safety is fixed exclusively by externalist factors* (I.e. factors which could obtain even if S were not aware that they obtain).

I will question the view that safety is fixed only by externalist factors. I will introduce a general consideration for doubting the view that only externalist factors fix safety.

A general consideration for doubting that safety is fixed solely by external factors is the close association between safety and risk (Pritchard, 2016). I will apply the notion of safety and risk to activities before considering the application to the activity of believing. Imagine that the safety of an activity is proportionate to the degree of risk involved in undertaking that activity, where the activity is dangerous if the degree of risk involved is appreciably high. How might we assess the degree of risk associated with an *activity*? Take the case of a parachute jump. To assess the safety of the jump, we may look to the environmental conditions — for example, the height of the jump and the reliability of the equipment. We presumably also look to precautions taken to eliminate or reduce the risk; we may enquire about a range of facts – including behavioral and psychological dispositions. Our assessment of risk refers to both the objective and subjective features of a situation. To assess the risk of the activity without reference the precautions the participant has taken, offers an incomplete picture: it does not allow us to extrapolate how the overall risk associated with the activity applies to any given participant.

The same moral may be true of belief. Assume that the relative safety of one's belief that p is roughly proportionate to the degree of risk involved in believing that p, where if the risk is high then the belief is unsafe. What then explains the risk associated with the activity of believing? A first attempt at addressing that question is to state that when one is a safe distance from falsity, the risk of falsely believing that p is less than when one is not a safe distance from falsity. However, that account makes the notion of risk solely a function of the safety of one's belief. Making our judgments about risk a

---

₆₁ In comparing naïve chicken-sexer's (who have no awareness of their method of belief formation) with enlightened chicken sexer's (who truly believe that their method of belief formation is olfactory), Pritchard understands both cases as safe and free from veritic luck providing they do in fact have the ability (p. 175).

function of our judgments about safety seems unsatisfactory, for one cannot assess the safety profile of actions independently of understanding the risks. Risk, as Pritchard (2016) points out, has a fundamentality in that one naturally explains judgments about luck in terms of risk (p11).

Some activities are not inherently risky but are inherently dangerous. In such cases, no precautions put one a safe distance from danger. However, the activity of believing is inherently risky rather than inherently dangerous. It carries a risk of falsely believing, but many of our beliefs are safely held. By appreciating that believing is an inherently risky activity, can one take steps to make it safer. For instance, the risk of falsely believing is something we arguably mitigate for by basing our beliefs on good evidence, being aware of the grounds for our beliefs, being able to adduce evidence and being able to revise those beliefs when necessary. Basing one's beliefs on the available evidence, ignoring prejudice and wishful thinking, seem paradigmatic ways in which one reduces the risk of falsely believing. Similarly, being *aware* of what one believes, being *aware* of the fact that one's belief is true and being *aware* of the method which one uses to form a belief, equip one with the ability to carry out those epistemic tasks. If facts of which one is aware, for someone who engages in an activity are relevant to the risk of an activity, then awareness may – equally - be relevant to the risk of the activity of believing. Safety need not be fixed only by externalist conditions because we find that our assessment of risk more generally refers to the subjective features of a situation.

§ 3.2. Norman as Veritically Lucky

Chapter 1 (§ 3) offered a stipulation of (awareness) internalism. The stipulation was that justified belief requires that the believer has a second-order awareness of the non-epistemic properties which their first order belief has, and specifically the fact that a belief is reliably formed or formed in an environment conducive to having true beliefs. In this chapter, I have so far examined the view that cases where one lacks internal justification are merely reflectively lucky and we have seen how this presents a sceptical challenge, one which undermines the argument for awareness internalism which appeals to our anti-luck intuitions. In § 3.1. I considered difficulties with the claim that Norman is merely reflectively lucky. I further considered the idea that our judgments about safety may be sensitive to our judgments about what a believer is aware of.

In light of this, I will consider the possibility that NORMAN is veritically lucky, and whether this can be explained by a failure to have second-order awareness about the non-epistemic properties which his first order belief has. In what follows, I will

assume that doxastic awareness is required for internal justification, i.e. that one must *believe* (either occurrent or non-occurrent) that one's first-order beliefs has the given properties (Chapter 1, § 3), returning to this question in Chapter 3.

The argument for thinking that having a non-veritically lucky true belief requires second-order awareness (i.e. awareness over and above believing the target proposition) concerns our comparative intuitions about counter-part cases. To motivate this, I will adapt BonJour's NORMAN counterexample. In the adapted case, imagine a possible world in which clairvoyance is a faculty possessed not only by Norman but by a small subset of the population, where there is a scientific consensus that clairvoyance is a reliable method of belief formation. Consider now the following two counterparts (across both cases, the externalist facts are held constant while facts about what the agent is aware of are varied):

> ENLIGHTENED NORMAN: Norman, who frequently forms true beliefs about the geographical whereabouts of the President of the United States, one day has a true clairvoyant belief that the President is in Washington. Norman believes that he is using a reliable method of belief formation and he is aware of how the mechanism of clairvoyance operates reliably much like our other perceptual faculties. Norman fastidiously checks to see whether his clairvoyance hunches are correct, on this occasion finds that a reliable news source corroborates his belief that the President is in Washington.

> UNENLIGHTENED NORMAN: Norman, who frequently forms true beliefs about the geographical whereabouts of the President of the United States, one day has a true clairvoyant belief that the President is in Washington). He neither believes that clairvoyance is a reliable method of belief formation, nor does he think that his belief that the President is in Washington is true; nor is he aware of credible media reports which corroborate his belief.

Both ENLIGHTENED NORMAN and UNENLIGHTENED NORMAN have reliably true beliefs – both are externally justified in Goldman's sense. However, both cases differ in what the believers are aware of. ENLIGHTENED NORMAN has a second-order belief that his (first-order) belief that the President is in Washington is true and he further believes that clairvoyance is reliable, and he is aware of a reliable media report corroborating his belief. He satisfies the awareness condition of Chapter 1 (§ 3). By contrast,

UNENLIGHTENED NORMAN has no such second-order awareness. ENLIGHTENED NORMAN is internally justified but his naïve counterpart is not.

The risk of having a false belief seems appreciably higher for UNENLIGHTENED NORMAN, who lacks internal justification, than for his naïve counterpart. This is because UNENLIGHTENED NORMAN is unable to take any precautions against falsely believing the proposition. For instance, if UNENLIGHTENED NORMAN'S evidential situation were to change, if for example he were to acquire evidence *against* the proposition, then it is hard to sense how his belief would be proportioned to his new evidential situation precisely because he does not currently have (or is aware of) any evidence or that his belief is reliably formed.[62] Similarly, if UNENLIGHTENED NORMAN were to learn via reliable testimony that his belief is false, it is hard to judge whether he would revise his belief according to that testimony; when we specify that the believer is unenlightened, then we know nothing about his psychological dispositions which would reasonably allow us to pass such a judgment. The risk of falsely believing seems higher for UNENLIGHTENED NORMAN vs ENLIGHTENED NORMAN.

Note there is no difference in *reflective luck* between the two cases: UNENLIGHTENED NORMAN is no more reflectively lucky than ENLIGHTENED NORMAN. For example, ENLIGHTENED NORMAN'S DECEIVED TWIN in the demon world possesses the same awareness of his belief as ENLIGHTENED NORMAN, even though in the demon world the Deceived Twin lacks reliably true beliefs.[63] Thus, it seems that reflection alone will not allow ENLIGHTENED NORMAN to tell whether or not he is reliable. From the perspective of ENLIGHTENED NORMAN, it is a matter of luck that he does not inhabit a demon world, hence his belief that the President is in Washington is reflectively lucky. [64] We can therefore understand both ENLIGHTENED NORMAN and UNENLIGHTENED NORMAN as reflectively lucky.

A difference in Veritic Luck?

We can, however, observe an intuitive difference in *luck of some kind* between the two cases: ENLIGHTENED NORMAN seems less lucky than UNENLIGHTENED NORMAN. The difference in luck between the two cases cannot be a difference in reflective luck –

---

[62] Where evidence can be understood here not in terms of justification (for example, the probability of a belief being true).

[63] The demon world is a possible world in which an omnipotent demon ensures that Norman has unreliable beliefs but has no way of knowing that his beliefs are unreliable. Importantly, the demon world is phenomenally the same as the non-demon world. Chapter 4. Considers demon scenarios.

[64] I am grateful to an anonymous referee at the Australasian Journal of Philosophy on this point.

because both are reflectively lucky on the non-standard ordering of possible worlds. Similarly, as explained in § 2, the standard treatment of NORMAN is that it is not veritically lucky, in which case UNENLIGHTENED NORMAN will also not be veritically lucky, for he too uses a reliable method of belief formation and is not aware of evidence either against the reliability of the method or the truth of the belief. The standard view does not explain the intuitive difference in epistemic luck *of some kind* between ENLIGHTENED NORMAN and UNENLIGHTENED NORMAN.

The above difficulty is amplified when we consider putative cases of knowledge. For example, if UNENLIGHTENED NORMAN is *merely* reflectively lucky but not veritically lucky, then there is no relevant difference in luck between UNENLIGHTENED NORMAN and putative cases of knowledge possession, such HANDS ("I have hands"), which are also not veritically lucky. Yet, intuitively there is a difference in epistemic luck of some kind between UNENLIGHTENED NORMAN and HANDS. The standard interpretation fails to unify our intuitions about counterpart cases.

By contrast, if both BonJour's NORMAN and the adapted UNENLIGHTENED NORMAN are veritically lucky, then we would be able to explain the pre-theoretical difference in luck between NORMAN, UNENLIGHTENED NORMAN and HANDS. On this view, neither HANDS nor ENLIGHTENED NORMAN would be veritically lucky, but BonJour's NORMAN and our case of UNENLIGHTED NORMAN would be veritically lucky. Notice that since all the cases are examples of reflective luck, a difference in veritic luck would explain the comparative differences in luck I have described, but a difference in reflective luck would not.

A difference in Safety and Veritic Luck?

If there is a difference in veritic luck between ENLIGHTENED NORMAN and his naïve counterpart, then there will need to be some way of that judgment being reflected on the standard ordering of possible worlds the safety condition employs. Let us examine whether comparative differences in awareness may confer differences in safety, that is, whether the one case is *safer* and less veritically lucky than the other.

> **Comparative safety**: for two cases (A & B) of belief that p, A is safer than B just in case the closest (Bp&~p) worlds to α, on a standard ordering of possible worlds, are *further away* from α in A, than the closest (Bp&~p) worlds are from α in B.

Comparative safety seems plausible. Some beliefs are very safely held and free from veritic luck, others are more precariously held, but still be safe and cases of knowledge.

If comparative safety is true, then there may be observable differences in comparative safety between two cases even if both cases are safe: A might be safer than B even though both A and B are safe because the closest (Bp&~p) worlds to α are closer in B than in A, even though in both cases the (Bp&~p) worlds fall within domain of close worlds.

Consider whether ENLIGHTENED NORMAN has a safer belief than naïve counterpart. This judgment - that one case may be safer than another - *concerns comparative safety.*[65] For the cases to reveal a difference in (comparative) safety, where safety is understood modally, we would need to reveal that the comparative judgment is explained by differences in how the possible worlds are ordered between the two cases.

Does a difference in second-order awareness confer a difference in comparative safety between ENLIGHTENED NORMAN and UNENLIGHTENED NORMAN? In considering this, we can imagine possible worlds - for both cases - in the believers falsely believe that the President is in Washington. For instance, there are possible worlds in which the President is playing golf in Aberdeenshire instead of attending to his duties, and in which Norman falsely believes that the President is in Washington. We may suspect that some of these worlds are close for UNENLIGHTENED NORMAN. At least it seems less would need to change – modally speaking - for UNENLIGHTENED NORMAN to falsely believe the President is in Washington, than would need to change for his naïve counterpart to falsely believe the President is in Washington. It seems intuitive to say of ENLIGHTENED NORMAN that he would not easily falsely believe the proposition *to the same extent* as UNENLIGHTENED NORMAN would falsely believe the proposition.

To examine this further, we may bear in mind how internal differences will impact the sphere of relevant close possible worlds. Specifically, there are facts about which internal states hold for ENLIGHTENED NORMAN, but which do not hold for his naïve counterpart. The fact that ENLIGHTENED NORMAN has beliefs about the evidence and the fact he believes that his first-order belief is consistent with the evidence and formed using a reliable method of belief formation. These states hold for ENLIGHTENED NORMAN but do not hold for his naive counterpart. The phenomenal and doxastic states of the enlightened believer may be *non-factive* (it might *seem* to the enlightened believer that his belief is true when in fact it is not). These are internally or phenomenally individuated facts, but they are facts, nonetheless.

---

[65] Comparative safety has been distinguished from other notions, such as relative safety and absolute safety. Absolute safety implies that the risk of harm has been eliminated, whereas relative safety implies that the risk of harm has been reduced or controlled to a tolerable level (See: Moller et al., 2006p. 420).

Consider now the ordering of possible worlds: the closest worlds (beyond the actual world) will be those in which we keep fixed the initial conditions – crucially, holding fixed the facts about which internal and phenomenal states the believers are in for both cases. [66] Notice the closest worlds (beyond the actual world) in which ENLIGHTENED NORMAN believes that p[President is in Washington] are worlds in which Norman has the same doxastic states as in the actual world, and in which he has the same second-order beliefs. These are not *merely* possible worlds in which Norman uses the same method of belief formation that he uses in the actual world. Keeping-fixed these internal facts, for ENLIGHTENED NORMAN the closest (Bp&~p) worlds are one's in which Norman's second-order belief (that his first-order belief is reliably formed) is also false, and they are worlds in which Norman's accessible evidence has misled him (for example, in which the newspaper has printed a typo about the President's whereabouts). Importantly, there are possible worlds for ENLIGHTENED NORMAN in which (Bp&~p) obtains and in which Norman *lacks* second-order awareness, but these are not the *closest* worlds.

By contrast, for UNENLIGHTENED NORMAN, the (Bp&~p) worlds will presumably be closer to actuality, because the closest (Bp&~p) worlds need not be worlds in which Norman's evidence has misled him. There are no further internal facts which need to be held constant when assessing the close worlds for UNENLIGHTENED NORMAN. For UNENLIGHTENED NORMAN, the closest (Bp&~p) to actuality trivially are not worlds in which he has false second-order beliefs or misleading evidence, because UNENLIGHTENED NORMAN lacks any such second-order awareness in the actual world). It is plausible that there is at least a difference in *comparative* safety between the two cases, as one seems *safer* than the other.

Thus far, we have examined how the notion of no difference in either veritic luck or reflective luck between counterpart cases fails to explain our comparative judgment that there *is* a difference in epistemic luck between the two cases. BonJour's NORMAN and the UNENLIGHTED Norman considered in this section are epistemically lucky in a substantive way that does not affect uncontentious propositions that we intuitively know and justifiably believe. Since reflective luck is ineliminable, only if we regard the difference between Norman and uncontentious cases as a difference in veritic luck can we explain that contrast. §3.1 described further difficulties for treating BonJour's counter-example as *merely* a case of reflective luck. This section has indicated how the premise that internalists are targeting veritic luck may be reflected in the standard

---

[66] The importance of keeping the initial conditions fixed when judging modal similarity is emphasized by Pritchard 2015.

ordering of possible worlds, the ordering the safety principle employs to eliminate veritic luck.

However, there is alternative way of understanding the counter-parts that needs examining. The alternative does not entail any difference in veritic luck between ENLIGHTENED NORMAN and his naïve counterpart. In particular, it may be that epistemic intuitions mislead us when comparing counterparts. Pritchard (2005), for instance, does not interpret any difference in *reflective luck* or *veritic luck* between enlightened and unenlightened chicken-sexers. Pritchard compares our verdicts about the chicken sexer who does not know how she forms reliably true beliefs about the sex of hatchling chickens with a chicken-sexer who is aware that she is reliably using her olfactory senses to form true beliefs, and he acknowledges that *it seems* there is a difference in reflective luck between the two cases. However, he contends there is a way in which our ordinary language disguises the fact that there is no difference in epistemic luck. Specifically, when attributing internal justification, Pritchard explains that we implicitly "bracket" sceptical worlds and that this disguises the fact that all our empirical beliefs are reflectively lucky (pp. 208-212). [67] The same may be said of ENLIGHTENED NORMAN and his naïve counterpart. On this view, what *seems* like a difference in luck is explained by the fact that we commonly bracket sceptical scenarios when approaching the cases.

Do we bracket sceptical possibilities when making ordinary attributions? One difficulty is that we are now in a context of considering epistemic luck and find that when we approach counterparts, considering whether the cases are epistemically lucky, it seems there *is* an intuitive difference in luck between counterparts. I suspect a view about the nature of knowledge attributions would not fully reveal why, when confronted with enlightened and unenlightened counterparts, there is still an intuition that the unenlightened believer is luckier than the enlightened believer. There is, pre-theoretically, an intuitive difference in epistemic luck *of some kind* between the counterparts which needs elucidating. According to Pritchard's proposal, however, there is not strictly a difference in luck between the two cases to explain.

Consequently, a difference in veritic luck between enlightened and unenlightened counterparts is not the only explanation of the intuitive difference in epistemic luck between the two cases. It does, however, have the advantage of taking our comparative luck intuitions at face-value; there is a difference in luck which needs

---

[67] My interpretation of Pritchard's account is that it is an error-theoretical approach to our intuitions about the case: it seems there is a difference in luck between the two cases when, in fact, there is not.

explaining and a difference in veritic luck is the most straightforward and parsimonious account of our intuitions.

Qualifying the View and Sources of Resistance

I will summarise the last two sections. In doing so, I identify an important qualification to the view, and I identify why one might be resistant to the interpretation offered in previous sections. The suggestion was that our comparative intuitions indicate that veritic luck and *belief* safety is a function, in part, of one's second order awareness. The account I have offered has the advantage of explaining intuitive differences in luck between counterpart cases.

However, there is an important qualification. Specifically claim that safety is in part a function of one's second-order awareness is not equivalent to the claim that every safe states of affairs requires second-order awareness. For safety is a modal state which holds in non-doxastic contexts. For example, an aircraft may be safe if the situation in which it dangerously malfunctions is sufficiently remote: no reference to the safety of a subject's belief is implicit, because we are not examining safety in a doxastic context. However, in the epistemic case, the context of assessing the modal profile of a belief, safety does depend on at least one internal condition – namely that S believes that p. If S does not at least believe that p, then trivially S does not safely believe that p. Thus, the claim is that facts about second-order awareness of non-epistemic facts, such as awareness that one is using a reliable method, are contextually relevant to the safety of a belief.

Notwithstanding, one may still be resistant to the proposal. One source of resistance to the proposal is the high likelihood that Norman's beliefs are true. Note that whenever Norman uses the method of clairvoyance it is highly unlikely that he would falsely believe the target proposition. This fact may lead us suppose that his beliefs are epistemically safe and free from veritic luck. However, for the safety subjunctive (Bp□→p), the closeness of worlds in which one falsely believes that p (or "Bp&~p" worlds, for short) will not hinge only on the probability of the consequent "p" obtaining conditional on the antecedent "Bp" obtaining. The LOTTERY case recommends against ordering the closeness of possible worlds probabilistically. Consider someone S* who truly believes that [I own a losing lottery ticket] where the lottery result has been announced but where she has not yet checked the result. Despite the probability of her belief being true, her belief seems epistemically unsafe – because the world in which she falsely believes that same proposition is very much like the actual world. On this view, the mere high probability of the truth of a belief does not make it

safe. By bracketing probabilistic considerations, we may overcome a source of resistance to the present proposal.

It remains possible, however, that all the counterparts described so far are absolutely safe. This is because two cases may differ in <u>comparative safety</u>, even if both cases are absolutely safe: this would happen if for both cases there are no close (Bp&~p) worlds, such that they are absolutely safe, but for the one case, the closest (Bp&~p) are further from actuality than for the other case. In both cases, the beliefs would be safe because the beliefs would be true at all or most close worlds.

Still, a difference in comparative safety will have a non-trivial implication. For if second-order awareness contributes to the safety of a belief, then UNENLIGHTENED NORMAN is not as safe as it could be. If it is preferable for a belief to be safer rather than less safe, then it is better from an epistemic standpoint to have second-order awareness than to lack second-order awareness. Thus, comparative differences in safety would suggest that the kind of epistemic luck which awareness internalism targets is not merely reflective luck. I think this is crucial because it supports the claim that internalists need not only be interested in eliminating reflective luck, and the proposal offers a response to the Reflective Luck Argument which took as its starting point the claim that awareness internalism *merely* targets reflective luck.

Third, even if both cases are understood as absolutely safe, despite differences in comparative safety, it does not necessarily follow that Norman is not veritically lucky. Recall that it was argued that UNENLIGHTENED NORMAN is, first and foremost, a case of veritic luck, and that that judgment is supported by our comparative intuitions about luck, and the need to differentiate the way UNENLIGHTENED NORMAN is lucky from putative cases of knowledge (which are veritically lucky but not reflectively lucky). However, having a safe belief does not entail that the belief is not veritically lucky - to the extent that safety is a necessary and rather than condition for being free from veritic luck, a claim which is stronger than what is needed to explain a wide class of other cases such as STOPPED-CLOCK.

Chapter 5 will defend the idea that safety is a necessary anti-veritic luck condition. I will support a weaker notion of safety which does not purport to explain every case of veritic luck; avoiding the claim that having a belief free from veritic luck *just is* being epistemically safe.

§ 4. Objections and Responses

The account developed in Chapter 2, so far, encounters at least two objections. The objections are as follows:

4.1. Internally justified beliefs can be unsafe and therefore veritically lucky.

4.2. Subjects may know or justifiably believe propositions without satisfying an awareness condition for justified belief.

This section will be structured around addressing those two objections. I will argue that by appealing to Stephen Hetherington's (2001) distinction between minimal and non-minimal ways of knowing, awareness internalism has the resources to accommodate a class of opposing intuitions about knowledge and justification.

§ 4.1. Internally Justified but Unsafe Beliefs

The account of justification I have been examining is awareness internalism. I have defined this as the view that one needs awareness of externalist conditions, and minimally that one needs to be aware of the fact that one is reliably formed. A more specific way of stating that latter requirement is that one's belief is justified only if one's token *B* results from a reliable method M and one believes (occurrently or non-occurrent) that B is the result of M (Chapter 1, § 3) I have argued that failure to satisfy that condition can explain veritic luck (for a class of cases), and can do so consistently with the safety principle.

However, a belief may satisfy the internal condition and yet still be unsafe and therefore veritically lucky. The reason for this is because a reliable method need not be 100% reliable to yield reliably true beliefs. As such, for any true belief (that p) via method M (where M is an imperfectly reliable method), there will be at least one close possible world in which one falsely believes that p via M.[68] Suppose, further, that S truly believes that p via M and S believes that M is reliable and that belief-tokens resulting from M are likely to be true (S is internally justified). If M is imperfectly reliable, then S may have an internally justified but unsafe belief.

For example, consider the STOPPED-CLOCK case, where S forms a luckily true belief that the time is 12.00 by *looking* at the clock face. Here, if we construe methods

---

[68] I am indebted to an anonymous reviewer for helping to clarify this.

broadly, then *sense perception* – the method used by S - may be individuated as a "method" of belief-formation (a point considered in Chapter 5, § 1). The method may function as a reliable method, albeit it an imperfectly reliable method. Notice that we can imagine in STOPPED-CLOCK that S uses the method of sense perception and is internally justified (with some further amendments, we can imagine that the subject *truly believes* that he has formed the belief via a reliable method). However, in STOPPED-CLOCK, the belief is unsafe (there is at least one close world in which S falsely believes) and therefore veritically lucky. Therefore, holding a belief which is internally justified, does not imply that one's belief will be safe and free from veritic luck.

A response to this worry is to re-affirm that internal justification need not be a *sufficient* condition for having a safe belief. Recall that what was problematic about NORMAN was that he failed to have any second-order awareness about the credentials of his first order belief. The claim to make about NORMAN is that he violates a necessary condition for internal justification. We need not, therefore, claim that internal justification is *sufficient* safety. The proposal allows that other features of an agent's doxastic situation are relevant to the safety of a belief; this may include, for example: whether one's belief is formed in a normal environment, the fact that one's belief results from truth-conducive dispositions, or the fact that there is evidence for the proposition.

However, a difficulty for the above response is as follows. If having an internally justified belief is *necessary rather than sufficient* for safely believing, then how do we understand the relationship between that second-order awareness and having a justified belief that p? In particular, would the other non-epistemic facts which are relevant to our judgments about epistemic luck, for any particular case, be factors which are relevant to the *justification* of that belief? We have two options here:

i) The non-epistemic facts which determine safety-failure in STOPPED-CLOCK (and cases like it) *are* facts about justification.

ii) The non-epistemic facts which determine safety-failure STOPPED-CLOCK (and cases like it) *are not* facts about justification.

Opting for (i) commits us to the verdict that STOPPED-CLOCK is unjustified on account of STOPPED-CLOCK being unsafe. Conversely, opting for (ii) means that the facts which explain why STOPPED CLOCK is unsafe are irrelevant to justification. Option (ii) leaves it hard to see why epistemic luck *alone* motivates awareness internalism. This is because we can cite cases such as STOPPED-CLOCK in which one *has* justification even though one's belief is unsafe and therefore veritically lucky.

I take it that option (i) is counterintuitive. For most of would be included to say that STOPPED-CLOCK is a case of justification even though it is not a case of knowledge. The agent uses a generally reliable method when methods are construed broadly and is unaware of any evidence against his belief or the reliability of the method.[69] On that view, not every non-epistemic fact which is relevant to our judgment about safety will be relevant to our judgment about justification. However, if we opt for (ii) – i.e. allowing that one can have justified but unsafe beliefs - then one worry is that anti-luck considerations *alone* do not show that we *need* an internalist notion of justification.

I suspect the above worry rests on confusing the direction of dependence between internal justification and safety. Cases such as NORMAN show (for a class of cases), I think, that a belief needs to be internally justified to be free from (veritic) epistemic luckand that without internal justification in such cases, one's beliefs will not be free from veritic luck. [70] Furthermore, we can explain this through the safety condition. The moral is that we should not want to count certain cases (such as NORMAN) as examples of knowledge because they are unsafe and that they are unsafe (at least in part) because they are not internally justified. However, this is compatible with some internally justified beliefs being unsafe and veritically lucky (for example, in as the STOPPED-CLOCK example attest). On the current account, it is being safe which depends (at least for a class of cases) on having an internally justified belief, not vice-versa.

§ 4.2. Knowing Without Awareness

There is an objection to the account I have offered. The objection is this: one may *know* and safely believe a proposition *"p"* even if one's belief that p is not internally justified. In this section I will introduce some cases as illustrative. I will then respond by explaining how the account I have offered may be reconciled with the possibility of one can *minimally know* a proposition in cases where is not internally justified.

First, recall the condition:

> **Internal (justification) condition:** S has a justified belief B *only if*
> S has used a reliable method of belief formation in arriving at B and S is
> aware that she has used a reliable method of belief formation in arriving
> at B.

---

[69] I consider the issue of method individuation in Chapter 5, and the reasons for a broad construal of methods. As an indication of what will be argued, the idea is that one needs a broad enough construal so that methods can function in subjunctive conditionals. See Chapter 5 § 1
[70] I consider this in Chapter6

Let us further imagine that the mode of awareness is *doxastic*, and that to satisfy the internal condition a believer must at least *believe* (occurrently or non-occurrently) that her belief is reliably formed.

If the internal condition requires *believing* that one's belief is reliably formed, then the proposal is vulnerable to counter-examples; specifically, cases of knowledge which do not satisfy the internal condition. I will provide the following example where the internal condition is violated but where knowledge *may* survive.

> BILL. Bill walks into a room and *sees* a red wall and believes that the wall is red, but Bill is later told (by a generally reliable informant) that the wall was lit up with a red light which would make any surface look red. However, Bill *saw* that it was red. Furthermore, Bill maintains his belief that the wall is red because Bill *remembers* that it was red. However, having heard the testimony, Bill does not believe that his belief was reliably formed - Bill is quite ambivalent about whether or not he was deceived. Intuitively Bill knows that the wall is red because he remembers it was red and still believes that it was red, but Bill is not internally justified. By ignoring the testimony and holding the belief, as Bill does, Bill lacks internal justification and still knows.

I want to concede that BILL knows but that he fails to have internal justification. The testimony BILL received defeats his justification because it leads him to suspend judgment about the reliability of his belief-forming method. Similar cases exist, where one is not sure that one's belief is formed via a reliable method.[71]

One concern about BILL is that he may have a defeater – a reason to think that the wall is *not red*. However, this does not appear to be the case. For example, Bill has not been informed that the wall is "white but illuminated with red light" – the informant does not make such a claim. Indeed, it seems that Bill loses his justification, without acquiring defeater, but still retains his knowledge.[72]

---

[71] There are other cases (for example, Colin Radford, 1966), where subjects know a proposition even though they are unsure of the truth of the proposition and do not believe that they know it. Interestingly, in one of the examples which Radford discusses (see pp. 5-6), Radford states that such a person has knowledge which is a "a *poor* thing, sparse, uncertain, unwitting, and therefore *unimpressive* and of little use" (pp. 5-6. My italics). Though Radford does not present this in terms of the gradualist response that I sympathize with in § 4.2 (namely, Hetherington 2001 proposal), I suspect Radford's verdict about this case (that it's an unsatisfactory item of knowledge) is not dissimilar to Hetherington's because Radford highlights a poor form of knowledge – that we may think of some knowledge as better or worse.

[72] A further worry is that clauses such as "remembers that", where *remembers that* is a factive state which is explained in terms of *knowing that* (Following Williamson, 2000). To avoid this,

Is BILL safe? I am unsure that Bill is safe, or at least, that he is as safe as he would be, had he not lost his justification. It seems *quite easily the case* that Bill *could* have had a false belief. This is because the situation in which the wall is red but illuminated by white light is one in which Bill still thinks that he remembers the wall being red; and in this case Bill falsely believes that it is red. Given the generally reliable testimony of his friend, then the possible worlds in which Bill falsely believes are relatively close. In this way, justification failure may explain the failure of his belief to be as safe, or as safe as it would have been had he not lost his internal justification. However, as indicated, I have avoided the stronger claim that safety requires internal justification in all cases.[73]

Other cases abound. For example, one may cite reliable chicken-sexer's, who can identify the sex of a hatchling chicken without knowing their method, as examples of safe belief (Pritchard; 2005) or as examples of knowledge without justification (Lewis, 1996).[74] Crucially, one could also reject the claim that NORMAN fails to know. For example, it might be said that if BILL is a case of knowledge, then we have no reason not to count NORMAN as a case of knowledge.

The intuitions I have described cast doubt on the *necessity* of internal justification for knowledge. Indeed, they deny the very epistemic intuition about *knowledge failure* which I have – so far – relied on.

§ 4.3. A Non-Absolutist Response

How to respond to such cases? The class of cases is broad, including BILL, NORMAN, CHICK, ANIMALS. Etc. The intuition I wish to accommodate is that they are cases of k*nowledge* without internal justification. I want to explain how this intuition may be reconciled with the claim that knowledge requires internal justification. A denial that they are cases of knowledge *in any sense* would be dogmatic. Conversely, a non-dogmatic response would explain how they are *in some sense* cases of knowledge, but in a way consistent with the intuition that in *some other sense* they are cases of knowledge failure.

The response I will issue is that there is a *type* of knowledge which requires internal justification. However, this does not imply that *every* type of knowledge requires internal justification. There may be other, normatively far from ideal, senses in which BILL, NORMAN, and other cases still *know* despite lacking internal justification.

we might want to amend the case as involving "Bill still remembering the red wall". My thanks to Rory Madden for clarifying this.
[73] I have avoided the stronger claim *that all cases* of safe belief require internal justification.
[74] See Lewis (1996, p.551). Lewis's claim arises in the context of his contextualist project.

In support of this, let us consider Stephen Hetherington's view in his book *"Good Knowledge, Bad Knowledge"* (2001). Stephen Hetherington construes knowledge on a sliding-scale, where knowledge of a fact can be better or worse, depending on the strength of the believer's justification. [75] The thesis that one's knowledge of a fact can be better or worse is termed *"non-absolutism"*. By contrast, "absolutism" is the view that "it is impossible for a person to have better, or to have worse, knowledge of a fact" (p. 3). Non-absolutism is the rejection of absolutism.

There are two components of Hetherington's position, which are essential for our present dialectic:

(**1**) Knowledge of a fact can be either minimal or non-minimal

(**2**) The quality of one's non-minimal knowledge depends on the quality of one's justification (Hetherington calls this position epistemic gradualism).

I shall suggest that on a certain construal of minimal knowledge, the distinction between minimal and non-minimal knowledge can help meet the intuition that BILL, NORMAN, CHICK, and other cases, are instances of knowledge. Finally, I contend that the response need not commit us to the further element (**2**) of Hetherington's methodology.

Non-absolutism is the rejection of the view that knowledge of a fact cannot be better or worse. By contrast, absolutism is the view that, even if justification comes in degrees, knowledge does not. However, as Hetherington explains, non-absolutism alone is compatible with a justificatory condition for all grades of knowledge, and so is compatible, in principle, with the idea that all knowledge needs to be internally justified. For instance, a non-absolutist view could hold that for a belief B to constitute knowledge, B must meet a threshold for knowledge, where that threshold requires that B is *justified*. Here, the non-absolutist would maintain that, once B meets the knowledge threshold, then the further justification for B would increase the quality of B as a case of knowledge. On that view, one might have better or worse knowledge of a fact depending on one's justification. [76]

NORMAN and BILL lack *any* justification (at least in the internalist sense), and so appealing to non-absolutism *alone* will not explain the opposing intuition that they

---

[75] On this view, justification is understood as the gradational component of knowledge and as the component which makes knowledge gradational. Cases in which one has poor justification are thus cases in which one has poor knowledge on Hetherington's view.

[76] On this view, for any two counterparts X and Y (where both have knowledge that p) but where X is more justified than Y, then X will have *better* knowledge that p than Y.

are cases of knowledge. This is because it is compatible with the non-absolutist view that there is a minimal justificatory threshold which both subjects fail to meet.

Applying minimal knowledge

However, Hetherington appeals to a category of minimal knowledge, claiming that some knowledge is "not good at all" and describes cases where one *barely knows*, and in these cases we "grudgingly" ascribe knowledge (p. 133).[77] Hetherington regards true beliefs without justification as "minimal knowledge" in which one barely knows that p. For Hetherington, minimal knowledge represents the worst possible type of knowledge. *Mere true belief*, on this view, is sufficient for minimal knowledge, which is knowledge of the worst possible kind (a view he calls *epistemic minimalism*). Schematically,

> Minimal knowledge: S has minimal knowledge that p just in case S has a true belief that p which lacks justification.[78]

Conversely, if one has non-minimal knowledge (i.e. knowledge with justification), then the quality of that knowledge – on Hetherington's view - would automatically be of a higher quality than minimal knowledge (p. 146). Furthermore, on Hetherington's view, the overall quality of that non-minimal knowledge will be proportionate to the strength of one's justification. [79]

First, consider how Hetherington's notion of minimal knowledge – as mere true belief – would meet BILL, NORMAN, and so on. The response would hold that true belief

[77] Hetherington writes that "some knowledge is not good at all; if you believe p without having any justification for it, yet your belief is true, then – albeit grudgingly – we should concede your knowledge that p. You would barely know that p; to have minimal knowledge is barely to know. Nevertheless, barely to win a race is still to win it."(p. 133), in supporting this, Hetherington claims that a true belief still meets its "primary epistemic goal" of truth (Ibid.,)
[78] Hetherington defines epistemic minimalism as follows "For any epistemic subject x, time t, and proposition p: (1) At t, x has minimal knowledge that p =df At t, x has an unjustified true belief that p; (2) at t, if x does have justification for the true belief that p, then (other things being equal) x's true belief that p is better knowledge that p to the extent that x's justification is good justification. (The better the justification for p, the better the knowledge that p – the better the true belief is as knowledge that p)." (Hetherington, 2001, p. 132). As Hetherington observes, non-absolutism does not entail *epistemic minimalism*. For one might argue that the minimum knowledge-threshold amounts to *more* than mere true belief, also requiring that one's belief is formed using a reliable method of belief formation, but that the quality of one's knowledge can be increased depending on whether one is justified and the strength of one's justification (for example, the degree of one's reliability).
[79] Hetherington describes true belief as "the first step" in knowledge, writing that "although the first step in the spectrum of knowledge does not include justification (since the step is mere true belief), each subsequent stage does include some degree of justification, each step represents a possible quality of knowledge that p – and thus some one or more way(s) to know that p more or less well." (p. 146). Hetherington argues that any justification for a true belief makes a case of knowledge much better than it would were it merely true belief (p. 148).

is a type of minimal knowledge, and that internally justified safe belief represents non-minimal knowledge. For all the cases I have described, they lack non-minimal knowledge (I.e. knowledge with justification) but still have minimal knowledge (knowledge without justification).

The proposal is, I think, consistent with our normative intuitions. For example, when we consider counterpart cases, intuitively it seems there is something *problematic* about Norman's epistemic situation, even if we think he still *knows*. For example, there is an intuition that we would rather be in the situation of the enlightened believer, who believes or knows that he is reliable, has evidence for the belief than in the situation of the unenlightened counterexample. Intuitively, there is a normative difference between the cases even if we think that both cases are examples of knowledge. Any knowledge which the unenlightened believers have seems far from *ideal*. By distinguishing between minimal and non-minimal knowledge, we can explain that contrast.

To see this more clearly, we may understand the dialectic as follows (for now, let us focus just on NORMAN). There are *knowledge-deniers* about NORMAN and are (or may be) knowledge-affirmers about NORMAN. If we distinguish non-minimal knowledge from minimal knowledge, then both claims would be partially correct. However, even though each party would be correct, they would, respectively, be correct about a different grade of knowledge. More specifically, the *knowledge affirmers* would be correct that NORMAN has low-grade (minimal) knowledge, and the *knowledge deniers* would be correct that NORMAN lacks higher-grade (non-minimal) knowledge. If this view is correct, then the claim that knowledge requires internal justification is restricted as follows: an epistemically desirable kind of knowledge requires internalist justification.

However, I think there are difficulties with minimal knowledge, as it is currently presented. First, I outline the intuition that reliability may still be necessary for knowledge. Second, I outline the need to accommodate the intuition that NORMAN does not merely *know* but moreover has *justified* beliefs.

Do we need reliability?

One worry is that *mere true* is not always sufficient for knowledge. Thus, if we apply minimal knowledge to our present cases (NORMAN, CHICK), then the explanation for why those cases are examples of (minimal) knowledge is that they are *merely cases* of true belief. This is a problem if we doubt that *mere true belief* is ever sufficient for

knowledge (even for knowledge of the worst possible kind). To understand the difficulty, consider the following example:

> JUROR: A juror concludes that the defendant committed theft, but the Juror has paid no attention to the defense barrister's case. Instead, the Juror has been conducting internet research into the defendant's prior convictions, searching through the press-clippings about the defendant's past crimes (the juror perverts the course of justice, but does not get caught). Imagine that, as it happens, the defendant *is* guilty. The juror has a true belief, but little else. We would have little inclination to think that the juror knows the defendant's guilt, not even knowledge of the worst kind.

It seems, to me, that the Juror does not know the defendant's guilt. More generally, that knowledge may frequently be sub-optimal does not imply that any true belief - however, disastrously formed - constitutes a knowledge. The believer is highly blameworthy, and this seems enough to defeat knowledge.[80]

To accommodate JUROR, I suggest we depart from Hetherington's notion of minimal knowledge as mere true belief. By doing this, I suspect there is a way of accommodating JUROR while preserving the view that NORMAN, CHICK are instances of knowledge (but of the worst kind). In particular, note the differences in the *reliability* of the method of belief formation across the cases. Principally, the juror is not using a reliable method of belief formation – she relies on press-stories about the defendant's prior convictions, rather than representations made by the defense barrister. Hence it seems the Juror has a true belief but that she cannot be said to have knowledge, not even knowledge of the worst kind. By contrast, the case of NORMAN, (and CHICK etc) involve a subject using a reliable method of belief formation, even though the belief is unjustified. Norman is like our Juror in that Norman fails to possess second-order awareness but is unlike Juror in that Norman uses a reliable method of belief formation. Since there is a difference in the reliability of the method of belief formation used in NORMAN compared to JUROR, there is a way of explaining why NORMAN is a case of minimal knowledge. Specifically, Norman has minimal knowledge because he uses a reliable method of belief formation, yet Norman lacks *non-minimal knowledge* because he lacks internal justification. More formally:

---

[80] The claim that *mere true belief* does not amount to knowledge is dialectically benign: if Hetherington is correct that mere true belief is the most minimal way of knowing, then there may not be an epistemic difference between JUROR and NORMAN that needs explaining.

> Minimal knowledge (with reliability) S has minimal knowledge that p
> just in case S has a true belief that p via a reliable method.

Adding a reliability condition to an account of minimal knowledge is important – if being reliably true is the most minimal way in which one can know a proposition. Adding reliability to minimal knowledge at least excludes JUROR from the extension of knowledge.

Explaining Justification

The second problem is that, insofar as I have adopted Hetherington's definition of *minimal knowledge* as a type of knowledge *without justification*, I have not allowed differences in reliability to confer differences in justification. This is a problem, insofar as it may seem that NORMAN (and CHICK) holds a *justified belief because the belief is reliably f*ormed. In short, one can deny the intuition that cases such as NORMAN and are unjustified.

Therefore, while an account of minimal knowledge in terms of *reliably true belief* accommodates the intuition that NORMAN is a case of knowledge, it still involves denying an intuition about justification. The problem is that we should need to accommodate opposing intuitions about knowledge *and* intuitions about justification for NORMAN. More specifically, we should need to accommodate opposing intuitions about justification for cases which do not satisfy an internalist condition for justified belief.

To accommodate this, note that it is compatible with non-absolutism (the idea that knowledge of a fact can be better or worse) that the reliability condition constitutes a *justificatory condition*. For example, while the best kind of (non-minimal) knowledge requires internal justification, one may have a lower grade of knowledge if one has a true belief which is reliably formed. Such knowledge may be *externally* justified because disapprobation seems inappropriate and because we can at least say something positive about the belief: namely, it is reliably formed. However, – normatively speaking – our comparative intuitions about enlightened counterparts suggest that the knowledge is far from ideal because it is not internally justified. On that view, minimal knowledge denotes *unsatisfactory* knowledge, because it is externally justified but not internally justified, rather than the egregious knowledge which Hetherington characterizes as mere true belief.

This revised account of minimal knowledge would hold that having a reliably true belief is a way of knowing and justifiably believing a proposition, though knowledge was has unsatisfactory (rather than shockingly bad), and a sort of knowledge

which – as I have argued in this chapter – which carries a risk of falsely believing that p.

The response captures, I think, the normative contrasts we make between enlightened and unenlightened counterparts. In particular, counterparts may both be instances of knowledge and justified belief, being enlightened (and internally justified) confers, intuitively, a normative difference. The difference is, I think, captured by the non-absolutist approach because such a person has, according to that view, a higher grade of knowledge because he has a higher grade of justification; in our case, he has internalist justification.

In Chapter 6, I will support the idea that there is more than one kind of knowledge as a response to how animals can have a robust kind of knowledge, despite failing to be internally epistemically justified. This will be the view that are *different kinds of propositional knowledge*. By contrast, Hetherington's gradualism is the view that the quality of one's knowledge exists on a sliding scale and is a function of the quality of one's justification.

Hetherington's view will be consistent with the position later defended, in the sense that Hetherington thinks there are both minimal and non-minimal ways of knowing a fact (i.e. his rejection of epistemic absolutism), and "gradualist" insofar as he holds that non-minimal knowledge of a fact can be better or worse. By contrast, I suspect one can endorse the view that there are *different kinds of knowledge* without gradualism. For instance, one might hold that there is minimal and non-minimal knowledge, but that non-minimal knowledge is normatively satisfactory knowledge and non-minimal knowledge is unsatisfactory knowledge, without holding that the strength of one's justification will further affect the quality of one's non-minimal knowledge. Such a view would hold that if one has non-minimal knowledge of a fact, then counterfactual situations in which one has a higher degree of internal justification (for example, a greater degree of awareness) need not be counterfactuals in which one has a higher quality of non-minimal knowledge. Notice that the claim that knowledge can be better or worse (the claim about non-absolutism) is still retained since one will have better knowledge if it is internally justified (i.e. one has non-minimal knowledge of a fact) than if one knows without internal justification.

The cases I have considered in this section are *consistent* with the sliding-scale conception of knowledge, in that a slid-scale conception would explain the cases, but I suspect the cases neither entail the sliding-scale account nor directly support it. The reason for this is because I think we need only the premise that some cases of knowledge are normatively much better than other cases of knowledge to explain the opposing intuitions about the cases, and that they are plausibly better because they are internally

justified. The comparative normative intuitions I have considered, I suspect, are consistent with this – they support the idea that enlightenment makes a normative difference and a difference to one's knowledge, even if there is still a residual intuition that one can know and justifiably believe, without satisfying an internalist standard for justification.

Chapter 3

Internal Awareness and Regress Scepticism

I have, so far, understood awareness internalism as the view that a belief's justification depends on the believer's awareness of externalist states of affairs that are necessary for the belief to be justified. More specifically, the suggestion was that reliability is necessary but insufficient for justified belief. I examined counterexamples to externalism (such as BonJour's) and used them to motivate awareness internalism. Chapter 1 conceded that there might be further, normatively impoverished ways in which a subject can know a proposition without having to have an internally justified belief. However, if normatively enriched knowledge requires an internally justified belief, then the requirements for justified belief will need to be requirements that believer can satisfy. If the standard for justification cannot be satisfied by actual believers, then awareness internalism will imply scepticism. If awareness internalism implies scepticism, then there is an intuitive objection to awareness internalism.

This Chapter considers a sceptical objection to awareness internalism. The objection is that awareness internalism either i) commits us to regress scepticism or ii) avoids regress scepticism but at the cost of not accommodating the anti-luck intuitions that motivate awareness internalism. The argument I will examine (Bermann's, 2006) is that awareness internalism is committed to either (i) or (ii), and that we do not therefore have grounds to support awareness internalism over justificatory externalism. In this chapter, I will argue that we can accommodate Bermann's objection by endorsing a *non-epistemic doxastic awareness requirement* instead of an *epistemic doxastic awareness requirement*. I then consider an objection to the *non-epistemic doxastic awareness requirement*: namely the *non-epistemic doxastic requirement* is too permissive because it allows intuitively unjustified beliefs to count as justified. I argue that there is scope to meet Bergmann's objection to awareness internalism.

Finally, I consider how the regress problem poses a meta-epistemological challenge, one which is not diffused by an awareness condition for justified belief. I explain the significance of this for both awareness internalism and justificatory externalism.

<u>This chapter is structured as follows:</u>

§1. I distinguish between epistemic and non-epistemic doxastic awareness requirements for justified belief. I then consider motivations for endorsing a doxastic awareness requirement for justified belief.

§2. I explain Bergmann's argument that strong (epistemic) doxastic requirements generate a sceptical regress. I then explain his argument that non-doxastic requirements avoid scepticism but at the price of not providing a reason to support awareness internalism over an externalist account of justification. I explain Bergmann's contention that non-doxastic requirements abandon an anti-luck motivation for internalism, which Bergmann identifies as the Subject Perspective Objection to externalism. I explain why Bergmann finds non-epistemic versions of awareness unsatisfactory.

§3. I argue that a non-epistemic doxastic awareness can avoid the sceptical regress while meeting the Subject Perspective Objection. However, I argue what the awareness requirement needs to incorporate an externalist condition on first-order justification.

§4. I consider a meta-sceptical argument that is unresolved by awareness internalism and I reflect on the implications of this for the internalism/externalism debate.

Preliminary Remarks

Before developing the view announced at the start of this chapter, recall from Chapter 1 (§ 3) the awareness condition:

> **(AI) Awareness internalism (with an externalist condition):**
> Subject S has a justified belief B only if: 1) B has some property Φ AND S is aware that B has Φ (for a Φ that may hold for B in circumstances where S is unaware that B has Φ).[81]

AI is neutral on which non-epistemic property/properties "Φ" a believer would need to be aware of to hold a justified belief. For example, **AI** permits that Φ is some internal state, such as a mental state, or a fact about one's phenomenal experiences,

---

[81] In circumstances where B has Φ, but S is not aware that B has Φ, the principle predicts that B will not be epistemically justified.

such as the fact it *seems* or *appears* to be the case that p. However, the argument in this chapter supports an awareness condition incorporating an externalist reliability condition; more specifically, the fact that one's belief is formed using a reliable method of belief formation. By considering Bergmann's argument, the present chapter aims to test the hypothesis that reliability is needed for justified belief. (Chapter 4 will further test that hypothesis).

§ 1. Doxastic Awareness Requirements: epistemic and non-epistemic

I wish to first distinguish between epistemic and non-epistemic doxastic awareness requirements. In general, a doxastic awareness requirement is a second-order constraint on whether one's first-order belief is justified. It says, roughly, that the justification of one's belief B is dependent on S having a belief that B has a property φ that is necessary for B's justification.

We may distinguish between two types of doxastic awareness requirements. first, an epistemic doxastic requirement holds that one's second-order belief (that B has some φ) must be justified. By contrast, a non-epistemic doxastic requirement does not require that one's second-order belief must further be justified. Let us present the distinction between epistemic and non-epistemic doxastic requirements as follows:[82]

> Epistemic doxastic awareness requirement: S has a justified belief B only if: B has some property Φ and S justifiably believes (either occurrently or non-occurrently) that B has Φ.

> Non-epistemic doxastic awareness requirement: S has a justified belief B only if: B has some property Φ and S believes (either occurrently or non-occurrently) that B has Φ.

On the assumption that knowledge requires belief, satisfying the epistemic (doxastic awareness) requirement entails satisfying the non-epistemic requirement but not vice-versa. For if S satisfies the epistemic requirement because her belief B has φ and s justifiably believes that B has φ, then she trivially satisfies the non-epistemic requirement in virtue of believing that B has φ.

---

[82] For now, let us be neutral on whether Φ is an internalist or externalist state of affairs.

By contrast, the non-epistemic requirement is more permissive than the epistemic requirement: for cases in which B has φ and in which S *believes but without justification* that B has φ, satisfy the non-epistemic requirement while failing to satisfy the epistemic requirement. The epistemic requirement is strictly stronger (and less permissive) than the non-epistemic requirement.

Let us consider motivations for imposing a doxastic awareness requirement of either kind (whether epistemic or non-epistemic) on the justification of one's first-order beliefs. Specifically, I wish to consider grounds for thinking that the justification of our first-order beliefs depends on further, second-order beliefs. Here, recall BonJour's case of NORMAN:

> NORMAN**:** "Norman, under certain conditions which usually obtain, is a completely reliable clairvoyant with respect to certain kinds of subject matter. He possesses no evidence or reasons of any kind for or against the general possibility of such a cognitive power or for or against the thesis that he possesses it. One day Norman comes to believe that the president is in New York City, though he has no evidence either for or against this belief. In fact the belief is true and results from his clairvoyant power under circumstances in which it is completely reliable." (BonJour, *Structure of Empirical Knowledge*, p41).

An intuition suggests that Norman has an unjustified belief, or at least, Norman's situation is normatively impoverished (Chapter 2, § 4). However, an externalist theory that makes reliability necessary and sufficient for justification (Goldman, 1979) would fail to explain the case of NORMAN. The case challenges the sufficiency of externalist conditions, rather than the necessity of externalist conditions. So, Norman's belief has something going for it - epistemically speaking - and perhaps it meets a necessary condition for justified belief, but it still falls short of being justified (I will bracket that consideration for now, returning to it in § 3).

What is needed to explain cases such as NORMAN, on BonJour's view, is an internalist condition for justification. Specifically, BonJour argues that for a belief to be justified, the believer must at least believe with justification that B has a high likelihood of being true (ibid, p. 32).[83] When we apply this to NORMAN, what is notable about NORMAN is that he fails to satisfy that condition. For example, if Φ denotes the use of a reliable method, or the fact that the belief has a high likelihood of being true, Norman does not believe that B has Φ nor does he believe with justification that B has

---

[83] BonJour's specific claim (SEK, p. 32) is that B can be justified only if one "believes with justification the premises from which it follows that the belief is likely to be true".

Φ. A second-order awareness condition for justified belief thus explains the verdict that NORMAN is unjustified.

To see this further, imagine an enlightened counterpart version of Norman, ENLIGHTENED NORMAN, who - unlike his naïve counterpart - is aware of how clairvoyance operates and believes that his clairvoyant hunch on any given occasion results from his reliable faculty. It is no longer as intuitive that ENLIGHTENED NORMAN is unjustified. ENLIGHTENED NORMAN seems justified and seems to satisfy a second-order, doxastic requirement for justified belief. Thus, an awareness condition for justified belief may explain our comparative intuitions for the cases under consideration.

§ 2. Regress Scepticism and Bergmann's Subject Perspective Objection

BonJour's (1985) condition for justified belief is not merely doxastic but is furthermore epistemic. That is to say, BonJour requires that not only must S believe that B has Φ for B to be justified, but S must moreover believe with justification that B has Φ (ibid, p. 32).

However, epistemic requirements lead inexorably to scepticism. The reason for this is because, on this view, to have a justified belief, one must justifiably believe that one's belief has Φ. That is to say, one must have a justified second-order belief (B2) that one's first-order belief (B1) has Φ. In order for B2 to be justified, B2 must have some property Φ and one must have a further justified belief (B3) that B2 has Φ. In order for B3 to be a justified belief, B3 must have some property Φ and one must have a justified belief (B4) that B3 has Φ. And so on, ad infinitum.

BonJour raises the regress problem in relation to "basic beliefs", beliefs which provide a "secure foundation" for knowledge (ibid, p.30). BonJour concludes the regress problem undermines the possibility of basic beliefs because it yields the "disturbing result" that a belief cannot be basic because its justification depends on a further belief, which in turn needs to be justified by another belief, and so on, ad infinitum (ibid, p. 13). For BonJour (in 1985 at least) this provides a reason to abandon foundationalism and support coherentism.

We can understand foundationalism as the view that all beliefs receive their justification inferentially from other beliefs, where the chain of justification must be traced back to some belief that does not receive its justification from other beliefs (such beliefs may be self-justified). By contrast, coherentism holds that a belief's justification is a function of whether it coheres with other beliefs within a system of

beliefs. Justification, on the coherentist model, does not require a terminus. The finer-details of this distinction are beyond the scope of this paper, but at least we have two different metaphors for knowledge. John L. Pollock (1974), preceding BonJour, uses the metaphor of a pyramid in describing foundationalism, where propositions which one knows in the lowest tier of the pyramid are those that one knows without requiring an independent reason for believing them to be true (p. 24). By contrast, Pollock's metaphor for coherentism is of a "vast nebula" in which justification does not terminate anywhere (p. 26).

Thus, it seems that an epistemic doxastic requirement results in a sceptical regress. If sceptical results are implausible, then the regress implies the pyramid metaphor is not the right one, and that we should instead conceive of justified beliefs as existing within a network or nebular of mutually consistent beliefs. By contrast, by adopting an externalist account of justification and abandoning the awareness requirement for justified belief, the regress does not get started. If the regress does not arise, then we needn't remove the pyramid and replace it with a nebula. For example, providing that one is, *in fact,* reliable in Goldman's sense, then one's belief will be (externally) justified.

Is there a way of avoiding the regress while maintaining an awareness requirement for justified belief? If epistemic properties supervene on non-epistemic properties, then it may seem the regress has a terminus (consistently with Van Cleve, 1985).[84] According to the epistemic supervenience principle, any two worlds W and W* which are the same in their non-epistemic properties are *necessarily* the same in their supervening epistemic properties. [85] It seems attractive to suggests that epistemic properties supervene on non-epistemic properties.[86] For instance, could the

[84] Van Cleve (1985) uses epistemic supervenience to address the regress problem. Van Cleve maintains that epistemic supervenience implies that there are self-justified beliefs, writing that "A self-justified belief is not a belief that bestows upon itself the justification it already has; it is a belief that has for its justification-generator (or justifier) the fact of its own occurrence (together, of course, with its being the kind of belief that it is)" (1985, p). A self-justifying belief will be a belief which is justified in virtue of standing in relation to non-epistemic properties. I cannot do full justice to Van Cleve's view here. It is important to emphasise, however, that my concern that the regress problem cannot be met by epistemic supervenience is benign, because the regress problem arises as a challenge to the awareness condition I am supporting.
[85] More formally, epistemic supervenience (for epistemic properties "E" and non-epistemic properties "NE") is as follows: $\Box NE(W \equiv W*) \rightarrow E(W \equiv W*)$. This is a strong supervenience principle (following Kim, 1984, "Concepts of Supervenience"). Note that epistemic supervenience is non-symmetric — I.E. two worlds which differ in non-epistemic properties can be the same in their epistemic properties. Thus, given non-symmetry, epistemic supervenience implies $\Diamond NE(W \not\equiv W*)\&E(W \equiv W*)$.
[86] As Van Cleve writes, "that there could be another belief just like the original in all natural respects - directed at the same proposition, caused by similar causes, accompanied by similar experiences, related in the same ways to other beliefs of its subject, and so on - yet not justified. This, too, is absurd". (ibid, p. 99).

ramblings of a conspiracy theorist be justified in possible worlds that are the same as the actual world in their non-epistemic properties? The unintelligibility of the claim that conspiracy theories could be justified in such a world is consistent with epistemic supervenience. For it is hard to believe that someone could hold that such a world exists while *sincerely* believing that conspiracy theories are unjustified in the actual world. If epistemic properties supervene on non-epistemic properties, could this fact terminate the regress? James Van Cleve (1985) explains that if epistemic supervenience is true, then there has to be such a phenomenon as the *generation* of justification involving self-justified beliefs. Van Cleve construes a self-justified belief as one that "bestows upon itself a justification it already has", contending that epistemic supervenience implies there must be such a thing as *the generation* of epistemically justified belief" (p.100).

A concern with using epistemic supervenience to terminate the regress, is that epistemic supervenience is plausibly ascriptive rather than ontological. If epistemic supervenience is merely ascriptive, then this would not imply that the justificatory-chain has a terminus.

To understand why, we need to distinguish between ascriptive and ontological supervenience (Klagge, 1988). *Ascriptive supervenience* places a logical constraint on the truth of normative attributions (ibid, p.464). In the epistemic case, ascriptive supervenience would appeal to the implausibility of denying that two worlds could be identical in their non-epistemic properties and yet differ in their supervening epistemic properties, as seen where it seems unintelligible that a possible world, the same as the actual world in non-epistemic properties, could be a world in which conspiracy theories are justified. By contrast, *ontological supervenience* would affirm something stronger for the epistemic case, namely: that there *are* non-epistemic states of affairs that bear responsibility for epistemic states of affairs. Ontological supervenience is the claim that when our ascriptions are *correct,* then that is because there are such epistemic properties, standing in a dependence relation to non-epistemic properties in virtue of which the ascriptions are correct. The relevant non-epistemic properties would, on this view, instantiate or bear responsibility for the epistemic properties. Importantly, Klagge observes that ontological supervenience cannot be derived from ascriptive supervenience. The reason for this is because – as Klagge puts it – ascriptive supervenience is a constraint on our judgments and does not entail that the world is constituted in any particular way (ibid, p. 464).[87]

---

[87] As Klagge characterises this: "from the fact that it is reasonable to place certain constraints on our judgments, it does not follow that the world is constituted in any particular way." (Klagge, 1988 p. 464) Klagge's observation is that examples such as Hare's (1952) example of

If the regress problem remains, then it is a problem only to the extent that conditions for justification *ought* not to imply global scepticism. By describing the sceptical regress as a *problem*, I have assumed what Chapter 2 (§ 2.3) identified as a "particularist" approach to the conditions for justified belief. [88] The particularist approach holds that any criterion or condition *c* for justified belief ought to accommodate the intuition that we can have justified beliefs in propositions, where if *c* has radical sceptical consequences, then that implies *c* is not a necessary condition for justified belief. By contrast, methodism addresses the question of whether a given belief is justified by asking whether that belief satisfies the correct criterion or conditions *c* for justified belief. In Chapter 2, I considered particularism in the context of anti-luck conditions. By contrast, the particularist stance in this chapter concerns conditions for justified belief.

I will assume a particularist approach to the conditions for justified belief, just as I assumed a particularist approach to anti-luck conditions in Chapter 2. The reason for this is the same: radical scepticism seems revisionist of our intuitions and is an exorbitant price to pay for awareness internalism. However, the particularist assumption is benign, because it implies that regress scepticism *is* a problem for awareness internalism – this makes the present task of supporting awareness internalism harder, not easier.

§ 2.1. Bergmann's Dilemma for Awareness Internalism

Bergmann (2006) shares BonJour's worry about the sceptical regress, explaining how the regress arises because finite creatures cannot comprehend infinitely long chains of reasoning (p. 15). Bergmann argues that the internalist is torn between a version of awareness internalism that accommodates the central intuitions which support internalism but at the price of generating scepticism, or a weaker version of awareness internalism that avoids the regress but fails to accommodate the intuitions which motivate internalism.

To understand Bergmann's argument, let us first state the doxastic awareness principle, which Bergmann discusses:

St Francis, the entailment is between the specification of some standard and the evaluative judgment (p. 463).[87]

[88] See Chisholm (1982, pp. 65-68) for the distinction I am broadly appealing to.

Actual doxastic strong awareness requirement (ADSAR): S's belief B is justified if only if (1) there is something, Φ, that contributes towards the justification of B and 2) S is actually aware of Φ in such a way that S ***justifiably believes*** that Φ is in some way relevant to the appropriateness of holding B. (Bergmann. p. 15). My emphasis.

Recall the difference between epistemic and non-epistemic doxastic awareness requirements from §1. Notice that ADSAR is an epistemic doxastic awareness requirement because it requires that S *justifiably believes* that B has Φ, and therein is the source of the regress.[89]

Bergmann examines versions of ADSAR which do not require that S *actually* holds a second-order belief, but which instead require that S could *potentially* have a justified second-order belief (that one's first-order belief has Φ). Bergmann observes how even weaker principles generate a regress. The reason weaker principles lead to a regress is because they require one *could* grasp an infinitely long chain of reasoning, of the sort required by ADSAR; again, this requirement does not seem satisfiable.[90] In summary: it is not the occurrence of the second-order belief which generates a regress, but instead, the requirement that one *could* comprehend an infinitely long chain of reasoning.

This leads us to consider less demanding forms of awareness which may avoid a skeptical regress. Specifically, Bergmann argues that non-doxastic awareness requirements may avoid the regress. The idea is that if one's awareness does not have propositional content, then a further question about whether that awareness is justified does not arise. The problem Bergmann identifies is that less demanding forms of awareness internalism avoid the regress at the expense of failing to provide a motivation for awareness internalism over externalism.

To understand why, let us consider what Bergmann calls "non-conceptual weak awareness" (NCWA). Bergmann argues that one might be conceptually aware of Φ because one can categorise Φ "according to some classificatory scheme", without having a belief about Φ (p. 17). NCWA is a form of awareness which reflectively unsophisticated subjects may have of objects and their environments. As Bergmann describes it, such awareness "doesn't involve the application of any concepts to the objects of awareness" (p. 19), and he cites the fact animals may be said to be aware of

---

[89] As Bergmann (2006) presents the concern, principles such as ADSAR require that "one has a justified belief only if one has an infinite number of justified beliefs of every-increasing complexity" (p. 15),

[90] Bergmann considers "potential [epistemic] doxastic strong awareness" and explains in detail why they still give rise to a regress (see 2006, p. 16).

the fact that they are in pain, for example, because they have the experience of being in pain and an awareness of being in pain.[91]

The difficulty with weaker requirements is that they are vulnerable to the following objection:

> The Subject Perspective Objection "If the subject holding a belief isn't aware of what that belief has going for it, then she isn't aware of how its status is any different from a stray hunch or arbitrary conviction. From that we may conclude that from her perspective it is an accident that her belief is true. And that implies that it isn't justified" (Bergmann, 2006, p. 12).

Bergmann identifies the Subject Perspective Objection as the single most compelling motivation for internalism. The objection is, at root an objection about epistemic luck or "accident" as Bergmann characterises it. The idea here is that one can satisfy an externalist condition for justified belief and yet it can still be a matter of luck that the belief is true. To support this, Bergmann appeals to BonJour's NORMAN counter-example, claiming that this is a case in which one can put forward the Subject Perspective Objection. Bergmann's claim that Norman is vulnerable to the Subject Perspective Objection is consistent with BonJour. For example, BonJour claims that, in the absence of awareness, it is an *accident* from Norman's perspective that his belief is true (ibid, pp. 43-44).

Weaker awareness requirements, it seems, would fail to meet Subject Perspective Objection, because weaker requirements – such as non-conceptual weak awareness –may be satisfied in cases where one can still make the Subject Perspective Objection. For example, notice that we can imagine NORMAN having at least has *some experience* of his clairvoyant abilities. We can imagine that whenever he has a clairvoyant hunch, he would be able to classify that hunch as indicating that the President is in New York City, and Norman might form the belief on the basis of that experience. Norman, I suspect, may satisfy Bergmann's notion of non-conceptual weak awareness.

Notice here that Bergmann's strong doxastic awareness requirement, ADSAR, *does* meet the Subject Perspective Objection. For if Norman were enlightened and if he believes with justification that his clairvoyant beliefs are reliably formed, then it does not seem a matter of luck that his beliefs are true. However, we have seen that a strong doxastic awareness requirement generates regress scepticism. Consequently,

---

[91] Bergmann (2006), p.19 (footnote).

awareness internalism faces a dilemma: either opt for a weak awareness condition which avoids regress scepticism, but does not meet the Subject Perspective Objection, and therefore does not provide a reason (based on anti-luck considerations alone) to support internalism over externalism) or, alternatively, meet the Subject Perspective Objection by opting for a stronger awareness requirement which results in regress scepticism. On the proviso that global scepticism is false, we have no reason (based on anti-luck considerations alone) to support awareness internalism over externalism.[92]

Finally, note that Bergmann provides a number of distinctions: weak/strong awareness (p.13), doxastic/non-doxastic (pp.15-17), conceptual/nonconceptual (pp. 17-20); some of these principles avoid the regress by not requiring second-order beliefs. However, Bergmann argues that anything short of strong awareness fails to meet the Subject Perspective Objection while avoiding the regress. [93] I wish to concede that Bergmann is right in concluding that less demanding principles would fail to accommodate the regress problem while meeting the Subject Perspective Objection.

§ 2.2. Non-Epistemic Doxastic Awareness Requirements

However, I think there is a principle which may avoid the regress problem while meeting the Subject Perspective Objection. Recall the non-epistemic doxastic awareness requirement.

> Non-Epistemic Doxastic Awareness Requirement (NEDAR). S has a justified belief B only if: B has some property Φ and S believes (either occurrently or non-occurrently) that B has Φ.

We saw how epistemic requirements generate a sceptical regress. However, the regress will not get started if one abandons the *epistemic* component of the principle, retreating instead to NEDAR. The regress is generated by the claim that one "must

---

[92] The specific premises of Bergmann's (2006) argument are as follows (quoting Bergmann): "1) The essential feature of internalism is that it makes a subject's actual or potential awareness of some justification-contributor a necessary condition for the justification of any belief held by that subject. 2). The awareness required by internalism is either strong awareness or weak awareness. 3) If the awareness required by internalism is weak, then internalism is vulnerable to the Subject Perspective Objection. 4) If internalism either leads to radical scepticism or loses its primary motivation for imposing the awareness requirement (i.e. avoiding the Subject Perspective Objection), then we should not endorse internalism. (Bergmann, 2006, pp. 13-14).
[93] As Bergmann (2006) describes this, "all such awareness will either involve *conceiving* of the justification-contributor that is the object of awareness as being in some way relevant to the justification or truth of the belief or it will not. Let us say that if it does involve such conceiving, it is a strong awareness and that if it doesn't, it is weak awareness" (p. 13).

justifiably believe" (Bergmann, p. 15) that B has Φ or, analogously, one must "believe with justification" that B has Φ (BonJour, 1985, p. 32). By abandoning the epistemic component of the awareness condition, an infinitely long-chain of beliefs is avoided; one does not need a further, third-order belief that one's second-order belief is justified (and so on, ad infinitum). If one's second-order belief does not need to be justified, then the regress terminates at the level of one's second-order belief.

All that would now be required to satisfy NEDAR is that one's belief B does, in fact, have Φ and that one believes (either occurrently or non-occurrently) that B has Φ. If the second-order belief does not have to be justified, then the regress does not get started.

Is NEDAR able to meet the Subject Perspective Objection? In answer, it seems that if Norman genuinely believes he is a reliable clairvoyant, then his belief will not be luckily true in the same way as if he has no beliefs about the status of his first belief. What it seems we want to say is that there is an awareness failure which undermines his knowledge, but which does not undermine putative cases of knowledge where we are aware, even if not occurrently, that our belief-forming methods are reliable.

For example, consider the case of ENLIGHTENED NORMAN, who truly believes that he is a reliably clairvoyant. We seem *less inclined* to raise the Subject Perspective Objection in this case. Of course, merely believing that he is a reliably clairvoyant may not be enough to insulate ENLIGHTENED NORMAN from the Subject Perspective Objection. For example, we may require that ENLIGHTENED NORMAN also believes that his particular belief token results from that faculty. What is crucial here is that ENLIGHTENED NORMAN does not have evidence against the reliability of clairvoyance or evidence that his belief token is false. When he believes that his first-order belief tokens are true, then when they are true, it does not seem a matter of luck that they are true from his perspective or from anyone else's.

Bergmann is aware of requirements such as NEDAR, but he finds them unsatisfactory for reasons independent of regress scepticism and the Subject Perspective Objection. Bergmann finds NEDAR unsatisfactory because he worries that it allows intuitively unjustified beliefs to constitute cases of justified belief. Let me quote what Bergmann writes:

> "The attentive reader will notice that, in addition to specifying that the awareness in question requires an actual belief, ADSAR stipulates that the required belief must be justified. The reason for this is that if the internalist has the intuition that merely having a justification contributor isn't enough - that the subject must also believe that she has a justification-contributor - it seems highly doubtful that the internalist

> will be impressed by the mere belief (no matter how unjustified or insane) that the thing of which she is aware is a justification contributor" (Bergmann, 2006, p. 15)

Bergmann's legitimate worry is that if one's second-order belief does not need to be justified, then the awareness requirement will allow intuitively unjustified beliefs to count as justified. Since Bergmann mentions irrational beliefs, let us use the following example:

> PARANOID: S has a paranoid belief B which results from a delusion and S has believes that B is reliably formed.

It seems clear that S's delusional belief (about whatever) will not be justified. Adding a second-order belief does not make S's belief epistemically justified.

Two points are important. First, there is an intuition that justification, like knowledge, should be neither too difficult nor too easy to acquire[94] – if a theory makes demands that cannot be satisfied by believers, then regress scepticism threatens. Conversely, if a condition for justification allows PARANOID to count as justified, then that theory makes justified beliefs too easy to acquire. Thus, the worry is that *non-epistemic* awareness requirements are, therefore, too permissive. Equally, however, we do not want an awareness requirement to be so restrictive that it fails to include in the extension of justification beliefs which should, intuitively, be included in that extension. In sum, while we wish to explain PARANOID as unjustified, we do not want an explanation which results in scepticism.

Second, PARANOID is an epistemically blamelessness belief due to facts about the believer's pathology. However, it involves a blameless belief the believer *ought* not to have.[95] The case suggests that epistemically justified belief requires more than merely blameless believing. Though blamelessness is necessary, it doesn't seem enough. However, at the moment, it may seem that without insisting that one's second-order beliefs are *also justified*, then there is nothing to stop PARANOID from being internally justified.

[94] Cohen (2002) highlights a similar intuition about knowledge being too easy to acquire on certain accounts of knowledge.
[95] Chapter 4 develops this idea further, in particular contends that the relevant form of blamelessness involves the believer being exempted.

§ 3. A Way Out?

I will argue that we can retain the non-epistemic awareness requirement (NEDAR) and avoid the Regress Problem, while avoiding the principle being so permissive that it does not explain intuitive cases of justification failure. I argue that if what one needs to be aware of are externalist facts, such as the fact that one's belief is reliably formed, then there is scope to meet both the Subject Perspective Objection without committing one to a regress and without making one's theory so permissive that it fails to explain intuitive cases of justification failure such as PARANOID.

Note that the only reason NEDAR has so far been discounted is Bergmann's legitimate worry that abandoning an epistemic awareness requirement would make justified beliefs too easy to acquire. Thus, if Bergmann's concern can be met, then we will have a more defensible awareness requirement, one which meets the Subject Perspective Objection and which does not automatically lead to a regress.

In pursuing this line of inquiry, the first point to make it that NEDAR (like the other principles) allows for a broad supervenience base. Recall the principle:

> Non-Epistemic Doxastic Awareness Requirement (NEDAR): S has a justified belief B only if: B has some property Φ and S believes (either occurrently or non-occurrently) that B has Φ.

The principle is compatible with the possibility that the property Φ necessary for justified belief, which the agent must be aware of, is a mental representation or a phenomenal state. For example, what might make one's belief justified is that it *seems* to be the case that the proposition is true.

However, if Φ holds for B for whenever it *seems* for S that the proposition is true, then I suspect we would be unable accommodate cases such as PARANOID. For example, in PARANOID, it may seem from the believer's perspective that the proposition is true. Specifying that Φ is a phenomenal state would not straightforwardly predict that PARANOID (and cases like it) are unjustified.

However, there are other non-epistemic states of affairs which can feature in the right-hand conjunct NEDAR. For example, some of the relevant non-epistemic properties may be naturalistic relations. They may be, for example, the non-epistemic relations which externalists think are necessary and sufficient for justification, and which can, in principle, be investigated by the natural sciences (see Chapter 1, §3). The non-epistemic conditions may include causal relations, environmental normality, probabilistic relations, evidence, and the reliability of the belief-forming method.

Could any of these non-epistemic conditions explain the case of PARANOID? First, we can see that many of these states of affairs do not hold in PARANOID. Her belief has a high probability of being false, and she uses a highly unreliable method of belief formation.

To explain this verdict, however, we need to be clear that the notion of justification we are invoking amounts to more than merely blameless believing. Although PARANOID describes a blameless believer, the belief is epistemically unjustified. Goldman (1988) distinguishes merely blameless belief from epistemically justified belief; being blameless, i.e.. not ignoring countervailing evidence for one's belief, may be necessary for justification even though it is not sufficient.

Importantly, notice that we do not need to explain why PARANOID is unjustified at the level of the second-order belief. Instead, we are explaining why her first order belief (B) is unjustified by noting that it fails to satisfy an externalist condition for justified belief. It so happens that her second-order belief is not epistemically justified, and if we wanted to explain why that is so, we could appeal once again to her failure to satisfy an externalist condition together with her failure to be aware that an externalist condition holds. However, that does not mean we need a second-order explanation for why it is her first-order belief is unjustified.

Further questions arise. First, does an appeal to externalist conditions imply that justified belief does not require second-order awareness? I suspect it does not. We may consider this question in the context of the earlier case of NORMAN, which motivates a second-order awareness requirement to meet the Subject Perspective Objection. The case of NORMAN did not challenge the necessity of externalist conditions for justified belief, but instead challenged the sufficiency of externalist conditions for justification. NORMAN is intended to question the sufficiency of externalist accounts of justification. Consequently, an internalist awareness condition does not need to supplant necessary externalist conditions.[96] Conversely, the problem with PARANOID is not that she lacks awareness (justified or otherwise), but rather that her belief lacks the necessary properties required for her belief to be justified. While our explanation of PARANOID is first-order, our explanation of Norman is second-order (specifically, a failure to have an awareness that he is reliable)

The vital consideration is that if awareness internalism incorporates an externalist reliability condition as a necessary condition for justified belief, sometimes the explanation of justification failure will be at the first-order level of belief (e.g. PARANOID), and sometimes it will be at the second-order level (e.g. NORMAN). Being

---

[96] Again, we may distinguish supervenience internalism from awareness internalism (see Chapter 1, Section 3, for reasons why).

justified at the second-order level does not need to be a requirement of being justified at the first-order level because, as indicated, because we do not need justified second-order beliefs to meet the Subject Perspective Objection. We may avoid making awareness internalism too permissive by incorporating externalist constraints on first-order justification.

Secondly can one's first-order belief be defeated by lacking a justified second-order belief? BonJour (1985) describes cases which may be interpreted as instances in which one has a B2 but where there is countervailing evidence against the reliability of the method used for B1 (see Cases, 1-3 Chapter 2 §1 in this thesis). Those cases are like Norman in all the relevant externalist facts, except the believer either has evidence against the truth of the belief token or against the reliability of the method of belief formation. In such a case, one would satisfy the awareness requirement for justification, but one's first-order justification may be defeated by a blameworthy second-order belief. The blameworthiness of the second-order belief may be sufficient for the first-order belief to be similarly blameworthy. Indeed, BonJour's suggestion is that the believers in these cases should abandon their first-order belief and instead regard it as an unfounded hunch. In which case, a believer's blameworthiness at the second-order level may be enough to defeat one's justification at the first-order level. However, the subjects we have described do not have blameworthy second-order beliefs. For PARANOID, we need a first-order explanation of justification failure.

In summary, I have argued that there is scope to construe the awareness requirement doxastically so that it accommodates the Subject Perspective Objection while avoiding the regress problem. I have argued that if justification depends on externalist conditions, then a non-epistemic awareness requirement does not automatically make justification too easy to acquire.

§ 4. Meta-Epistemological Scepticism

The epistemic doxastic principle implied regress scepticism by imposing on S conditions which are unsatisfiable. By contrast, the (non-epistemic) doxastic principle avoids the regress by not imposing on S unsatisfiable conditions. Thus, attributor scepticism – the view that we speak falsely when we attribute justification – is implied by the epistemic doxastic requirement. Furthermore, by placing an externalist constraint on the justification of one's first-order beliefs, we can avoid the implication of attributor scepticism without risking making justification too easy to acquire.

The claim of this chapter, however, has not been that awareness internalism may surmount every type of sceptical argument facing it. Such a claim would be

implausibly strong. In the present section, I outline a sceptical argument which remains in situ, irrespective of whether we construe the doxastic requirement epistemically or non-epistemically. The scepticism I will discuss arises when the regress problem is presented as a first-person concern about whether or not we *do* have justified beliefs. For example, when one asks in the first person "Is my belief B justified" or "am I justified in believing that B has the property which makes it justified", we want something stronger than a condition for justification which *makes it possible* that one has a justified belief by not automatically generating a regress. It seems that what we want to know, when inquiring in this way, is that we do have a justified belief.

 To have that assurance, we would want to be sure our belief does, in fact, have the property which we think it would need to be justified. Suppose I believe that my belief B has Φ and that Φ makes B justified. If I am curious about whether I am justified in that second-order belief (B2), then I can ask the same question about B2, asking whether, in fact, it does have the property needed for it to be justified. Suppose I form another belief (B3) that my B(2) has the Φ which makes it justified, then I can ask the same question about whether B3 is justified, and so on ad infinitum.

The sceptical worry underscores a meta-epistemological anxiety that arises in the first person context of wanting assurance that we *do* have justified empirical beliefs (Fumerton, 1995). Indeed, insofar as we want to find out whether we have justified beliefs, then merely supporting awareness internalism (epistemic or otherwise) over externalism is not going to offer assurance that our beliefs do have the properties which we think their justification depends on. Equally, however, abandoning doxastic awareness and making justification only a function of reliability, causality, tracking, or other paradigmatic externalist conditions, will not supply one with an assurance that one *does* have justified beliefs.

Fumerton identifies, at length, the difficulty for externalism. In particular, he contends that any specification of a condition for justification in externalist terms is not going to provide us with an assurance that the condition is satisfied. The issue now is that the version of awareness internalism I have supported, make the question of whether one's belief B is justified contingent on whether B has some externalist property "Φe". The concern is that the externalist conditional that "if B has Φe, then B is justified" and the awareness conditional that "if B has Φe and S believes that B has Φe, then B is justified", offers no assurance that the antecedent of the conditional obtains for B. [97] Hence, we can inquire whether or not we are justified in believing that

[97] Note that we needn't analyse awareness internalism of the latter conditional. For instance, we can allow that one's justification to be defeated in circumstances in which B has Φe and S

Φe obtains for B, and the externalist answer will be that B is justified providing one's (second-order) belief that B has Φe also has some property Φe, and so on, ad infinitum. Fumerton contends that the ease with which externalist views can allow for second-order justified beliefs indicates that externalism fails to explicate epistemic concepts that are of philosophical interest (ibid, 170). Reciting the conditional would speak past the sceptic who is doubting that the antecedent of the conditional obtains.[98]

The problem for the awareness condition I have supported is that the condition, to the extent that it offers a response to scepticism, does not remove the doubt that the antecedent of the conditional obtains. Since Fumerton's argument against externalism is extensive, and allied to an arguments against naturalized epistemology, I want to focus just on his claim that the externalist offers at most a conditional response to the first-person concern about whether we do in fact have justified belief, which is equally an objection to the version of awareness internalism I have supported in this chapter. We may understand the concern with whether our beliefs do *in fact* have the properties they would need to have to be justified as *meta-epistemological scepticism*.

The awareness principle I have supported does not offer a response to meta-epistemological scepticism. Instead, it has been offered as a condition to explain our third-person epistemic evaluations, about cases in which one can intuitively be said to lack a justified belief and cases in which one can intuitively be said to have a justified belief. The awareness condition avoids scepticism only to the extent that it does not impose unsatisfiable doxastic demands on justified belief; more specifically, scepticism is avoided by abandoning the requirement that the believer should be able to comprehend an infinitely long chain of reasoning for every justified belief she holds. However, it offers no assurance that our beliefs are, in fact, justified or reliably formed.

However, the fact that the version of awareness internalism supported offers no assurance that our beliefs are in fact justified need not count against it as an explanation of justified belief. We need to distinguish here between avoiding attributor scepticism and avoiding meta-epistemological scepticism. Attributor scepticism arises if knowledge ascriptions are false because a theory has imposed

---

believes that B has Φe (per §4). Notwithstanding, if B has Φe and S believes that B has Φe, and there are no normative defeaters present, then it is at least possible that B is justified, even though we have no assurance that B does, in fact, have Φe.

[98] As Fumerton describes this, "...if as a philosopher I start wondering whether perceptual beliefs are accurate reflections of the way the world really is, I would not dream of using perception to resolve my doubt...the use of perception could never satisfy a *philosophical curiosity* about the legitimacy of perceptual beliefs" (p. 177). For an opposite view to Fumerton's, see Kornblith (2002), Chapter 6

unsatisfiable conditions on justified belief. The awareness condition avoids attributor scepticism to the extent that it offers a framework that *makes it possible* that one can justifiably believe, and it does so by not *implying* that we are speaking incorrectly when we attribute justification. However, the awareness condition does not avoid meta-epistemological scepticism because it does not *demonstrate* that our beliefs have the properties they would need to have to be justified.

On this point, it is often said that the regress does not arise on an externalist notion of justification.[99] However, a version of the regress problem does arise whether or not one endorses internalism and externalism. The reason for this, is because any desire to have philosophical assurance that we *do* have justified beliefs cannot fully be adequately answered by reciting the conditional that "if a belief B has a given externalist property Φe, then, B is justified".[100] If the regress problem is the problem of demonstrating that we justifiably believe or know a proposition, the present account of awareness internalism fares no better than an externalist account of justification – both are vulnerable to the meta-epistemological challenge.

We may, therefore, take it that the appropriate way in which awareness internalism (specifically, the version supported in this thesis) offers a solution to the regress problem is that it allows for the *possibility* of justification and knowledge.[101] That is, it imposes, on justification, conditions which are at least consistent with having justified beliefs. It does so by allowing that when one attributes justification to oneself or others, one is not necessarily systematically mistaken in thinking that the attributions are correct.

To further understand the distinction between attributor scepticism and meta-epistemological scepticism, we may note how they involve different – indeed opposite – meta-epistemological presuppositions about justification and knowledge. For example, when we ask of ourselves, whether our beliefs are justified, a regress threatens. For we are now no longer presupposing that we are justified in believing what we think we ordinary know or justifiably believe. Instead, we are attending to sceptical possibilities to see whether they are eliminated by the justification we have for our beliefs. Here, we find that neither the conditional "if B has Φe and S believes that B has Φe, then B is justified" nor the bare externalist conditional "if B has Φe,

---

[99] For example, BonJour concedes that if the conditions for justification are entirely external (if, they involve no further beliefs or other cognitive states, SEK p. 35), then no regress of justification is generated (SEK, p. 35).

[100] We can view both externalist conditions and awareness internalism in the context of Nozick's project of *explaining how knowledge is possible* rather than *proving* that we know (see Nozick, 1981, pp. 15-15). Nozick notes that he does not want to *refute* the sceptic but instead wants to explain how knowledge is *possible*.

[101] Nozick (1981), pp. 15-16 on the distinction between refuting the sceptic and offering an explanation which allows for the possibility of knowledge.

then B is justified", offer assurance that *in fact* we do have justified beliefs. Both accounts are vulnerable to meta-epistemological scepticism.

By contrast, the avoidance of attributor scepticism does not take as its starting point a question about whether or not we have justified beliefs but instead asks whether our theories of knowledge or justification can accommodate the intuitive fact that we are justified in our beliefs. I suspect it is attributor scepticism which paradigmatic externalists accounts - which reject awareness requirements - seek to avoid, for they offer an account of how these epistemic possibilities can be reconciled with the intuition that we are justified or do know quite a lot. The presupposition, in that case, is that we do know and justifiably believe much of what we think we do and that definitions of what it takes to be justified should reflect this.

In summary, whereas meta-epistemological scepticism arises once we abandon the presupposition that we are justified in our beliefs, attributor scepticism – and the need for an account of justification which avoids it – makes the very presumption which meta-epistemological sceptical arguments dispute. By keeping attributor scepticism and meta-epistemological scepticism distinct, we can see that there are not clear anti-sceptical grounds to support an externalist notion of justification over the version of awareness internalism supported so far in this thesis. However, the meta-epistemological challenge shows how the anti-sceptical import of internalism is, to some extent, limited.

Chapter 4

Supervenience Internalism and the Reliability Condition

This chapter considers the view that unreliable belief-forming methods may yield justified beliefs. It does so by introducing the Demon Problem. The Demon Problem describes a possible world in which one uses an unreliable method of belief formation, but where one may, intuitively, have justified beliefs. I explain how the Demon Problem is a motivation for a view of justification called "supervenience internalism". I explain how supervenience internalism allows that one may have justified beliefs in circumstances in which one's belief-forming methods are unreliable.

The Demon Problem initially suggests that the conception of "awareness internalism" defended in previous chapters may be incorrect. The reason for this is because the version of awareness internalism defended in previous chapters makes justification partially dependent on external factors, specifically reliability. Supervenience internalism is incompatible with *the account of justification* which I have defended. Thus, if only supervenience internalism offers a satisfactory response to the Demon Problem, it will count against the internalist account of justification I have supported.

This chapter responds to Demon Problem. It finds that the Demon Problem does not offer a conclusive reason to support *supervenience internalism* over the current brand of *awareness internalism*.

The chapter is structured as follows:

§1. I explain the New Evil Demon Hypothesis. I explain the challenge, define supervenience internalism and consider how content externalism about beliefs bears upon the case for supervenience internalism.
§2. I consider externalist responses to the Demon Problem. I consider analogies between the justification of a belief and the justification of actions. I consider responses to the Demon Problem offered by Goldman (1988) and Williamson (2007), and find the case for supervenience internalism inconclusive. However, I explain that a satisfactory response to the Demon Problem needs to be able to accommodate our

comparative normative intuitions. I explain how there is a modal luck involved if at most we are *merely blameless* in the demon world, and examine a way in which we may be able to appeal to Williamson's insights and LittleJohn's (2009) insights to explain our comparative judgments.

§ 1. The Demon Problem and Supervenience Internalism

The Demon Problem starts with the hypothesis that one is an incorporeal victim of deception. For example, imagine a possible world in which an omnipotent and malevolent demon ensures that all of our empirical beliefs are false, but in that world, we have no means of knowing that we are deceived in this way. The world is, in phenomenal respects, the same as the actual world (we might equally substitute the demon scenario with the (1981) Putnam's brain-in-a-vat hypothesis; a possible world in which one is a bodiless brain-in-a-vat).The phenomenal properties are the relevant "internal" properties which are held constant in both the sceptical world and the non-sceptical world. The Demon, being an omnipotent and deceptive creature, arranges the possible world so that it looks the same from the subject's perspective. Similarly, if one were a disembodied brain-in-a-vat (BIV), artificially stimulated to have the experiences which are phenomenally the same as the actual world, one would think one was in the actual world. Externalist conditions, such as the fact that one is using a reliable method of belief formation, fail to hold in the demon world. The Demon ensures that our belief-forming methods are unreliable; they invariably lead to false beliefs in empirical propositions.

Consequently, we have described a scenario where one lacks externalist justification but where one may nevertheless, intuitively, be said to hold justified beliefs. The verdict that beliefs in the demon world are justified conflicts with externalist accounts of justification. It also conflicts with the account of internalism which I have defended as that account incorporates externalist conditions necessary for epistemic justification.

The demon hypothesis and the brain-in-a-vat hypothesis also arise as a sceptical challenge. For example, given the possible demon world, how can we be sure that we do know or justifiably believe what we ordinarily think we know or justifiably believe? In what follows, I will not be discussing the demon world as a sceptical challenge. Instead, what is called the *New Evil Demon Problem* (henceforth the *Demon Problem*) concerns the *status* of epistemic justification. It does so by appealing to the claim that *intuitively* our empirical beliefs in the demon world are epistemically justified. Since this intuition conflicts with the implications of

externalist accounts of justification, which predict that demon beliefs are not epistemically justified (Cohen, 1984). The demon hypothesis also conflicts with the version of awareness internalism supported in Chapter 3, which makes externalist conditions necessary for justified belief.

The Demon problem suggests that externalist conditions for justification, conditions such as reliability, are unnecessary for justified belief. We may present the Problem as follows, taking the general form from Cohen (1984).

i.   In the Demon World, we have the same justified empirical beliefs which we have in the actual world.

ii. In the Demon World, our empirical beliefs do not satisfy the reliabilist standard for justified belief.

iii. Therefore, the reliabilist standard for justified belief is not the correct standard for justified belief.

Note that I have focused on the "reliabilist standard" for justification above. However, the Demon Problem arises for any theory of justification which makes a belief's justification depend on conditions which fail to hold in the demon world. For example, for an empirical proposition "p", if having a justified belief that p requires knowledge that p (the knowledge norm for belief - Williamson 2007), then since one falsely believes that p and thus fails to know that p in the demon world, the knowledge norm would predict that empirical beliefs in the demon world are unjustified.

To elicit the relevant intuition which supports the Demon Problem, we should need to focus on some empirical proposition "p" which *is* a case of knowledge or justified belief in the actual world. We may focus on some reasonable p, such as a subject's belief "I have hands". Holding this belief in the demon world may seem reasonable and justified. If so, then the Demon Problem suggests that justification does not require reliability.

One way of meeting the intuition is to endorse a form of internalism called supervenience internalism (e.g. Audi, 2001). Supervenience internalism holds that any two possible worlds which are the same in phenomenological respects are necessarily the same with respects to epistemic justification. Since the demon world is ex hypothesi phenomenologically identical to the actual world, this phenomenological likeness may *explain* the intuition that the demon world is the same with respects to justification.

We may express the supervenience claim as follows:

If W and W* are identical in phenomenal respects, then necessarily W and W* are identical in justificatory respects.

One way of responding to the demon is to argue that factors which make one justified in any given belief are factors which hold across both worlds. Since there are certain phenomenal features of one's situation which are held constant across both the actual world and the demon world (for example, what one is aware of, or how the scenario *seems*), these facts emerge as an explanation of the demon case. The aim in what follows is to argue that the Demon Problem does not provide a sufficient reason to abandon the view that justification depends on having a reliably true belief.

§ 1.1. Questioning the Intuition

The first task is to inquire as to whether it is obvious that we *are* epistemically justified in the demon world – i.e. to inquire whether we should accept the first premise of the Demon Problem: "In the Demon World, we have the same justified empirical beliefs which we have in the actual world."

The first thing to note here is that beliefs in the demon world are not *successful*. They are unsuccessful because they are false and not cases of knowledge, and also because they are not arrived at using belief-forming procedures which are conducive to having true beliefs or knowledge.

The Demon Problem implies that the failure of beliefs to be successful does not undermine their justification. Stewart Cohen, for example, holds that beliefs in the demon world are just as justified as beliefs in the actual world. Cohen contends that this verdict relies on viewing justification as a normative concept, and he regards the verdict as supported by the fact that the envatted victim is not *responsible* for circumstances of which he is unaware of (Cohen, 1984, p. 282).

Similarly, Robert Audi (2001) supports the view that beliefs held in radical sceptical scenarios are justified, and he does so by arguing that normative judgments about justification belief reflect normative judgments about justified action. For example, Audi equates the normative situation of the victim of sceptical scenarios to that of a skilful surgeon who loses a patient. In Audi's example, a competent surgeon has chosen to operate on a patient but unfortunately loses the patient. Audi (p.-23-24) contends that the surgeon should explain his actions without apologizing,

concluding that we would think of the surgeon's actions as reasonable and that we would *approve* of them.102

However, our intuitions concerning the demon case may not align with those of Audi and Cohen. Williamson (2007), for example, claims that there is an inherent normative difference between having successful beliefs and unsuccessful beliefs and that this supports the idea that belief norms are *success* norms. Williamson does not share Audi's view of the skilful surgeon who loses a patient, claiming that while we do not blame the surgeon, neither we do not praise him, and we think in some sense the decision to operate was the wrong decision (p. 23). The surgeon has an *excuse*, on Williamson's view, but is not justified. Similarly, Williamson presents the subject in the demon world as having a "cast-iron excuse, but not justification" (2007, p. 18), Williamson's statement suggests that our intuitions about justification in the demon case are far from uniform.

What seems evident, however, is that beliefs in the demon world are *blameless*. The view that demon beliefs are blameless is compatible with the internalist account of justification defended so far, because the account stated that being *blameworthy* (having a defeater) is enough to undermine justification, but that we may nevertheless have *blameless beliefs which are not epistemically justified*. For example, we considered cases in which one forms *paranoid* beliefs as examples in which the subject is blameless but unjustified.

The previous points do not imply that the Demon Problem does not pose a challenge to the account of justification offered so far. In particular, the point to make is that many beliefs in the demon world do not seem *merely blameless* in the way that other beliefs (for example, paranoid beliefs) are *merely blameless*. I therefore think that a satisfactory response to the Demon Problem would at least need to explain the proper way in which the believer has an *epistemic excuse* which is not conferred by the beliefs being *merely blameless*. § 2 of this chapter will attempt to offer an account which allows us to accommodate the view that many beliefs formed in the demon world are *not merely blameless*.

---

102 It may be helpful to have the full quote here. Audi contends, of the sceptical case " I am like a surgeon who skillfully does all that can be expected but loses the patient. There I should feel regret, but not guilt; I should explain, but need not apologize; and when we know what my evidence was, *we approve of what I did*. We consider it reasonable. The evidence, moreover, was not subjective in the sense that others would not respond to it as having the same force I have taken it to have. This kind of evidence is, however, ultimately internal. It centrally involves my sensory states, memory impressions, inferences, and the like. Such internal grounds seem fundamental in justification. They are indeed *partly constitutive of the concept*: one could not fully understand the concept without seeing the relevance of such grounds to the justifiedness of belief" (2001, p. 23.-24). (My italics)

§ 1.2. The Content Externalist Challenge

To consider the claim that we have the *same* justified *beliefs* in the demon world and the actual world, I will discuss the nature of belief content, rather than the question of what justifies a given belief.

If we lack *the same beliefs* in the demon world as we do in the actual world, then trivially we do not have the *same justified beliefs* in the demon world as we do in the actual world. The issue I wish to raise is that of how two very different environmental situations we are describing (the demon and the non-demon world, presumably the actual world) may imply that the beliefs in each situation will, respectively, have different contents and be different beliefs, which would undermine the claim that we have the same justified beliefs in both scenarios. Consider Cohen's presentation of the demon hypothesis:

"…on the demon hypothesis, we would have *every reason for holding our beliefs that we have in the actual world.* Moreover, since we actually have reason to believe that our cognitive processes are reliable, it follows that in the demon world we would have every reason to believe that our cognitive processes were in fact reliable." (p 281, Cohen 1984). (my italics).

Implicit, above, is the presumption that we would have *the same beliefs* in the actual world α and the demon world. Namely, that beliefs in the demon world are *our* beliefs. This view suggests that first: all of the beliefs which we have in α are also beliefs which we hold in the demon world, and second, that those *same beliefs* in the demon world are no less justified than they are in α.

The content externalist challenge (See, for example, Williamson – 2007) is that we do not have the same *beliefs* in the demon world as we do in α. Furthermore, if we do not have the same beliefs α and the demon world, then trivially we do not have *the same justified beliefs* in α and the demon world (and, importantly, we do not have the same *unjustified* beliefs in both worlds either). According to that counter, we do not need to accept that belief norms are not externalist norms.

Let us briefly consider some motivations for content externalism. Thought experiments are consistent with content externalism. I will first consider an externalist view of meaning before applying it to the case of beliefs. Here is a line of reasoning often described in the Philosophy of Mind: consider the property "water" and the compound substance H2O. What we term "water" is *necessarily* any substance with 2 hydrogen atoms and 1 oxygen atom, such that all and only samples of H2O are

samples of water. Let us further imagine that the necessary and sufficient properties of water are not its macroscopic properties (i.e. the fact that water is odourless, transparent, tasteless) but are instead its microscopic properties – (i.e. that it has a given molecular structure).

Let us now imagine another world, which is phenomenally the same as the actual world but in which there is a compound which macroscopically is the same as water, but which differs in microscopic respects, for example, by being composed of 3 rather than 2 hydrogen atoms. Such a substance would, on Putman's (1973) view, not be water despite its apparent macroscopic identity to water. If so, on this view, when an inhabitant of the possible world (the ~H2O world) stumbles across the substance that is macroscopically the same as water and makes utterances about what she terms "water", her exclamation would be referring to a substance other than water. Putnam appeals to our intuitions to support this account. Putnam's focus is on the twin earth scenario, another planet which is mostly the same as ours. Putnam's account predicts that if someone who knows "water is necessarily H2O" were to visit the planet in which the watery-like compound is ~H2O, they would see that the inhabitants of that planet are using the term water differently to how inhabitants of our world use the term. According to this view, what a subject means when she states "that is water" is determined by her environment. As Putnam summarises this thought, "meanings just ain't in the head" (1973, p704).

The above thought experiment may be extended to the claim that the contents of *beliefs* are also determined by environmental factors. For example, if the inhabitant of the twin-earth scenario *believes* "that is water", then given that what she *means* by "water" is not what we mean by water, it seems that her belief would be about a different proposition to ours of water back on earth.

On the externalist view of belief content, phenomenally identical duplicates may have different beliefs because belief content may be sensitive to environmental conditions. We have thus described a world which is environmentally *radically* different from the non-demon world. Given this radical difference, in the demon world, the content of our beliefs cannot be expected to be the same as the content of our beliefs in the non-demon world. If content externalism is supported, then the demon problem does not show that we have the same justified beliefs in the demon and non-demon worlds, as we do not necessarily have the same beliefs.

Accommodating The Challenge

I will concede that supervenience internalism may accommodate the content externalist challenge. (This concession is dialectally benign, as I will argue that there are other reasons why the Demon Problem is inconclusive).

To motivate the idea that supervenience internalism may be able to handle concerns regarding content externalism, consider that Audi (2001) says of the demon case:

> "Whether the internalist can find some proposition that both duplicates actually believe is not crucial; the point is that if the two are internally identical, then what they are justified in believing is the same. The case could be accommodated by maintaining that even for a proposition with external content, justificatory grounds are-in the accessibility sense-internal, even if their content (as where they are beliefs of premises for the proposition), is external. (Audi, 2001, p32).

There are two claims here which may be distinguished. First, the claim that whether beliefs in the demon world have the same content as non-demon beliefs does not matter to the internalist intuition. Second, we may instead focus on what subjects are *justified in believing*.

I suspect Audi is appealing here to a notion of justification similar to Engel's (1992b) notion of personal justification, where personal justification focuses on the normative status of the cognisors rather than the question of whether the beliefs themselves are justified. Engel distinguishes personal justification from doxastic justification, where personal justification concerns the attribution "S is justified in believing that p", and doxastic justification concerns the attribution "S's belief that p is justified" (pp. 135-136). Since the inhabitant of the demon world is at least epistemically blameless in many of her beliefs, she is personally justified in those beliefs.

However, if personal justification is what Audi has in mind, then there is a problem with appealing to personal justification to preserve the justificatory symmetry between the demon and non-demon worlds. The problem being that it is doubtful whether the two notions are equivalent. As Engel has argued, for many cases, these have different extensions. In particular, attributions about personal justification may hinge on our assessment of whether or not a belief is epistemically blameworthy; hence it is compatible with the personal justification of the demon victim that she lacks doxastic justification. Appealing to personal justification would, therefore, allow

the externalist to claim that while the demon victim is personally justified, her *beliefs* are unjustified because they violate externalist conditions for justification.

If personal justification only a version of blameless believing, then an externalist view of justification (and awareness internalism) need not deny that demon beliefs are personally justified. This is because externalist conditions for justification are not satisfied for whenever an agent blamelessly believes a proposition. Therefore, the externalist view need not entail that the agent in the demon world is not in *some sense* personally justified; instead, the view is that epistemic blamelessness is insufficient for one's belief to be justified. This may be extended to the conception of justification which I defended in Chapter 2 and Chapter 3; we need not deny that believers in the demon world are epistemically blameless in their beliefs, even if we deny that believers in the demon world are epistemically justified because they fail to satisfy a reliability condition.

The supervenience internalist would now need to reveal a difference in *doxastic* justification between the demon and non-demon worlds, but – so far – this line of reasoning conflicts with content externalism.

I would suggest that, instead of retreating to personal justification, we may instead question the import of content externalism for the demon problem (again, this is a benign concession because I will argue that there are other reasons for finding the Demon Problem inconclusive). Recall that if we assume a content externalist view of belief, it follows that the agent in the demon world would have very different beliefs to the agent in the actual world – this is because if one's environment determines belief content, then given the victim of the demon world is in a very different environmental situation to the believer in the actual world, she would have different beliefs to the believer in the non-demon world.

Is content externalism enough to diffuse the Demon Problem for our account of justification? I suspect not because someone might argue as follows: beliefs formed in the demon world (even if they have a different content to beliefs formed in the non-demon world) are still epistemically justified. Furthermore, the justificatory status of many of those beliefs (different in content to actual world beliefs) may, intuitively, seem equal to the justificatory status of beliefs in the actual world. That is to say, *beliefs* in the demon world are intuitively justified, even though there is not a success norm, in the vicinity, which is satisfied. Thus, one might argue that there is a possible world "*W*" in which subjects form intuitively justified empirical beliefs, but where no externalist belief-norms are satisfied. Thus, the account of justification I have defended is not immune to the demon problem.

To clarify this, consider the following standard presentation:

(1) "• Necessarily, for every justified belief in the actual world α, one has the same justified belief in world *W* (where *W* is phenomenally the same but very different in external respects)".

However, in formulating the demon problem (which, I emphaise, is a problem for the account of justification I am defending), I concede that one may replace the necessity operator in (1) • with an existential quantifier. Specifically, the demon problem would now be as follows: *there exists* some possible world *W* in which S intuitively has a justified belief B, even though there are no externalist conditions which B satisfies in W. More schematically:

(2): ∃*W* (B is justified & the reliability condition for B is not satisfied).

Notice that (2) does not imply (1). The reason being that (2) does not imply that we have *the same beliefs* in *W* as we do in the actual world α. Specifically, (2) does not entail that "• *for all justified beliefs in the actual world α, one has the same beliefs in W (where W is phenomenally the same but very different in external respects)".*

Framing the demon problem in terms of (2) does not require that we have the same *beliefs* in α and W. Instead, (2) expresses that at *W* the agent has *some* empirical beliefs which are justified. This allows the propositional content of beliefs to vary between α and W, respectively, as one would expect if content externalism is correct.

If the demon problem is formulated as (2) rather than (1), then the Demon Problem does not presuppose content-internalism. Hence, supporting content-externalism will not necessarily allow us to meet the demon problem, which is a problem for the awareness internalism I have defended.

Reconciling Content-externalism with normative-internalism (the claim that beliefs in the demon world are *justi*fied) may be possible if we formulate the challenge as (2). However, as indicated, the Demon Problem is a problem for the account of justification I have been defending. Therefore, any claim that normative-internalism can meet the content-externalist response to the Demon Problem is, dialectically, benign. However, I will argue that there are *other* ways of responding to the problem. I will argue that we may distinguish between being epistemically justified and having an epistemic excuse. I will argue that we can accommodate a class of intuitions about the demon case without having to accept the claim that beliefs formed in the Demon World are epistemically justified.

§ 2. Reliabilist Responses

In this section, I will examine attempts to explain the intuition that (many) demon beliefs are justified in a way compatible with an externalist account of justification, and in a way which is compatible with the awareness internalism I have defended; incorporating as it does a reliability condition on justification. I will cover some of the primary responses in the literature, and will support the one which I find most promising. The response I will prefer appeals to a distinction between having an excuse and an exemption as two ways of being blameless. I contend that many of the beliefs we form in the demon world are instances of *excusable* beliefs rather than *exempt* beliefs.

First, consider Alvin Goldman's reliabilism. The early Goldman endorsed reliabilism but within the framework of normal worlds (1979). *Normal-world-reliabilism* (**NWR**). As Goldman writes, "a normal world is understood as a world consistent with our general beliefs about the actual world, beliefs about the sorts of objects, events, and changes that occur in the actual world." (1988 p. 61). NWR holds that we should judge reliability only with reference to possible worlds which resemble the actual world. Goldman used NWR to explain counterexamples to process reliabilism. For example, Goldman considers a possible world W in which wishful thinking is a reliable method of belief formation. He says we would be inclined to think that beliefs which result from wishful thinking are epistemically unjustified. NWR explains this using the fact that at normal worlds (such as the actual world) wishful thinking is unreliable and results in unjustified beliefs.[103]

However, Goldman's NWR would not accommodate the view that the demon world is the same in justificatory respects as the non-demon world, nor does it purport to. The problem is that even if NWR accommodates our intuitions about cases of *unjustified* beliefs which satisfy the reliability condition (e.g. wishful thinking), NWR does not accommodate our intuitions about cases of justified belief which violate the reliability condition. Since the demon case concerns the *necessity* of reliability as a condition for justification, NWR does not fully explain the demon intuition because it still makes reliability as a necessary condition for justification.

---

[103] A further legitimate worry about *Normal-worlds-reliabilism* is that what constitutes a normal world will be a function of general beliefs, in the actual world, about "the sorts of objects, events, and changes that occur in the actual world" (Goldman, 1988, p. 61). But *which* beliefs are relevant here? General beliefs about what sorts of objects, events, and which changes occur in the actual world are subject to social, cultural, and historical variability. As Goldman (1988) points out, beliefs about normality are likely to exhibit considerable divergence.

Goldman later renounced NWR, replacing it with a distinction between *strong and weak justification* (Goldman, 1988) which I discussed in Chapter 2 (§ 1), and Goldman applies this distinction to the demon case. Goldman argues that demon beliefs are weakly justified but not strongly justified, where being weakly justified is necessary but insufficient for being strongly justified. A belief is *weakly justified* if it is *ill*-formed but *blameless* and *non-culpable* (1988, p. 53).[104] By contrast, a belief is strongly justified just in case it is blameless, *and* it results from a reliable method of belief formation (Goldman, 1988, p. 52). Goldman thus concludes that weak justification is a necessary but insufficient condition for strong justification: epistemic justification requires both that one is weakly justified in one's belief (I.e. that one is not blameless) and that one is strongly justified. Presenting weak justification as *mere blamelessness* allows weak and strong justification to be different notions. Goldman argues that beliefs in the demon world (or at least, many of them) are blameless and therefore weakly justified, even though they fail to be strongly justified (Goldman, 1988, p. 59-60).

The distinction between strong/weak justification also explains our intuitions about other cases. Recall the earlier example of the flat-earther (Chapter 1, § 1) who is unaware of any evidence for the proposition that the earth is round due to living in a culture isolated from the scientific community. The flat-earther is not a conspiracy and is *not irrational* in any empirical sense. Her belief is blameless, but not epistemically justified. Goldman's account predicts that we experience a "tug in opposite directions" with our epistemic intuitions (p.52). On the one hand, it seems that the flat-earther fails to have justified beliefs because she is using an unreliable method of belief formation; on the other hand, she is blameless.[105] The distinction between weak and strong justification explains the tug in intuitions. Furthermore, the fact that the flat-earther's beliefs are unreliable, and therefore unsuccessful, explains the intuition that they are not epistemically justified.

[104] Goldman argues that each conception of justification, strong and weak, seems to be a legitimate conception (p.53). However, for Goldman, strong justification is that species of justification required for *knowledge* (I.e. it is an epistemic justification) because it allows us to have true beliefs about matters of fact. For example, Goldman informs us that beliefs which have no chance of qualifying as knowledge cannot be justified in this stronger sense (p. 52). To illustrate weak justification, Goldman focuses on a "scientifically benighted culture" as a case of blameless but unjustified belief. Note that not everyone shares the intuition that cases of cultural isolationism describe necessarily blameless beliefs. For example, Steup (1988) see Alston (1988) *The Deontological Conception of Epistemic Justification*, pp. 146 for a response to Steup. Alston (1988) argues that denying blamelessness betrays an insensitivity to cultural differences (Alston, 1988, p. 146).
[105](Goldman (1988), p. 56).

However, I will suggest that strong/weak justification fails to accommodate our *comparative intuitions* and that we need to find some other way of accommodating the demon problem.

The difficulty for strong/weak justification may be illustrated by considering our comparative intuitions. Consider again the earlier case of PARANOID (Chapter 3, § 3) which involved a paranoid believer having a blameless belief which is not epistemically justified. Notice that, in the actual world, PARANOID is weakly justified according to Goldman's account. Notice that when we focus only on the actual world, strong/weak justification can explain the normative contrast between paranoid delusions and an eminently reasonable belief, such as HANDS ("I have hands"). For in actual world, PARANOID is weakly justified but not strongly justified, whereas HANDS is strongly justified (and a-fortiori) weakly justified. In the actual world, Goldman's strong/weak justification captures our verdicts about and HANDS and PARANOID.

The difficulty with the two cases I have described, HANDS and PARANOID, may hold in the demon world as well as the actual world. Goldman's account predicts that both HANDS and PARANOID are merely blameless and weakly justified in the demon world. However, it seems, in the demon world, there is intuitively a normative difference between PARANOID and HANDS. Yet, if beliefs such as HANDS are strongly justified in the actual world and only weakly justified in the demon world, then given that weak justification is a blameless non-culpable belief, we are committed to the verdict that both HANDS and PARANOID are *weakly justified* in the demon world. This verdict does not explain, however, the normative contrast that holds between these two cases in the demon world. It seems sheer modal luck that HANDS and PARANOID could be normatively different in the actual world, and yet normatively the same in the demon world, and the suggestion here is that an adequate response to the demon problem needs to exclude such modal luck.

§ 2.1. Excuses

According to Williamson's account, subject S has an epistemically justified belief B only if B is a case of knowledge (2000, 2007). Therefore, Williamson denies that the believer in the demon world has a justified belief. The subject in the demon world has, on Williamson's view, "cast-iron excuse, but not justification" (2007, p. 18). Having an *excuse* is not, we will see, equivalent to merely being blameless.

More Recently, Williamson explains away *the intuition about justification* by distinguishing between primary norms and derivative norms for belief. [106] The primary belief norm for Williamson is the knowledge norm. Here, one has a justified belief that p just in case one knows that p. However, there are further norms which are derivable from the knowledge norm, in particular, Williamson contends that one can be *disposed* to comply with the primary norm but fail to do so due to abnormal circumstances. Williamson contends that the victim in a sceptical scenario *intends* to comply with the knowledge norm and is *disposed* to comply with the knowledge norm, because in normal (non-sceptical circumstances), she *does* comply with the knowledge norm.[107]

Williamson argues that the satisfaction of norms which are derived from the knowledge norm explains the "residual sympathy" we feel towards the subject in the demon world.[108] The fact the believer satisfies a norm which is derivative from the knowledge norm - a norm which is sufficient for having an excuse. Therefore, demon beliefs are *excusable but not justified.*

Williamson argues that primary and dispositional norms are distinct, and that complying with a primary norm is neither necessary nor sufficient for complying with a derivative norm (p. 8). :Let us consider primary and dispositional norms, more generally, with an example (this case is my own, but I hope it illustrates the distinction between primary and derivative norms):[109] consider a dinner party where Samantha and her guest both have empty wine glasses and where a full, opened bottle of wine stands in the middle of the table. Samantha knows her guest would like another glass, and that it would be polite to replenish her guest's glass before filling her own. She knows there is an etiquette norm here that she *ought to replenish her guest's glass first* (call this "G"), and on most occasions, she adheres to this norm. However, on this occasion, Samantha fills her glass first. She thereby violates a primary norm (G) – violating this norm does not invite strong condemnation, but it is a norm nonetheless. However, Samantha satisfies a dispositional norm: she has the disposition to comply with G, even though on this occasion that disposition is not manifest. In which case, her action is excusable but not justified because she ought to have complied with the norm. Conversely, if Samantha lacked the disposition but still complied with G, then her action would be justified.

---

[106] In a draft paper Forthcoming in Julien Dutant and Fabian Dorsch, eds., The New Evil Demon (Oxford: Oxford University Press).
[107] Williamson, (Ibid., p. 13)
[108] Ibid., p 14
[109] Williamson emphasises that what is true of primary and dispositional beliefs norms applies to norms more generally and appeals to non-epistemic examples.

Considering the epistemic case, Williamson argues that in the case of sceptical scenarios, such as the demon world, there are derivative belief norms which are satisfied even though the primary ones required for *justification* are not. On Williamson's view, the (primary) norm for *justification* is *knowledge*; hence one's *belief* is justified if and only if it conforms to the primary knowledge norm. In the demon world, one has a false belief, and one violates the primary belief norm. However, dispositional norms are satisfied because the inhabitant of a sceptical world is *the* sort of person who would normally know the proposition (p.13). A derivative norm is that one is disposed to knowing the demon world because one does know that same proposition in normal worlds. Subjects in the demon world have excuses because they are *disposed* to know, on Williamson's view, but their beliefs are not justified.

Williamson argues that derivative norms are satisfied in a sceptical world because of the abnormality of the world, and the fact that one's disposition would be manifest in ordinary circumstances (p. 13). Williamson observes that they are disposed to knowledge because, in *normal circumstances* (normal worlds), they comply with the knowledge norm. The victim of the sceptical scenario would *usually* know (i.e. if he did not inhabit a very *abnormal* situation). If the derivative norm is satisfied, if one has the disposition to know normal circumstances, and if knowledge is the primary norm for justification, then this derivative norm, on Williamson's view, that explains the case as an instance of excusable but unjustified belief.

The norms for justification in the demon world, as Williamson (2007) emphasises, need not require that the beliefs are *successful* in the sense of being truth conducive. For Williamson, the norm for epistemic justification is a *success norm*; that is, it is a truth-conducive norm which is satisfied for a given belief only if that belief is a case of knowledge. The moral of the demon case for Williamson is that "excusable failure is not normatively equivalent to success" (2007,p. 37.).

Similarly, the view I have defended links justification to success. For one must at least use a reliable method of belief formation to have internally justified beliefs. There is, however, a difference between Williamson's knowledge norm and the internal condition for justification I have supported. Specifically, the knowledge norm does not admit of cases of knowledge failure as instances of justified belief. By contrast, the internalist condition supported in Chapter 2 and 3 of this thesis allows that internally justified beliefs may still sometimes be false and therefore not cases of knowledge. Therefore, the internal condition is a *truth-related norm*, but a *non-factive truth related to the factive norm*. Specifically, the relationship between believing that p needs to be *related* to the truth that p because one's belief that p needs

to result from a reliable method of belief formation; but the suggestion for the account I have been advancing is that the connection between belief and truth is not inviolable.

Appealing to Williamson's distinction between primary and derivative norms, we might ask: what kind of secondary norm is derivable from the primary norm I have in mind (the truth-related, non-factive norm), which might be satisfied in the demon case? Again, we may appeal here to the fact that if the primary norm is the norm which is satisfied when one meets the internalist standard of justification (I.e. when one's belief is reliably formed, and one is *aware* that that is so), then a corresponding derivative norm is just being *disposed* to conform to the primary norm. One may be disposed to confirm if in normal circumstances one would conform. We may appeal here to the fact that in the demon world one is in an environmentally abnormal situation. Someone in the demon world will satisfy the derivative norm (in believing that p) to the extent that they are internally justified (and therefore reliable) in the actual world.

§ 2.2. Exemptions

As indicated in the last section, there is scope to understand beliefs in the demon world as *excusable* while still unjustified. What I think one now needs is some means of distinguishing between beliefs which are blameless because they are *excusable* and beliefs which are blameless for other reasons.

To see this further, let us again consider the case of the Paranoid cognisor (assume the construction from Chapter 3, §, 3, and further that the believer is not *free* to have believed otherwise) The case of Paranoid was an example of a blameless but *intuitively unjustified belief*. I wish to again contrast Paranoid with a belief which seems justified in the actual world, specifically the following proposition Hands ("I have hands"), where this belief satisfies the internal condition for justification.

Two points: first, the last section indicated: Hands is excusable but not justified in the demon world. Second, let us imagine – for argument's sake – that Paranoid is also excusable. In which case, we would have the following verdicts:

> (1) Paranoid is excusable in the actual world
>
> &
>
> (1*) Paranoid is excusable in the demon world

Furthermore, of Hands we have the following verdicts:

(2) HANDS is justified in the actual world

&

(2*) HANDS is excusable in the demon world.

Given the above verdicts, then following would be true:

(1*) PARANOID is excusable in the demon world

&

(2*) HANDS is excusable in the demon world.

However, surely there is an intuitive a contrast between (1*) and (2*)? Even though both are blameless, case (1*) seems normatively impoverished, whereas case (2*) has more going for it – normatively speaking. It is hard to see why having an excuse should be subject to this degree of modal luck. Again, the mere fact we inhabit one world rather than another phenomenally identical world, should make the difference between whether or not our common-sense beliefs *differ* normatively to paranoid beliefs.

What we need to say is that while both cases (1*) and (2*) are *blameless*, there is a sense in which (1*) involves the believer *having an excuse* rather than being *merely blameless*. In the demon world, PARANOID, it *seems, is merely blameless*, but HANDS is blameless in some *other* sense.

I will support the case (consistently with Clayton Littejohn, 2009) for distinguishing between excuses and exemptions to explain our comparative intuitions about the demon case. In particular, it seems that in virtue of not a normative defeater and not being blameworthy but *also* in virtue of having a truth conducive disposition (I.e. having a belief which would, in normal circumstances, be true) HANDS is an epistemically excusable belief in the demon world, as Williamson's view predicts. Conversely, the blameless believer in PARANOID lacks a truth-conducive disposition in the demon and non-demon worlds and may, therefore, be *exempted* from blame rather than *excused*.

Williamson's idea that satisfying a derivative norm may be sufficient condition for an *epistemic excuse*. Derivative norms may explain why beliefs in the demon world are excusable but not justified. Analogously, the competent surgeon who unsuccessfully performs a difficult operation is blameless because he is *disposed* to

successfully performing an operation. By contrast, his incompetent counterpart (imagine an imposter, disguised as a surgeon, who has never operated in his life) is not *disposed* to successfully operating; he is neither excused nor blameless. Unfortunately, however, the case of the paranoid believer is neither the same as the competent surgeon nor the incompetent surgeon. In particular, it does not seem that the paranoid cognisor has an *excuse* even though she is blameless.

I suspect the above difference is one between having an excuse and being exempt (following Littlejohn, 2009). When we consider an exemption, there are typically *mitigating circumstances* which mean that the agent could not *reasonably* be expected to have acted or believed otherwise, and should not, therefore, be *blamed*.

We may consider an analogy with action to further characterise this distinction between excuses and exemptions. In particular, we need an analogy with excusable and unjustified action. First, consider INTRUDER: in self-defence, S kills an intruder who has broken into her home, where killing the intruder was the only way in which S could save her own life. In INTRUDER, the action is *justified* because the action was in self-defence. The subject is excused from blame because the action was justified. However, INTRUDER seems to be an instantiation of the principle that if one has an excuse for acting, then one was justified in performing that action. This is not the analogy between belief and action which we want.

Clayton Littlejohn (2009) provides a more promising analogy. Littlejohn understands cases of excusable but unjustified belief as similar to cases of "imperfect self-defence". Imperfect self-defence is when a person harms someone in the mistaken belief that they were acting in self-defence. Such a person is not to be *blamed;* nevertheless, it plausibly remains the case that they *ought not* to have acted. Such cases, on Littlejohn's view, are like cases of insanity because there is a sense in which the agent *ought* to have acted otherwise, and they are like exemptions because the agent is not blameworthy in the way that they acted. The case differs from a case of insanity because in cases of imperfect self-defence, the agent is "rational and responsible" (p. 432). The agent has an excuse for her action, but it was nevertheless unjustified. Analogously, the believer in the demon world has an unjustified belief which she should be excused for having performed.[110]

---

[110] On excuses, Littlejohn writes: "To say that someone is to be excused from criticism (and not exempted) is to describe the agent as being reasonable while remaining neutral on the further question as to whether they ought to have acted or believed in ways other than they actually did. The agent who is to be excused is unlike the agent who is to be exempted insofar as the agent is rational and responsible. If we judge of such an agent that she failed to do what she ought and wish to remove blame, we have to think of her response to the demands she was under as being other than they ought to have been, but reasonable".(p. 430). An agent who is

So far, I have discussed ways in which an account of justification which makes justified belief depend on externalist conditions can explain a class of intuitions, without implying that all beliefs in the demon world are merely blameless. On this premise, the agent in the demon world with blameless beliefs has an excuse, provided she is successful in normal circumstances and is not exempt from blame in normal circumstances. I have argued that this distinction may allow us to explain normative contrasts between different kinds of beliefs in the demon world. In particular, I sought to eliminate the modal luck that in the actual world, our putatively true beliefs are justified, but that in demon worlds they have the same normative status beliefs which are intuitively unjustified in the actual world.

There is still one claim which has not been met by the response. Specifically, the robust claim that empirical beliefs in the demon world are epistemically justified.

In response to this claim, we should note that it attaches significant evidential weight to the intuition beliefs in the demon world *are* justified. However, the view that they are justified is controversial. For example, we can observe that in a wide range of cases, failure to be reliable can undermine one's justification (e.g. cases such as FLAT-EARTHER considered in Chapter 1). Similarly, the fact that the paranoid believer is unreliable in his belief-forming method seems to count against the view that the belief is epistemically justified, and the same intuition can be applied to the demon case.

On the above point, recall from § 1 that the intuition that beliefs in the demon world are justified needs to be balanced against the intuition that there is also a normative intuition that there is a normative difference between demon and non-demon worlds. The skilled surgeon Audi (2001) describes, having done all that could reasonably be expected of him, but nevertheless failing to save the patient is of course blameless. But should he be commended in the same way as the surgeon who successfully carries out the procedure? Williamson asks that question and I think rightly concludes that the normative picture is more complicated than this.[111]

When one draws upon a class of normative intuitions as we have done – e.g. that the demon case is blameless, should not be faulted, has done everything that could reasonably be expected, etc, then denying that the subject is blameless would come at an exorbitant price. Furthermore, when one looks to our comparative intuitions – that there are normative differences between different kinds of blameless beliefs – an account of justification which cannot explain those normative contrasts

---

excusable in having ϕ'd *and* who cannot be exempted in having ϕ'd is such as to have been reasonable in having ϕ'd (Littlejohn, 431, *the externalist's demon*).

[111] Williamson (2007), p. 116.

would be an impoverished account. Nevertheless, the strategies outlined here both accommodate the blamelessness of beliefs in the demon world and the normative contrasts between different beliefs in the demon world.

What conclusion can reasonably be drawn? First, we have seen that our intuitions regarding justification in the demon world are not uniform. We have seen (in § 1.1) that, while there is an intuition that the demon problem describes an instance of justified belief, this should be balanced against the competing intuition that justified belief is linked to success, either through knowing the proposition or, as I have suggested, by using a propitious method of belief formation. Second, we have seen that content externalism poses a challenge to the Demon Problem. If the content externalist challenge is indecisive, as conceded, then we have examined ways in which the view that belief norms are tied to *success* may accommodate the possibility that beliefs in the demon world are excusable. Finally, we have seen a normative contrast between different believers in the demon world, and we have observed that by distinguishing between an excuse and an exemption, we may explain that contrast. Having an excuse, on this view, is more than having a merely blameless belief, but less than having a justified belief. The reasonable conclusion to draw here is that the Demon Problem is an inconclusive basis to abandon the type of internalism defended in previous chapters.

Chapter 5

Methods of Belief Formation and Anti-Luck Conditions

The aim of this chapter is to offer a more detailed account of belief-forming methods and of anti-luck conditions for knowledge. I argued in previous chapters that justified belief requires the use of a reliable method and that knowledge further requires safety. Questions arise for both claims: how are methods to be understood? Why might one support safety over other principles, such as sensitivity? I will address both questions.

The chapter is structured as follows:

§ 1. Distinguishes between narrow and broad methods.

§ 2. Explains the safety condition and distinguish it from the sensitivity condition. It describes how we understand methods of belief-formation will affect which propositions we safely or sensitively believe.

§ 3. Explains the epistemic closure principle in the context of safety and sensitivity. It explains how safety, but not sensitivity, retains closure, and I present this as an advantage of safety over sensitivity.

§ 4. Examines non-standard versions of both safety and sensitivity and argues in support of a weak safety condition for propositional knowledge.

§ 1. Internal Justification and Methods

For awareness internalism to be plausible, we need to explain the reliability condition identified in Chapter 1 (§3) as a core component of an internally justified belief. What first needs articulating is the nature of that reliability.

On the present account, one's belief must be formed via a reliable method (M) of belief formation for it to be justified and one must further have an awareness (occurrent or non-occurrent) that M is reliable. Based on this, facts about justification

(for example, the truth-value of a statement such as "S has a justified belief that p") are, at least in part, determined by whether M is reliable.

1.1. Narrow Methods

But just *how* generally should M be construed? The widest construal of M holds that M is a method such as "sensory perception". By contrast, on the narrowest construal, M is individuated by facts which are very specific to the circumstances in which one is forming a given belief, which means there is not any single method of sensory perception.

To expand on this, let the reliability of a method be its truth-ratio, that is, the proportion of true beliefs relative to false beliefs which the specified method produces over the long term. (Reliability is being used here as a statistical notion, as employed by Goldman (1979) which is not equivalent to the modal notion of reliability). As Goldman observes, a method might be individuated so narrowly that only one instance of that method ever occurs, then, on the narrowest construal of methods, M will be either 100% reliable or 100% unreliable (ibid, p. 12).

The narrowest individuation is problematic. The first reason against the narrowest individuation is that it would not allow for justified false beliefs. The reason for this is because if a necessary (albeit insufficient on our account) condition of having a justified belief that p at time *t* is that one's belief that p is formed via some reliable method M, and if there is only ever one instance of M, then M is reliable only to the extent that it produces a true belief that p at t. On the narrowest definition, if M produces a true belief at t, then M is 100% reliable at *t*; if M does not produce a true belief, then M is 100% unreliable. On the narrowest individuation, therefore, there is no scope for justified false beliefs. The narrowest individuation thus puts pressure on the independence between the truth and justification conditions for knowledge (see Zagazebski, 1994, on that independence). To the extent that we may have justified false beliefs, the narrowest individuation of methods would not explain that intuition.

The second reason for finding the narrowest individuation problematic concerns the role which methods play in subjunctive conditionals, such as in safety and sensitivity. To illustrate this, imagine that in the actual world α, S forms a true belief that p at time *t* via M, where M is individuated so narrowly that M is 100% reliable at *t*. Now imagine a close possible world β which differs slightly to α but in which S forms a belief that p at *t*. On the narrowest individuation of M, the method M that S uses at *t* in β would differ from the method S uses to arrive at a belief that p at *t* in α; this poses a challenge to safety and sensitivity to the extent that we are wanting

to relativize the subjunctives to the method of belief formation which one is using in the actual world (a point I consider in § 3). Thus, if there are no close possible worlds in which one uses the same method as in α, then providing that one has a true belief that p in α via M, one's belief that p in α will trivially satisfy safety and sensitivity.

To understand the above consideration, the reason sensitivity (the subjunctive $\sim p \rightarrow \sim Bp$, when relativized to methods) will be trivially satisfied on the narrowest individuation of methods because the antecedent for the sensitivity condition would be an impossible antecedent (the antecedent, in this case, would "be ~p & S were to use M to arrive at a belief as to whether p"). The antecedent would be impossible because, on the narrowest individuation, the counterfactual situation is one in which one uses a different method to that which S uses in actuality. Similarly, for safety, if there are no close possible worlds in which S uses the same method M in believing that p which S uses in α, then trivially there are no close possible worlds in which S falsely believes that p via the same method as in the actual world.

The narrowest understanding of methods therefore poses a challenge if we want to allow that one can have justified false beliefs using a reliable method and if we want to relativize subjunctive conditions for knowledge to the method of belief formation which a believer uses in actuality. By contrast, let a "broad method" be any method which is not narrow and that can be used in counterfactual situations. Goldman opts for a broader notion of methods than the narrowest individuation but confesses that he cannot give a "precise explication of our intuitive principles" (Goldman, 1979, p. 12), and neither can I. However, I think we can examine a broader notion of methods and find it to be plausible for a wide class of cases.

Importantly, Goldman focused on cognitive belief-forming procedures rather than methods. According to Goldman, a cognitive belief-forming procedure generates a mapping from certain states onto other states. On this view, a process has an input and an output, where the output is the production of a belief token. He allows that the "inputs" may include external stimuli in one's environment, but may also include other beliefs, desires, hopes, and memories (p. 11). Goldman emphasizes that a "belief-forming process" occurs at the cognitive level – on Goldman's view, processes may be restricted to cognitive events within the nervous system of an organism. By contrast, the notion of a "method" need not be restricted in that way, as factors outside of one's nervous system may be relevant to how a method is individuated. The notion of a "method" is not equivalent to the notion of a "process", though the distinction between methods and processes is, admittedly, not easy to draw.

Furthermore, we may take methods to be *ways* in which one forms beliefs. If methods are individuated externally, then the method will be individuated by

environmental features. By contrast, according to the internalist notion of methods, how a method is individuated will be a function of how things appear, from the perspective of the believer using that method. Nozick (1981, pp. 184-185) opted for the latter internal view of methods, claiming that a method "has its final upshot in experience".

I suspect it is possible to come up with a list of methods – or ways – of coming to believe a given proposition. Some of these ways are very broad, including visual and olfactory perception, and hearing, touch and taste. These general categories cannot exhaust the list of methods, since many of our beliefs about the world are arrived at through other sources, such as testimony, and memory.

The advantage of defining methods broadly is that we would retain the idea that a method does not always need to result in true beliefs. Note, here, that if we wanted to use justification as a condition for excluding veritic luck, then the use of one of these broad methods would not allow one's belief to be free from veritic luck, and the formulation of awareness internalism I have supported would not eliminate veritic luck. However, in this present project, I have been using safety rather than justification to exclude veritic luck. Chapter 2 (§ 4) supported the view that epistemically justified beliefs may still be veritically lucky.

The next two sections will continue to examine methods but will situate methods in the context of the safety and sensitivity conditions. The remainder of this chapter will first explain (in § 2) safety and sensitivity in the context of benign and malign epistemic luck, defining these two notions. I argue that Nozick's adherence condition (p→Bp) is not necessarily needed to eliminate veritic luck, because it targets an arguably benign type of luck called evidential luck. I consider how safety and sensitivity may yield different results for our knowledge of the negation of radical sceptical hypotheses. In § 3, I explain how sensitivity violates closure and how safety retains the closure principle, and I examine why this is a difficulty for sensitivity. In §4, I argue that there are counterexamples to the necessity of sensitivity as a condition for knowledge. I support safety over sensitivity, but I argue that a weak safety condition is needed to avoid counterexamples.

§ 2. Safety and Sensitivity

I wish to start by considering the very basic versions of both safety and sensitivity.

The first condition is Nozick's sensitivity condition (1981, p. 176). Sensitivity holds that S's belief that p is sensitive only if *were it the case that p is false, then S*

would not believe that p (~p→~Bp). Sensitivity condition is now, usually, described as the condition (~p→~Bp).

However, we need to be careful to distinguish between the sensitivity condition and the condition that one knows only if one *tracks the truth*. For Nozick (1981, p. 178), tracking the truth requires satisfying the sensitivity condition *and* what he terms an *adherence* condition (p→Bp).[112] The adherence condition is the condition that if *p were true*, then *S would believe that p*. Nozick explains that the condition (~p→~Bp) "tells us only half the story" about how a belief may be sensitive to its truth value (p. 176). On Nozick's view, the condition (~p→~Bp), tells the other half of the story is how a belief can be sensitive to a proposition's truth. Notwithstanding, sensitivity is usually characterized as (~p→~Bp). Since Nozick does not give the condition (~p→~Bp) a unique name, I will adopt the modern usage, where sensitivity is the condition (~p→~Bp), and adherence is the condition (p→Bp). Here are Nozick's two conditions (for a subject S and a proposition p):

Sensitivity: S's belief that p is sensitive just in case "If P weren't true, S wouldn't believe that p" (Nozick, 1981 p. 172)

Formally, **(~p→~Bp)**

Adherence: S's belief that p is adherent just in case "Not only is p true and S believes it, but if it were true he would believe it (Nozick, 1981, p. 172)

Formally, **(p→Bp)**

Nozick appeals to possible world semantics for determining the truth conditions of both conditions.[113] On David Lewis's view of counterfactuals (1973a; 1973b), a

---

[112] For example, Nozick (p. 178) writes that "A person knows p when he not only does truly believe it, but also would truly believe it and would not falsely believe it. He not only actually has a true belief, he subjunctively has one"
[113] See Nozick, 1981, p. 173.

counterfactual of the form $p \Rightarrow q$[114] is true just in case the closest possible worlds in which p holds are worlds in which q also holds. Using Lewis's account, sensitivity is as follows:

> **Sensitivity**: S's belief that p is sensitive just in case the closest worlds in which p is false are worlds in which S does not believe that p.[115]

The adherence condition does not use the same possible worlds framework as sensitvity. Notice that if we, therefore, appeal to Lewis's possible worlds framework for counterfactuals, then for cases in which one truly believes that p, adherence will be trivially satisfied. The reason for this is because the closest world in which the antecedent of the adherence condition (p) obtains is the actual world, and in the actual world one truly believes that p. To avoid triviality, Nozick appeals to the idea that there are *close* worlds beyond the actual world (p. 176). Adherence now becomes the following condition:

> **Adherence:** S's belief that p is adherent just in case in all the close worlds in which p holds, S does not believe that p.[116]

Nozick relatvizes both sensitivity and adherence to the method of belief formation that the believer uses in the actual world.[117] Method relativization holds that we should only examine possible worlds in which one uses the same method of belief formation as in the actual world. For Nozick, there are counterexamples to un-relativized conditions. One example is the grandmother who sees that her grandson is well, but at the closest world in which he is ill, her relatives would tell her that he is fine to spare her grief (p. 179). Notice that the closest ~p world is one in which the grandmother would be using a different method of belief formation (testimony) to that which she uses in actuality (namely, *seeing him*). In what follows, I allow that the most

---

[114] Throughout this chapter, let ⇒ represent a subjunctive conditional (whether that be a counterfactual subjunctive or a non-counterfactual subjunctive) rather than a material conditional.

[115] Nozick does not formally state this. See p. 173 for his statements.

[116] Nozick does not tell us how many close worlds, but the suggestion that "he believes that p for some distance out in the p neighbourhood of the actual world" (p. 176), suggests that it is the stronger notion that he has in mind – i.e. that any close p&Bp world makes one's belief not adherent.

[117] Sensitivity becomes (for a method M), "If p weren't true and S were to use M to arrive at a belief whether or not p, then S would believe, via M, that p" (Nozick, p. 179). Adherence becomes: "If p were true and S were to use M to arrive at a belief whether or not p, then S would believe, via M, that p" (ibid.,).

defensible notion of sensitivity is Nozick's version, relativized to the belief of belief formation. Notwithstanding, I think it is doubtful that sensitivity is a necessary condition for knowledge.

§ 2.1. Benign Luck

Before comparing safety, sensitivity, and adherence, we need a notion of what kind of luck they target, respectively. Chapter 2 argued that some species of epistemic luck are benign, or at least compatible with knowledge with a lower grade of knowledge (see Chapter § 4 for the distinction between minimal and non-minimal knowledge).

However, what other kinds of epistemic luck are there, other than veritic epistemic luck and reflective epistemic luck? There is a type of epistemic luck that Engel calls *evidential luck* (1992a). Given what Engel writes about the notion, we may state evidential luck as follows:

> **Evidential luck**: S's true belief that p is evidentially lucky just in case it is a matter of luck that S has evidence for her belief that p. (Engel, 1992a).[118]

Evidential luck and veritic luck do not seem equivalent. For instance, imagine one momentarily glances at the sky to check the weather and truly believes that there is a jet aircraft flying above Oxford, forming a true belief in the proposition based on evidence. It is a matter of luck that one has evidence for this proposition – if one had looked southwards rather than northwards to check the weather, then one would not have formed a true belief. Nevertheless, given the fact that one *is* looking northwards, it does not seem a matter of luck that the belief is true. Furthermore, when we keep fixed the initial condition (i.e. that one *is* looking northwards) in the actual world, then – to the extent that one *knows* the proposition - the belief seems both safe and sensitive. For example, the closest worlds in which there is not a jet overhead are not worlds in which one falsely believes that there is. The belief does not, therefore, seem veritically lucky or a case of knowledge failure, even though it is evidentially lucky.

Engel maintains that evidential luck is benign, and he considers Harman's example of DICTATOR to illustrate why. In Harman's DICTATOR example (1973), Harman imagines that Jill buys a newspaper and reads that the dictator of her country has been assassinated, and she then forms a true belief in this proposition. We are then

[118] Engel defines this notion as follows, for a person who is evidentially lucky: "A person who is epistemically lucky in virtue of the fact that she is lucky to be in the evidential situation she is in" (p. 67)

to suppose that on national television it is announced that the assassination attempt has failed. Finally, we are to suppose that a reporter on the scene contacts the newspaper, which prints the real story in its final edition. Harman argues that Jill lacks knowledge because "everyone else has heard the televised announcement", claiming "it is highly implausible that Jill should know simply because she lacks evidence everyone else has" (p. 144). Engel observes that Jill's belief is evidentially lucky since it is a matter of luck that Jill has the evidence which she does have (1992a, p. 66). Notice that Jill's belief is sensitive because at the *closest* world in which the dictator is not assassinated, the closest ~p world, is one in which the real story would not have been printed in the final edition, and Jill would not have believed that p.

We can see how Nozick's adherence condition predicts that DICTATOR is a case of knowledge failure. Specifically, there are *close* possible worlds in which p obtains and in which Jill watches the televised announcement and comes to believe that the assassination attempt has failed. If Harman is correct that Jill fails to know, then that would motivate the idea that adherence is a necessary condition for knowledge. Indeed, Nozick uses the example of JILL to motivate adherence (p. 177).

Conversely, if Harman's DICTATOR example is understood as a case of knowledge possession because evidential luck is compatible with knowledge, then that suggests that we may not need adherence as a condition for knowledge. Notice that the difficulty of the DICTATOR example is the intuition that Jill lacks knowledge is controversial once we further consider Jill's situation. Note, here, that what is crucial to THE DICTATOR example is that everyone else who hears the televised announcement has *false* or *misleading* evidence. In which case, is it plausible that Jill should fail to know because she does not have false or misleading evidence? Is it not more plausible to say that Jill *almost* failed to know, but that, nevertheless, she *does* know?

Engel's claim that DICTATOR is an example of knowledge is plausible because the luck at stake is evidential rather than veritic luck. Just as it is benignly lucky that one can form a true belief about the jet flying overhead because one happens to look northwards instead of southwards. What we need to bear in mind when considering the example, is that if Jill did, *in fact,* read the final edition and she did not base her belief on misleading evidence. She almost failed to know, rather than failed to know.

Notice, finally, that we do not need the notion of adherence to explain typical Gettier case and example of veritic luck. Recall that the stopped clock example from Chapter 2 was explained as a safety failure and notice that we can explain that case as a sensitivity failure also – because the closest world in which the proposition is false is one in which the believer continues to falsely believe the proposition. Adherence, it seems, plays at most a marginal role in targeting veritic luck. Ultimately, I suspect this

leaves us with the two most plausible anti-luck conditions - safety and sensitivity. For many cases, there is not much to decide between them, as they yield the same predictions for a wide range of cases. However, I will argue that there are difficulties for sensitivity. These difficulties are:

1) Sensitivity but not safety predicts that we do not know the negation of radical sceptical hypotheses.

2) Sensitivity but not safety fails to retain the principle of epistemic closure.

3) There are counterexamples to sensitivity that support safety.

The rest of this chapter will be structured around explaining those difficulties for sensitivity and considering how sensitivity may respond to them.

§ 2.2. Sensitivity and Sceptical Hypotheses

Safety and sensitivity yield different implications for our knowledge of the negation of radical sceptical hypotheses (take, as paradigmatic, the hypothesis that one is *not* a disembodied brain-in-a-vat), henceforth ~BIV. The standard view is that safety is compatible with knowing that ~BIV ("K~BIV"), but that sensitivity is incompatible with K~BIV. I will explain why safety and sensitivity have different predictions for K~BIV, before examining the details of Nozick's proposal.

Consider, first, the safety condition. And consider the proposition HANDS (I have hands). To the extent that you know that HANDS, there are no close possible worlds in which you falsely believe HANDS, and so your belief is safe. Note here that we must focus on close worlds in which you believe HANDS and ask whether HANDS holds at those worlds. For example, there are more distant worlds in which ~HANDS. Yet, intuitively, and to the extent that you *know* that HANDS, it seems that ~HANDS worlds are distant. Importantly, there *are* radical sceptical worlds in ~HANDS holds & in which one falsely believes that HANDS holds. Specifically, the world in which one is a disembodied brain-in-a-vat (BIV), but on the standard ordering of possible worlds, such worlds are radically dissimilar to the non-sceptical worlds in which HANDS is true; the BIV worlds do not seem close.

Conversely, let us consider the sensitivity condition. There is an essential difference between safety and sensitivity here: for safety we are keeping fixed one's belief in a

proposition *p*, and assessing whether one's belief that p is true in close worlds. By contrast, for sensitivity, we are varying the truth-value of p at the closest worlds and are asking whether one doesn't falsely believe that p at the closest ~p worlds. For sensitivity, we are interested in the *closest ~p worlds* however distant they are. For many putatively true propositions which we know, the closest ~p worlds will be close, and in these cases one often (though not always) safely *and* sensitively believes that p. However, for some propositions, the worlds in which they do not obtain are *very distant indeed*. This is what happens in radical sceptical scenarios when the proposition at stake is the negation of a sceptical hypothesis. For example, if the proposition which one believes in the actual world is the proposition that one *is not a disembodied brain in a vat* (~BIV), then the closest possible world in which this proposition is false (and in which the antecedent of the sensitivity condition is satisfied) is a world in which one is a brain-in-a-vat. Sensitivity predicts that one's belief that ~BIV is insensitive because the closest world in which the proposition is false is a BIV world in which one falsely believes that ~BIV.

Nozick's View of Methods

While I have outlined the basic explanation of why sensitivity yields a different result to safety, the situation is in fact more complicated once we turn to the details of Nozick's proposal. Recall that Nozick relativizes sensitivity to the method of belief formation which the agent uses in the actual world. On this view, for a subject S and a proposition p, we should hold fixed the method (M) which S employs in the actual world, such that (for any p) if the closest ~p worlds are worlds in which S uses a different method to M. If, in these closest ~p worlds, S falsely believes that p, then we would need to move further out - modally speaking - to examine S's doxastic state at the closest ~p worlds in which S uses the same method of belief formation which S uses in actuality.

Nozick hopes to secure the implication that one doesn't know ~BIV by arguing that one's belief that ~BIV is insensitive because it does not satisfy that method-relativized notion of sensitivity. Nozick's view is that in both these worlds we are using the same method of belief formation. The reason for this is because Nozick tells us that a "method has its final upshot in experience on which the belief is based", and that two methods which are "the same from the inside" are the same methods (Nozick, 1981, pp. 184-185). By contrast, an externalist view of methods maintains that the identity of the method is determined by the environmental conditions in which one uses that method. On an externalist view of methods, since the sceptic's possibility is

far removed – environmentally speaking - from the non-sceptical world, we would be using a different method in both sceptical and non-sceptical scenarios.

The internal notion of methods is controversial. Indeed, Nozick's view of methods is, perhaps, one of the more surprising elements of his account of knowledge. For example, Michael Williams (1996) reminds us that, on Nozick's view, "one can use a method without awareness of what method one is using since that is a requirement of his subjunctives being externalist in the relevant sense" (p. 343). However, given that one can use a given method without being aware *that* one is using a given method, it is - as Williams describes – "unclear why experiential factors fix the identity of the method" (Williams, p. 343). However, there is also a sense in which the method of belief formation which one would be using in radical sceptical worlds is *not* the same method of belief formation which one uses in non-sceptical worlds.). In particular, we would be using an unreliable method in the BIV world and a reliable method in the ~BIV world (per Chapter 4).

Necessary vs. Sufficient Conditions

In what follows, let us abbreviate "*one's failure to know that ~BIV*: as follows: **~K(~BIV)**. I want to assume that sensitivity can explain ~K(~BIV). However I will explain why this arises as a difficulty for sensitivity, if sensitivity is used as an anti-luck condition.

If the sensitivity condition implies ~K(~BIV), then it is vulnerable to the objection that intuitively *we do know that ~BIV* ("**K~BIV**"). The neo-Moorean view - in the spirit of G.E. Moore's pronouncements - maintains one can know that ~BIV. One neo-Moorean argument is that since we know that p (for a putatively true p) and we can deduce that p being true entails that ~BIV, then we can come to acquire knowledge that ~BIV on the basis of knowing p and competently deducing that ~BIV. An analysis of neo-Mooreanism is beyond the scope of this thesis. However, there is an epistemic intuition that we *do* know that ~BIV that sensitivity theorists should need to explain.

The situation for sensitivity seems especially problematic once we bear in mind that we are examining the *necessity* rather than the sufficiency of the safety and sensitivity conditions for knowledge – specifically, when considering both conditions as methods of eliminating veritic epistemic luck. If the issue of which of the conditions to support concerns the *necessity* of safety as a condition for knowledge, then the neo-Moorean intuition that K~BIV seems more problematic for sensitivity than the converse intuition that ~(K~BIV) is for safety. The reason for this is because, on the safety conception, it is still possible that one fails to know that one is not-a-brain-in-

a-vat on *some other basis* (for example, that one fails to know ~BIV because one's evidence for this proposition is underdetermined). In sum, the sceptical intuition [~(K~BIV)] challenges only the *sufficiency of safety*. By contrast, the neo-Moorean intuition [(K~BIV)] challenges the very *necessity* of sensitivity.

Nozick (1981, pp. 227-230) is aware of the Moorean intuition and seeks to address it. Nozick addresses it by his insight that while one's belief that ~BIV fails the sensitivity test, we can still know conjunctive propositions which include, as at least one of their conjuncts, a proposition about the negation of a radical sceptical hypothesis. Nozick explains that we can *track* (and sensitively believe) conjunctive propositions. To understand this, let the proposition HANDS again denote "I have hands" and let "~BIV" continue to denote the proposition that one is not a disembodied brain-in-a-vat. The relevant conjunct is now the proposition that [HANDS & ~BIV]. Notice that one's belief that [HANDS & ~BIV] is sensitive. The reason why this is so is because the closest worlds in which the conjunction is false are worlds where HANDS is false (e.g. worlds in which they are lost) rather than worlds in which one is a BIV; and in these closest worlds, one does not believe the conjunction [HANDS & ~BIV] because one does not believe the first conjunct (HANDS) of that conjunction.

I will explain in the next section how safety but not sensitivity is consistent with the closure principle, and that since closure is attractive, this arises as a concern for sensitivity. I suggest that tracking conjunctive propositions does not, necessarily, ameliorate that concern.

§ 3. Safety, Sensitivity and Epistemic closure

Nozick's sensitivity condition abandons the epistemic closure principle. Conversely, safety is consistent with the epistemic closure principle. It has been argued that the epistemic closure principle is intuitively plausible (DeRose, 1995). If closure is plausible, then that counts in favour of safety condition which retains it rather than sensitivity that abandons closure (See: Sosa, 1999).

I will characterize epistemic closure. First, I will explain how sensitivity and safety differ concerning epistemic closure; I will further consider whether a method-relativized version of sensitivity, but where methods are individuated externally, can retain closure. Second, I will explain why closure is attractive.

The epistemic closure principle may be stated as follows:

EC: for some subject S, and for two propositions p & q, if S knows that p and S knows that p entails q, then S knows that q.

On this view, S will fail to know that p if she does not know proposition q which she knows (by deduction) to be entailed by p.

I wish to separate this from a related principle, which we may call transmission. Transmission has been discussed in the literature in the context of providing warrant for a cogent argument, where a warrant for a conclusion is *transmitted* from its premises to its conclusion when it would allow an interlocutor to overcome agnosticism about that conclusion (Wright, 2002). The notion of warrant may be related to closure, but at root, the idea of transmission is the ability to *acquire* or *extend* one's knowledge by knowing that p, and knowing that *p* entails q, and *therefore* knowing that *q*. Transmission is the idea that competent deduction allows one to expand one's knowledge. One way in which transmission may differ to epistemic closure principle is that closure may be satisfied in cases where one knows that *q* on some basis which is independent of one's knowledge that *p*, and one's knowledge that *p* entails *q*.

We may understand the epistemic closure principle (EC) as a view about the distribution of epistemic properties. More specifically, EC places a constraint on the truth-value of attributions about knowledge. The epistemic property for EC is knowledge. Specifically, EC implies that we would speak falsely in attributing, to S, knowledge that p and knowledge that p entails that q, while denying knowledge that q.

As Sosa (1999) demonstrated, the safety condition retains the epistemic closure principle for the beliefs under consideration. For example, if I safely believe and know that HANDS [I have hands], and I know that HANDS entails ~BIV, then I may safely believe and know that ~BIV. As we have seen, one safely believes that HANDS and safely believes that ~BIV, consistently with closure. By contrast, sensitivity denies the epistemic closure principle (see Nozick, 1981, pp. 197-211). The epistemic closure principle is violated for sensitivity because one can know and have a sensitive belief that HANDS, know that this entails ~BIV, and yet fail to know that ~BIV because one's belief that ~BIV is insensitive.

Is there a version of sensitivity consistent with the closure principle? There may be if there is a formulation of sensitivity which retains our knowledge that ~BIV. Let us now consider the possibility of sensitivity retaining closure.

Recall that Nozick relativizes his sensitivity condition to a method of belief formation M. However, Nozick argues that one uses the same method of belief formation (for any empirical p) in the sceptical world as in the non-sceptical world; this was because Nozick defines methods internally as having a "final upshot in experience"(Nozick, 1981, pp. 184-185).

However, if one were using a different method of belief formation BIV and ~BIV worlds, because they are externally very different situations, then there may be scope to retain closure because there may be scope to allow one to sensitively believe the negation of sceptical hypotheses. To understand why, this is because an externalist conception of methods makes the antecedent of the sensitivity subjunctive an impossible antecedent (if sensitivity is method-relativized). Furthermore, the sensitivity subjunctive will be vacuously satisfied if its antecedent is impossible, given the Lewisian view of subjunctive conditionals with impossible antecedents. Therefore, it is possible that sensitivity can accommodate our knowledge that ~BIV and may, therefore, accommodate the closure principle.

To examine this further, note that if there *are* no BIV worlds in which we are using the same method of belief formation (to believe that ~BIV) as we do in the actual world, then one's belief that ~BIV in the actual world will *trivially* be sensitive. Specifically, the subjunctive conditional would be vacuously true because it contains an impossible antecedent. For example, if we plug in the ~BIV proposition, sensitively believing that ~BIV requires the truth of the following subjunctive*:*

 "If ~(~BIV) and S were to use the method M which she uses in the actual world, then S would not believe that ~BIV via M."

However, on an externalist view of methods, the antecedent of that conditional is impossible. On the standard view, counterfactuals with impossible antecedents are vacuously true because they are not capable of being false (Lewis, 1973a, p. 26).

I suggest that an externalist view of methods will not necessarily allow sensitivity to retain closure. In particular, if the antecedent is impossible, it will be true that sensitivity is *not violated*, but it will also be no clearer that sensitivity is *satisfied*, to the extent that *satisfying* the condition requires that one does not believe that p at the *closest* possible worlds in which ~p obtains. Sosa (1999, p. 283) raises an analogous point when he observes, of safety, that only safety can accommodate the possibility of knowing necessarily true propositions. In the case of our beliefs in propositions which are necessarily true, we cannot entertain the possibility that the antecedent is false, and therefore we cannot expect discriminating intuitions about whether sensitivity is satisfied. Similarly, if we understand the sensitivity condition in this way, as vacuously true if methods are individuated external, then it would seem there are no firm modal intuitions about the belief's counterfactual responsiveness to its truth value.

Abominable Conjunctions and Closure

Why is closure intuitive? Keith DeRose (1995) highlights the implausibility of denying closure. DeRose writes that denying closure results in the "a*bominable conjunction that while you know that you have hands, you don't know that you're not a brain-in-a-vat*" (DeRose, 1995, p. 28 - my italics). As Sosa highlights, one of the advantages of a safety-based approach is that it avoids the abominable conjunction. To see this further, on Nozick's account, it would be correct to say that:

    I. [I know that (HANDS **&** ~BIV)],

    II. [I know that HANDS **&** I do not know that ~BIV].

Ascription (II) describes DeRose's abominable conjunction.[119] The conjunction is problematic because it seems counterintuitive that one could fail to know ~BIV in circumstances in which one knows HANDS alongside knowing the relevant entailment. By contrast, Sosa by replacing sensitivity with safety, the abominable conjunction is avoided. Could Nozick's appeal to the fact that we track conjunctive propositions to avoid DeRose's abominable conjunction? Recall Nozick's insight that while we cannot sensitivity believe that ~BIV, we can sensitivity believe conjuncts such as [HANDS & ~BIV], because the closest world in which the conjunct is false is a -HANDS worlds rather than a BIV world, and in this world (to the extent that one knows the conjunct), one does not falsely believe HANDS.[120] Tracking conjunctions may go some way to accommodating the neo-Moorean intuition (of. § 2.2), but it still leaves us with conjunctions such as the following:

    III. [I know that HANDS **&** I know that (HANDS &~BIV) **&** I do not know that ~BIV].

In ascription (III), the second conjunct is the one we can know by tracking conjunctive propositions. However, of ascription (III), it seems counterintuitive to say that one

---

[119] For DeRose (1995), the conjunction does not lead him to abandon sensitivity, but instead, to place sensitivity within a contextualist treatment of knowledge attributions, according to which in asserting the sceptical hypothesis one thereby raises the standards for knowledge to such a degree that we need to sensitively believe that ~BIV. But, in ordinary contexts, one can know HANDS and that this entails ~BIV, and one can know that ~BIV in ordinary circumstances. The closure principle is retained on this view. For DeRose, the act of asserting raises the standards to such a degree that one needs to sensitivity believe that ~BIV in order to know that ~BIV – given epistemic closure, one fails to know that HANDS in contexts in which the sceptic asserts her hypothesis.

[120] Nozick, (ibid, 227-230).

truly can be said to know both the first and second conjuncts, and yet fail to know the third conjunct.[121]

I wish to now examine counterexamples and whether there is a modal condition that can avoid them. To this end, the next section is a publication from the thesis supporting the view that safety rather than sensitivity can be differentially supported as the correct requirement for propositional knowledge (Neil, 2019).[122] In the process, we can observe that a weaker rendition of safety, one which requires truth-adherence at *most* rather than *all* close possible worlds is positioned to avoid influential counterexamples to sensitivity.

§ 4. Safety, Domination, and Differential Support[123]

In a recent paper "Safety, Sensitivity, and Differential Support" (*Synthese*, December 2017), Jose Zalabardo argues that (contra Ernest Sosa, 1999) sensitivity can be differentially supported as the correct requirement for propositional knowledge. Zalabardo argues that safety fails to *dominate* sensitivity; specifically: some cases of knowledge failure can only be explained by sensitivity. In this paper, I resist Zalabardo's conclusion that *domination failure* confers differential support for sensitivity. Specifically, I argue that counterexamples to sensitivity undermine differential support for sensitivity. Using Zalabardo's modal framework, I consider a less demanding modal condition, what I call *weak sensitivity*, and I explain how weak sensitivity avoids an influential counterexample to sensitivity. However, I argue that we can subvert that counterexample only by abandoning Zalabardo's case for *domination failure*. So either way, we cannot differentially support sensitivity.

The paper is structured as follows:

---

[121] I do not want to suggest that the closure principle is sacrosanct. For example, Dretske (1970) offers an account of the sentential operator "knows that" which is consistent with closure failing (see Dretske's zebra example, pp. 1015-1017). Dretske's example is of a visitor to a zoo who knows that she is looking at a zebra but fails to know she is not looking at a cleverly disguised mule. I suspect there is a safety response to Dretske's counterexample, one which holds that we do know the negation of the sceptical hypothesis Dretske describes. However, I will not pursue this case further here, since §4 indicates that sensitivity is wanting on grounds independent of closure.

[122] Neil, C. (2019). "Safety, domination, and differential support", *Synthese*. https://doi.org/10.1007/s11229-019-02091-2.
Department of Philosophy, University College London.

[123] I am grateful to two referees for this journal for their helpful comments.

§ 4.1. I foreground standard modal versions of the safety and sensitivity conditions. I retrace Zalabardo's argument that sensitivity entails safety.

§4.2. I argue that *domination failure* alone doesn't confer differential support for sensitivity; counterexamples to sensitivity affect differential support.

§4.3. I distinguish between *weak sensitivity* and *strong sensitivity*. I argue that safety fails to dominate sensitivity only if we endorse *strong sensitivity*, which succumbs to counterexamples which undermine differential support for sensitivity.

§4.1. Domination and Translation Keys

Consider two venerable conditions for (propositional) knowledge, the sensitivity and safety conditions. [124] Let us say that a subject S has a *sensitive* belief in some proposition p just in case:

"were it not the case that p, then S would not believe that p" [125]

Formally:

Sensitivity: $\sim p \Rightarrow \sim Bp$

(Throughout this paper, let "$\Rightarrow$" denote the subjunctive conditional and let "$\rightarrow$" denote the material conditional).

Robert Nozick (1981) regards sensitivity as a necessary condition for knowledge. In contrast, Ernest Sosa (1999) argues that sensitivity should be replaced with the following subjunctive condition, which he calls *safety*:

"S would believe that p only if it were so that p" [126]

Formally, Sosa writes safety as the contrapositive of sensitivity, as follows:

[124]Nozick also relativizes the sensitivity conditional to the method of belief formation "M" which S employs in the actual world to believe that p (p. 179). Similarly, safety has been relativized to the method or basis which one uses in the actual world (See: Sosa, 2007). In this paper, I will be concerned only with the un-relativized versions of sensitivity and safety, which are the focus of Zalabardo's paper.

[125] On an interpretive note, Nozick intended "sensitivity" to refer to the conjunction of (what I've called) the sensitivity condition *and* an adherence condition, namely $[p \Rightarrow Bp]$.(Nozick, p. 176). I will not be discussing adherence here. I shall use sensitivity to refer only to the subjunctive $[\sim p \Rightarrow \sim Bp]$.

[126]Sosa (1999, p. 142).

Safety: $Bp \Rightarrow p$

As Sosa reminds us, sensitivity and safety are not equivalent because subjunctive conditionals do not contrapose.[127]

Sosa argues that sensitivity cannot be differentially supported as the correct requirement for propositional knowledge. He does so in defence of neo-Mooreanism, the view that we can and do know the negation of radical sceptical hypotheses.[128]

Zalabardo summarises Sosa's argument for *the denial of differential support for sensitivity* as the claim that "every virtue that can be claimed for sensitivity as a necessary condition for knowledge would be matched by conferring this status on safety instead" (p. 2). Zalabardo argues that the denial of differential support requires the truth of two premises, the conjunction of which he calls *domination*:

> **(D1)** If a belief that has the status of knowledge is sensitive, then it is also safe.
> **(D2)** If a true belief that doesn't have the status of knowledge is insensitive, then it is also unsafe.[129] (p. 2).

More schematically:

> **(D1)** [Kp & (Bp)SENSITIVE] → (Bp)SAFE.
> And
> **(D2)** [~Kp & ~(Bp)SENSITIVE] → ~(Bp)SAFE.[130]

Domination fails if either **D1** or **D2** is false (for an inclusive disjunction). Zalabardo rejects *domination* by conceding **D1** and rejecting **D2**.

---

[127]Sosa (1999, pp. 149-150).

[128] Sensitivity, but not safety, is inconsistent with neo-Mooreanism; the view that we can and do know that radical sceptical hypotheses don't obtain (in the spirit of G.E. Moore's pronouncements (1918, 1925, 1939)). Take, for example, the hypothesis that one is a disembodied brain-in-a-vat (BIV). Notice that one's belief that ~BIV is safe but insensitive. For *not easily* would one falsely believe that ~BIV (hence safety is satisfied), but *if* BIV were true *then* one would falsely believe that ~BIV (hence sensitivity is violated). By replacing sensitivity with safety, Sosa blocks the argument against neo-Mooreanism based on the contention that one's belief that ~BIV is insensitive.

[129] Zalabardo (p. 2) qualifies the consequent: non-modal shortcomings may account for why a sensitive belief fails to constitute knowledge.

[130] D1 holds that *if* one knows that p (Kp) AND one's belief that p (Bp) is sensitive, *then* Bp is safe. D2 holds that *if* one fails to know that p (~Kp) AND Bp is insensitive, *then* Bp is unsafe.

I will now examine Zalabardo's argument for **D1**. **D1** is true because, Zalabardo argues, sensitivity entails safety. To understand why sensitivity entails safety, let us first examine how safety requires a different *translation key* to sensitivity.

What is a *translation key* for a subjunctive conditional? A *translation key* for a subjunctive conditional is a way of unearthing the truth value of that subjunctive. On David Lewis's view, for example, the subjunctive conditional "if it were the case that *p*, then it would be the case that *q*" is *true* just in case in the closest possible worlds to the actual world "α"[131] in which *p* obtains, *q* also obtains (Lewis 1973a,b). I will refer to this as the *standard translation key* (ST). More schematically:

> **ST:** $p \Rightarrow q$ is true just in case in all the worlds in which p is true that are
> at the shortest distance from α, q is also true.

Sensitivity theorists embrace **ST**. They say that S's true belief that p is sensitive just in case the *closest worlds* to α in which the antecedent is true (i.e. in which ~p holds) are worlds in which the consequent is true (I.e. in which ~B(p) holds).[132] However, if safety theorists adopt **ST**, they're in trouble. Their problem is this: if S has a true belief that p in the actual world α, then the closest possible world to α in which S truly believes that p *just is* α. The closest world to α = α, and therein lies the problem with **ST** for subjunctive conditionals with true antecedents (such as safety). Adopting **ST** for safety renders all true beliefs trivially safe.

Safety theorists, as Zalabardo observes, must use a different translation key to avoid triviality. Specifically, safety is usually made a function of what one believes at "close" possible worlds which lie beyond α. Let us call this the *non-standard translation key* (**NST**). Let us formulate this key as follows, where "*d*" denotes a sphere of *close* possible worlds:

> **NST:** $p \Rightarrow q$ is true just in case in every world in which p is true that is
> at a distance of *d* or less from α, q is true.

Zalabardo observes that by using **NST**, safety becomes the following non-trivial condition:

---

[131] Throughout this paper, let "α" denote actuality.
[132] See Zalabardo (p. 3).

> **Safety (NST):** S's belief that p is safe just in case, in every world in which S believes that p that is at a distance of *d* or less from α, p is true. (Zalabardo, p. 3)

At this point, Zalabardo argues that if we adopt **NST** for safety, then we should adopt **NST** for sensitivity too. Zalabardo reasons as follows:

> "it seems wrong to make the truth conditions of a subjunctive depend on the truth value of its antecedent. Take, for example, the subjunctive "if the currency were devalued, interest rates would go up." On the current proposal, if the currency is actually devalued, the truth of the subjunctive will require interest-rate rises, not only in actuality, but also in the range of conditions obtaining in all the worlds at a distance of *d* or less in which the currency is devalued, e.g. at a range of levels of taxation, inflation, etc., but if the currency is not devalued, it would suffice for the truth of the subjunctive that interest rates go up in the specific circumstances obtaining at the nearest worlds in which the currency is devalued, e.g. with the precise levels of taxation, inflation, etc. present in those worlds." (p. 5)

Zalabardo concludes:

> "I find this counterintuitive. If, in order to avoid making safety redundant, we look at a wide range of worlds to determine the truth value of the subjunctive, we need to apply the same approach to sensitivity." (p. 5)

Zalabardo argues that if we employ **NST** for sensitivity, then sensitivity entails safety. To understand why, let us reproduce Zalabardo's argument. Let CT(p) denote the distance from α of the closest possible world to α in which p is true. Notice that CT(p) = 0 for whenever p is true. Therefore, for actually obtaining beliefs, **NST** implies that a belief that p is safe just in case p is true in all worlds in which one believes that p up to *d* (p. 5). Conversely, adopting **NST** for sensitivity now yields the following condition:

> **Sensitivity (NST):** S's belief that p is sensitive just in case: in every world in which p is false that is at a distance of CT (~p) + *d* or less from α, S doesn't believe p. (Zalabardo. p. 5) [133]

On this construal, sensitivity entails safety. To see why, let us follow Zalabardo and understand safety and sensitivity as properties which exclude *worlds in which one falsely believes that p* – what Zalabardo terms *error worlds*. The only difference between safety and sensitivity, Zalabardo contends, is that safety but not sensitivity imposes a *fixed lower bound* on which error worlds knowledge excludes (the bound is fixed for safety because CT(p) = 0 for subjunctives with true antecedents). Safety prohibits the existence of error worlds up to distance *d* only. By contrast, sensitivity demands that at the closest ~p worlds (however distant!) one doesn't believe that p *and* one doesn't falsely believe that p up to *d*. Crucially, sensitivity will always require the elimination of error worlds up to *d*. Notice that, on this view, a belief that p will be safe but insensitive just in case the closest ~p worlds fall outside the *d* sphere and if in one of those ~p worlds one falsely believes that p. In this respect, sensitivity is the more demanding condition and **D1** is true.

Sensitivity and safety are manifestly *anti-luck conditions*. A widely-held intuition is that knowledge is incompatible with certain kinds of epistemic luck. Consider the familiar stopped-clock example. [134] Imagine that S forms a true belief that *the time is 12.00* (henceforth "*t12*"), but where unbeknownst to S the clock stopped exactly 24 hours beforehand. Intuitively S doesn't know that *t12* because her true belief seems luckily true. Both sensitivity and safety accommodate this verdict. After all, S's belief is *insensitive* because the closest worlds in which *~t12* (for example, the worlds where the clock stops at 11.59 or 12.01) are worlds in which S falsely believes that *t12*, and S's belief is *unsafe* due to the many close worlds in which S falsely believes that *t12*. Let us call the species of knowledge-excluding epistemic luck which arises in such cases, *veritic (epistemic) luck* (following Mylan Engel, 1992). [135]

---

[133] Nozick employs NST for his adherence condition ($p \implies Bp$). For the subjunctive is trivially true for all true beliefs unless one considers whether Bp obtains at *close* worlds in which p obtains. Nozick claims that ($p \implies Bp$) holds true if "not only does he actually truly believe that p, but in the "close" worlds where p is true, he also believes it" (p. 176).

[134] See Russell (1948, p. 170) for original example. For further discussion of the stopped-clock example in the context of safety and sensitivity, see Pritchard (2008). For an explanation of how true beliefs may be "Gettiered", see Zagzebski (1994).

[135] Engel (1992) defines a belief as being "veritically lucky" when it is a matter of luck, given one's evidential situation, that one's belief is true (p. 67). Note that anti-luck conditions needn't require the elimination of epistemic luck *simpliciter*. For example, Engel distinguishes "veritic luck" from "evidential luck", where the latter pertains to the luck that one is in an evidential situation E, which is to be distinguished from the luck which pertains to the truth of the belief given that one is in E (ibid). Engel argues that evidential luck is compatible with knowledge . See Unger (1968) for an analogous distinction.

Modalized sensitivity and safety eliminate a wide class of cases of veritic luck; this includes Gettier-type cases and true beliefs which result from wishful thinking or guesswork. Furthermore, for a wide class of cases, sensitivity and safety are on par in countering veritic luck (Pritchard, 2015).

Importantly, both sensitivity and safety eliminate veritic luck without committing one to a thoroughgoing scepticism about knowledge. For example, take your putative knowledge that H[I have hands]. Notice that, to the extent that you know that H, your belief that H is both *sensitive and safe*. Modally speaking, the closest ~H worlds are ones in which you don't falsely believe that H (for example, the worlds in which you're an amputee), and so your belief is sensitive. Similarly, H is true in most close worlds in which you believe that H, and so your belief is safe. Sensitivity and safety corroborate many cases of knowledge possession because subjunctive conditionals differ from logical entailments. For the sensitivity subjunctive, we need only examine S's doxastic dispositions at the *closest* ~H worlds. And for safety, we need only examine whether H holds at the *close* worlds in which S believes that H. Both proposals are thus compatible with the existence of radical error worlds, such as the world in which you're a disembodied brain-in-a-vat (BIV) and in which you falsely believe that H. However, BIV worlds are not the *closest* ~H worlds (hence sensitivity is satisfied), nor should they be included in the domain of *close* worlds in which you believe that H (hence safety is satisfied). The existence of radical error worlds doesn't render your belief that H insensitive or unsafe.

Both sensitivity and safety exclude (a wide class of) cases of veritic luck from the extension of knowledge, while also retaining much of our putative knowledge. But despite both conditions yielding identical predictions for a wide range of cases, I will argue that safety rather than sensitivity is the correct anti-veritic luck condition.

§ 4.2. Domination and Differential Support

Domination requires the truth of the conjunction [**D1** & **D2**]. Zalabardo denies the truth of the conjunction by rejecting the second conjunct, **D2**.

But the first thing to notice is that ~**[D1&D2]** *alone* cannot surely confer *differential support* for sensitivity. For ~**[D1&D2]** is consistent with counterexamples to sensitivity. Sosa argued for *domination*, but he also propounded counterexamples to sensitivity. Sosa *replaced* sensitivity with safety because safety

does not (ostensibly) succumb to counterexamples. On first blush: S knows that p only if S has a safe belief that p. Formally,

$$Kp \rightarrow (Bp \Rightarrow P)$$

But that schema tells only half the story. Because, so far, we have not eliminated the necessity ("□") of the sensitivity condition. Sosa argues that knowledge doesn't require sensitivity (i.e. [K$p \nrightarrow (\sim p \Rightarrow \sim Bp)$]). Thus, Sosa's safety premise is better represented as:

$$Kp \rightarrow (Bp \Rightarrow P) \& (\sim\Box(\sim p \Rightarrow \sim Bp))$$

The above principle allows insensitive but safe beliefs to constitute knowledge. In sum: a belief having the property of insensitivity doesn't affect its epistemic standing. It is therefore the conjunction of *domination* and Sosa's *counterexamples* which are used to *differentially support safety* as the correct requirement for knowledge. That is, Sosa argues that there are no virtues which are exclusive to sensitivity *AND* there are some virtues which are exclusive to safety. Sosa's best-known counterexample to sensitivity is this:

> CHUTE: "On the way to the elevator I release a trash bag down the chute of my high rise condo. Presumably I know my bag will soon be in the basement. But what if, having been released, it still (incredibly) were not to arrive there? That presumably would be because it had been snagged somehow in the chute on the way down (an incredibly rare occurrence), or some such happenstance. But none such could affect my predictive belief as I release it, so I would still predict that the bag would soon arrive in the basement. My belief seems not to be sensitive, therefore, but constitutes knowledge anyhow, and can correctly be said to do so." (Sosa, 1999, p. 145)

Does safety accommodate CHUTE? Safety theorists will need the following verdict: if the chute is a regular smooth chute, then the bag-snagging world is *not close*. Conversely, if the chute has *s*tructural irregularities which leave it prone to bag-snagging, then the bag-snagging world is *close*. Notice that if the chute is prone to bag-snagging*,* then the example may also fail to elicit the intuition that the subject *knows*. If this is the case, then our safety intuition in CHUTE is covariant with the epistemic intuition that the subject knows the target proposition.[136]

---

[136]My thanks to an anonymous referee for this journal on this point.

Nevertheless, even if the chute is smooth, the closest error worlds in CHUTE are not nearly as distant as the error worlds in which you or I falsely believe that we're not a disembodied brain-in-a-vat, for example. That is, on a similarity ordering of possible worlds, the error worlds in CHUTE are not *radically dissimilar* to the actual world. After all, not *that much* would have to change - modally speaking - for the bag to snag. The error world is not distant enough for it to be uncontroversial that safety is satisfied. The worry, therefore, is that there is no principled way of excluding bag-snagging worlds from the domain of *close* worlds. The case rests on our intuitions about whether the error worlds are "close", and here –I suspect - the case may fail to elicit clear reactions in either direction.

I want to concede that safety as Zalabardo formulates it fails to straightforwardly accommodate CHUTE (This concession is benign since I shall argue that sensitivity can't be differentially supported). If this is so, then there are construals of safety which *do* accommodate CHUTE. Note first that Zalabardo construes safety *strongly*. For example, Zalabardo argues that S's belief that p is safe just in case in *every* world in which S believes that p that is at a distance of *d* or less from actuality, p is true (p. 3). The inclusion of the universal quantifier "all" makes this safety condition *modally strong* because that quantifier prohibits the existence of *any* error world up to distance *d* or less from α. If CHUTE is a counterexample to *sensitivity*, then – so it goes - CHUTE is a counterexample to *strong safety* (for there exists at least *one* close possible bag-snagging world in which one falsely believes that the bag is in the basement). Hence, it has been claimed, CHUTE doesn't motivate strong safety over sensitivity. However, we can see straight away that by adopting the non-standard translation key **(NST)** safety theorists needn't endorse strong safety. Consider, for example, a more liberal version of safety:

> **Weak safety:** S's belief that p is weakly safe just in case in *most* worlds in which S believes that p that are at a distance of *d* or less from α, p is true.

CHUTE is not a counterexample to weak safety, because in *most* close possible worlds, one's belief in the target proposition is true (Pritchard, 2012, 2015). Indeed, in virtually all close possible worlds in which one believes that [the bag is in the basement], one's belief in this proposition is true. This matters. Because when we assume strong safety, it seems that both sensitivity and safety conditions are roughly on par in the number of counterexamples to their necessity. However, the question of how *many worlds* are relevant to safety arises because safety theorists embrace **NST**.

By embracing **NST** for safety, we can see straight away that there arises the question of how *strongly* or *weakly* safety ought to be construed. It is thus open to safety theorists to argue that *sensitivity fails to dominate* weak safety. To argue, in other words, that there are cases of knowledge possession which pass the *weak safety test*, but which fail the *sensitivity test*. In effect, to argue as follows:

(D1*) [Kp & (Bp) weakly safe] → (Bp) sensitive.

(D2*) [~Kp & ~(Bp) weakly safe] → ~(Bp) sensitive.

It cannot be said that sensitivity reaps every advantage reaped by weak safety. If sensitivity entails weak safety, while that guarantees **D1**, entailment doesn't guarantee **D1*.** Entailment doesn't stop sensitivity failing to dominate weak safety. And if sensitivity doesn't dominate weak safety, then we can't *differentially support* sensitivity.

Note, however, that there are cases of knowledge failure which are explained by strong safety but not by weak safety. Consider, for example, the following LOTTERY case. Imagine one buys a lottery ticket where the odds of winning are 60 million to 1. Despite these long odds, one's pessimistic belief that *L*[I am in possession of a losing ticket] intuitively falls short of knowledge before the result is announced. Nevertheless, in *most* close worlds one's belief that *L* is true, and so one's belief that *L* is weakly safe. In contrast, so it goes, there exists at least one close world in which one has a false belief that *L*, hence one's belief that *L* violates strong safety. [137] Strong safety succeeds here where weak safety fails, such that the retreat from strong to weak safety may seem ad-hoc.

Two points here. First, it is unclear that safety theorists are wedded to a construal of safety which can't accommodate both CHUTE and LOTTERY. Consider Duncan Pritchard's defence of safety (2008, 2012, 2015). Pritchard construes safety as an anti-veritic luck condition, one which explains cases of knowledge failure such as LOTTERY and Gettier-type cases. Importantly, Pritchard understands luck as a modal notion which admits of degrees, where the degree of luck pertaining to an event is proportionate to the modal proximity of worlds in which that same event fails to obtain. Pritchard's safety condition is motivated by an extensive modal account of luck (see Pritchard 2015). Let's focus on the account's implications.

---

[137] See Greco (2003). See also Pritchard (2008, 2012, 2015). The lottery case also militates against a probabilistic construal of modal similarity, because the high likelihood of *L* obtaining doesn't detract from the closeness of the ~*L* world.

What is arguably problematic about LOTTERY is the very close modal proximity of the worlds in which one falsely believes that *L*. After all, as Pritchard reminds us, the closeness of lottery error worlds explains the lure of the game. In this case, we need only *one* error world to render the belief veritically lucky in a way which precludes knowledge. In contrast, it is not obvious that CHUTE error worlds are *very similar* to the actual world (Indeed, it's controversial whether error worlds in CHUTE are even *close*) and hence the degree of luck at stake doesn't undermine knowledge. Safety may handle CHUTE by requiring the elimination of *all very close* error worlds, with that tolerance gradually increasing as one moves further away from the actual world (Pritchard 2015, p. 101). By the time we reach the *close* error worlds in CHUTE, safety needn't display complete intolerance of error.[138] Safety, at this point, still requires that one doesn't falsely believe the target proposition at *most close worlds*. It is just that LOTTERY violates the safety condition in virtue of safety displaying complete intolerance towards falsity at worlds which are *very close* to the actual world. Weak safety is now supplemented with a strong safety condition, one which applies only to worlds which are very close to the actual world.[139]

The second point, and notwithstanding Pritchard's proposal, is that we need to judge the plausibility of weak safety overall. While weak safety doesn't explain LOTTERY, weak safety does explain a class of cases of knowledge failure (Gettier cases, lucky guesses etc.) and – being a less demanding condition than strong safety – it trivially fares no worse than strong safety at corroborating intuitive cases of knowledge possession.[140] On this point, we must remember that the issue we are considering here is whether sensitivity can be supported over safety as a *necessary* condition for knowledge. CHUTE may undermine the *necessity* of strong safety, but LOTTERY – as a case of knowledge failure – undermines the *sufficiency* rather than the *necessity* of weak safety.

Weak safety is an improvement on strong safety as a candidate necessary condition for knowledge because weak safety succumbs to fewer counterexamples.[141] In sum: weak safety avoids at least one of the counterexamples which Sosa levied against sensitivity. This is not to vindicate weak safety as a necessary condition for knowledge, for it too might be vulnerable to counterexamples. However, at least one

---

[138] For discussion of weak and strong safety, see also Becker (2007), and Greco (2003), who poses the problem for safety which Pritchard (2015) addresses.
[139] This is closest to Pritchard's (2008) presentation of the issue.
[140] I am not saying that weak safety explains every Gettier-type case. On this point, see Greco, (2003).
[141] The lottery case implies that knowledge is not co-extensive with *weakly safe belief*. But safety theorists needn't *analyse* knowledge as safe belief - neither Sosa (1999), Pritchard (2012, 2015) or Williamson (2000) argue that having a safe belief is necessary and sufficient for knowledge.

prominent counterexample to sensitivity doesn't affect weak safety, such that Zalabardo's case for differentially supporting sensitivity is undermined.

§4.3. Weak Sensitivity and Strong Sensitivity

The distinction between strong and weak sensitivity is not raised in Zalabardo's paper. However, I want to raise it here and exploit its implications for the present dialectic. I will argue that in extensional respects weak sensitivity doesn't outperform weak safety.

We may construe Zalabardo's sensitivity condition as a strong sensitivity condition. Zalabardo's sensitivity condition is *strong* because it demands that in *all* worlds up to a distance of CT(~p) +d or less from α, S doesn't falsely believe that p. (The inclusion of the universal quantifier makes Zalabardo's condition *modally strong*). However, as with safety, sensitivity theorists who use the non-standard translation key could in principle adopt a weaker rendition, one which – like weak safety - drops the universal quantifier[142]:

> **WS**: S's belief that p is *weakly sensitive* in α just in case: at *most worlds* in which p is false at a distance of CT (~p) + *d* or less from α, S doesn't falsely believe that p.

Rather than Zalabardo's strong sensitivity condition:

> **SS:** S's belief that p is *strongly sensitive* in α just in case: at *every world* in which p is false up to a distance of CT (~p) + *d* or less from α, S doesn't falsely believe that p. (Zalabardo, p. 5)

I want to make two points in what follows. First, **WS** doesn't outperform weak safety in extensional respects. That is, both conditions include in the extension of knowledge cases which should be included (CHUTE), and both fail to exclude from the extension of knowledge cases which should be excluded (ROGER). Second, **WS** provides us with no reason for abandoning the standard counterfactual analysis for sensitivity: that **WS** yields the same predictions as weak safety is explained by the claim that **WS** just *is* a weak safety requirement for knowledge.

[142] See Greo (2012) who discusses strong restricted sensitivity and weak restricted sensitivity; both conditions materially differ to Zalabardo's condition (SS) and its weaker variant (WS), insofar as restricted sensitivity makes distant ~ p worlds irrelevant to the determination of sensitivity.

Let's see how **WS** performs in extensional respects by first examining whether **WS** can accommodate CHUTE. Let "S" denote the subject who dispatches the bag down the chute and let "p" denote the proposition that *the bag is in the basement*. In the actual world, S dispatches the bag and forms a true belief that p. Consider now a range of ~p worlds up to a distance of CT(~p)+d and consider whether or not S believes that p in most of these ~p worlds. Note that for **WS** we needn't examine only the "closest" ~p worlds (i.e. the bag-snagging worlds) – for we are now also including "close" ~p worlds in the domain of relevant ~p worlds. The **WS** condition will be satisfied if S doesn't falsely believe that p in *most* ~p worlds up to a distance of CT(~p)+d, where d denotes the range of close worlds. I think we will see that CHUTE satisfies **WS**.

The closest ~p worlds are bag-snagging worlds. But importantly, there is still a wide class of relevant ~p worlds (~p worlds in which S doesn't believe that p) which lie further away from actuality than bag-snagging worlds, but which are paradigmatically *close*. This is due to the wide class of close worlds in which ~p holds because S has not dispatched the bag down the chute, and in these close ~p worlds S doesn't believe that p. For example, consider the many close worlds in which rather than dispatching the bag down the chute, S attends to other mundane tasks; or consider, for instance, the relatively close worlds in which the chute is closed for maintenance. Modally speaking, these ~p worlds are not radically dissimilar to the actual world; indeed many of them are very similar to the actual world. To the extent that S knows that p in the actual world, S doesn't believe that p in a wide class of close ~p worlds. The close ~p worlds described are not the *closest* ~p worlds, but they are *close* nevertheless. Furthermore, intuitively there are more *close* ~p worlds in which S doesn't believe that p than there are the narrow band of *closest* bag-snagging worlds. Therefore, CHUTE isn't a counterexample to **WS**.[143]

We should take **WS** seriously as an anti-veritic luck condition because it's not implausible that **WS** yields the same prediction as weak safety for CHUTE. It may, therefore, be supposed that what sensitivity theorists need is **WS** rather than **SS**. However, I want to argue that replacing **SS** with **WS** comes at an exorbitant price for sensitivity. To see why, let us consider the following case:

ROGER: "Consider Roger, who believes ~BIV6(where "BIV6" is the proposition that *MI6 secretly keeps a collection of envatted brains, artificially*

---

[143] At a minimum, there doesn't seem to be any obvious sense in which CHUTE *violates* **WS**. However, the claim that CHUTE isn't a counterexample to **WS** is benign in the present dialectic, since I will argue that we can support weak safety over **WS** even if **WS** accommodates CHUTE.

stimulated to produce the experiences of normal embodied beings) but for slightly unorthodox reasons. Roger doesn't believe that brain envatment is technically impossible. In fact he believes it's a common occurrence, since he heard about Putnam's thought experiment and got the wrong end of the stick. However, he is convinced that MI6 doesn't engage in these activities, the reason is that he has a friend who tells him that he works for MI6 and is always prepared to answer his questions about this service. As it happens, Roger's friend is just a cleaner in the MI6 headquarters, with no access to any classified information, and gives random but coherent answers to Roger's questions, just to humour him. When Roger asked him if MI6 keeps any envatted brains, he assures him that they didn't. It is on these grounds that Roger believes ~BIV6." (Zalabardo, p. 8, my parentheses)

Zalabardo argues that ROGER is a case of knowledge failure explained as such by **SS** but not by safety. ROGER violates **SS** because we can expect that at *some* of the closest BIV6 worlds, Roger's friend tells him that MI6 don't keep envatted brains, and at these worlds, Roger falsely believes ~BIV6. Hence, Zalabardo claims that safety fails to *dominate* **SS**.

Once we substitute **SS** with **WS**, however, ROGER no longer elicits the intuition that **WS** is violated. After all, ROGER's belief is (strongly) epistemically safe because the BIV6 error worlds are no less distant than the distant worlds in which *you* or *I* are a BIV; or so that is Zalabardo's claim, and that is the claim he needs for ROGER to clearly fail the safety test.[144] Once ROGER is understood in this way, as strongly safe, and *a fortiori* as weakly safe, it is no longer uncontentious that **WS** does the explanatory work which weak safety fails to do. The problem at stake here is that ROGER doesn't yield the reaction that **WS** is violated. After all, ROGER is *safe* we are told, so it is unclear why ROGER is not also weakly sensitive.

For **WS** to do explanatory work here, there would need to exist *more* closest BIV6 worlds in which Roger falsely believes that ~BIV6 than there are *close* worlds (and worlds up to the closest BIV6 worlds) in which Roger doesn't falsely believe that ~BIV6; this suggestion is tenuous at best. Indeed, the consideration of doxastic

---

[144] Zalabardo observes "the sheer distance from actuality of the nearest BIV worlds – beyond any plausible value we might set for *d* – means that my belief in ~BIV – and indeed any belief in this proposition by an embodied subject in the kind of world we think we inhabit – will be safe" (p. 7). Zalabardo then argues that the closest BIV6 worlds are no closer to actuality than the closest worlds in which *I* or *you* are BIV's (p. 8). However, Zalabardo's claims on p. 8 suggest a slightly different view. I will assume however that the *closest* worlds in which *I* or *you* are a BIV are *radically distant* – this is the claim that Zalabardo needs, for otherwise it won't be clear that the belief ~BIV6 is safe and, a fortiori, it won't be clear that Roger's belief is *safe*.

dispositions at the closest BIV6 worlds (beyond the close worlds in which Roger doesn't falsely believe that ~BIV6) appears to make no difference to our modal intuitions about what Roger believes at *most* of the relevant possible worlds. Therefore, **WS** fares no better than weak safety at *explaining* ROGER, which is the very case intended to show the superiority of sensitivity over safety.[145] **WS** doesn't do the explanatory work which weak safety ostensibly fails to do for ROGER.[146]

Is there a way of adjudicating between **WS** and weak safety, given that both yield the same predictions for these cases? Do we have a safety/sensitivity impasse? I do not think we have an impasse. The impasse presupposes that we are dealing with two different conditions. However, treating **WS** as denoting the sensitivity counterfactual [$\sim p \Rightarrow \sim Bp$] violates the standard analysis of counterfactuals. On the standard analysis, for cases where the antecedent p of a subjunctive ($p \Rightarrow q$) is presumed to be *false* in α, the counterfactual ($p \Rightarrow q$) is true just in case the consequent q obtains at the closest worlds to α where the antecedent p obtains. On the standard analysis, the sensitivity counterfactual ($\sim p \Rightarrow \sim Bp$) is false if Bp obtains at one or more of the closest ~p worlds. **WS** deviates from the standard counterfactual analysis because **WS** permits ($\sim p \Rightarrow \sim Bp$) to be true in cases where Bp obtains at one or more of the closest ~p worlds, providing that one doesn't falsely believe that p at *most* worlds up to CT(~p) + d.[147] **WS** violates the standard counterfactual analysis, while at the same time supplying no reason to abandon the standard analysis.

The way to preserve the standard analysis of counterfactuals is to treat **WS** as expressing the safety subjunctive (Bp $\Rightarrow$ p) rather than the sensitivity subjunctive

---

[145] The important point here, with respect to our modal intuitions, is that ROGER doesn't elicit the modal intuition that **WS** is violated, and that **WS** is therefore an impoverished explanation of ROGER. **WS** fails to do the explanatory work required of it.

[146] Weakening the non-standard sensitivity condition also renders it unable to block the neo-Moorean argument that you can and do know that you're not a brain-a-vat. After all, your belief that ~BIV satisfies **WS** for much the same reason that it satisfies weak safety, namely that in *all* close worlds and all worlds up to the distant BIV worlds, one's belief in this proposition turns out to be true. **WS** is no less amenable to neo-Mooreanism than weak safety.

[147] On this point, I am indebted to a referee for this journal. **SS** also clashes with the standard counterfactual analysis, because [~Bp obtains at the closest ~p worlds] is insufficient for **SS** obtaining. For example, **SS** will won't obtain if at the closest ~p worlds one doesn't believes that p but if there is an error world in the d sphere. However, we may at least reconcile **SS** with the standard counterfactual analysis by taking **SS** to denote the conjunction [$\sim p \Rightarrow \sim Bp$ & Bp $\Rightarrow p$]. On this view, the sensitivity conjunct will be satisfied in those cases where [~Bp obtains at the closest ~p worlds], but the conjunction will be false if there is an error world within the d sphere. Zalabardo's view that sensitivity entails safety is consistent with this. As is his conclusion that "safety cannot do by itself all the work that sensitivity can do in excluding from the extension of knowledge beliefs that shouldn't be there"(p. 9). If this is so, then the claim of this paper is that the conjunction of sensitivity and safety can't – from an extensional point of view - be differentially supported over safety by itself. Note that a conjunctive view doesn't allow us to reconcile **WS** with the standard counterfactual analysis, since **WS** violates a necessary condition for the truth of the sensitivity conjunctive – namely the condition that in the closest ~p worlds, one doesn't believe that p.

(~p ⇒ ~Bp). Moreover, since **WS** and weak safety yield the same predictions for the cases under consideration, there is no reason to think that what we are dealing with here is a *sensitivity condition* rather than a *weak safety condition*.

But does weak safety also violate the standard counterfactual analysis? One final worry is that (Bp ⇒ p) may be true according to weak safety even if (~p) obtains at one or more of the "close" (Bp) worlds, which at first glance seems inconsistent with the standard counterfactual analysis. To allay this worry, we need to keep in mind that weak safety does not express a *counterfactual conditional*, but rather expresses a *subjunctive conditional with a true antecedent*. Note that the hallmark of a counterfactual is that its antecedent is presumed to be *false* in the actual world "α". For the sensitivity condition (~p ⇒ ~Bp), the antecedent (~p) is presumed to be *false* in α because we are concerned with a class of *true beliefs* which fail to constitute knowledge, cases such as ROGER. The root idea here is that a true belief fails to constitute knowledge if it doesn't have the property of being sensitive, such that sensitivity emerges as a non-trivial condition for knowledge, over and above the conditions that the agent must *truly believe* that p to know that p. By contrast, for weak safety (Bp ⇒ p), the presumption is that the antecedent (Bp) is *true* in α. For example, when weak safety excludes veritically lucky *true beliefs* from the extension of knowledge, the antecedent (Bp) is presumed to be true in α. Likewise, when weak safety backs-up cases of knowledge such as CHUTE, the presumption is that the antecedent (Bp) is *true* in α: because the agent has a *true belief* which constitutes knowledge. The point is that when weak safety eliminates veritic luck, and corroborates knowledge, the presumption is that the antecedent (Bp) is true in α.

We should therefore construe sensitivity as a *counterfactual subjunctive conditional* and weak safety as a *non-counterfactual subjunctive conditional*.[148] As we saw in Section 1, weak safety requires the non-standard translation key to avoid triviality, however adopting that key for weak safety does not relinquish the standard counterfactual analysis *for counterfactuals*. Providing that we keep in mind that the non-standard key is used by weak safety to unearth the truth conditions of a non-

---

[148] Subjunctive conditionals with true antecedents are sometimes called "counterfactual conditionals", and safety is sometimes described as a counterfactual condition in the literature. However, by calling safety "a non-counterfactual subjunctive" I am merely saying that its antecedent, for the cases which concern us, is presumed to be true in α. Lewis (1973a, p. 3) calls such counterfactuals with true antecedents "counterfactuals that are not counterfactual". DeRose (2004) aptly describes the safety subjunctive as a "true/true subjunctive"; this is perhaps the clearest statement of the safety condition, since insofar as truth is a condition for knowledge, the presumption is that the consequent (p) as well as the antecedent (Bp) is true in α.

counterfactual subjunctive, one can endorse weak safety alongside the standard analysis of counterfactuals.

In summary, the real contest is not between *weak sensitivity* and *weak safety* but is instead between the full-blooded strong sensitivity condition (**SS**) that succumbs to CHUTE, and a weak safety condition which overcomes that counterexample. In this contest at least, weak safety can be differentially supported as the correct necessary condition for knowledge.

Chapter 6

Internal Justification and Animal Knowledge

In summary, so far, the thesis argued in Chapter 2 that awareness internalism can be motivated by anti-luck considerations and responded to the Reflective Luck Argument. Chapter 3 further tested the hypothesis that the awareness requirement for justified belief is satisfied only when one's belief is formed using a reliable method. Chapter 4 continued by examining the hypothesis that justification requires reliability, considering the Demon Problem as a challenge to the reliability condition. Chapter 5 provided a more detailed account of the methods of belief formation relevant to the awareness condition, and further examined the nature of the anti-luck condition, supporting the safety condition over the sensitivity condition.

This chapter returns to the connection between justification and knowledge, to consider one final objection to the proposal that knowledge, at least knowledge of the best kind, requires believers to satisfy an awareness condition. This chapter outlines an objection to awareness internalism. The objection considered is that awareness internalism, when combined with a justificatory condition for knowledge, implies that a wide class of subjects lack knowledge. The claim that animals have knowledge is contentious, but the chapter explores the implications for awareness internalism if the claim is correct. The chapter responds by how knowledge attributions for reflectively unsophisticated believers need not be interpreted as counterexamples to awareness internalism.

This chapter is structured as follows:

§1. Outlines the objection that awareness internalism cannot explain how unsophisticated believers may correctly be said to possess knowledge.

§2. Examines the view that reflectively unsophisticated subjects can correctly be said to possess knowledge, despite lacking second-order awareness. I consider motivations for the objection. I respond to the objection by contending that unsophisticated believers may know despite lacking justified beliefs. I thereby introduce a notion of knowledge without justification.

§3. I explain the view that not all kinds of knowledge depend on justified belief, situating the view in the context of the conclusions of previous chapters.

Finally, in a separate section to this chapter, I explain how the thesis has addressed the research question of Chapter 1, outlining the conclusions that may reasonably be drawn.

§ 1. Unreflective Believers: A Problem for Awareness Internalism

The version of awareness internalism supported in previous chapters encounters a difficulty. Recall the condition: one has a justified belief only if one's belief is formed via a reliable method and one is aware of it being formed via a reliable method.

Internal (justification) condition: S has a justified belief B *only if* S has used a reliable method of belief formation in arriving at B and S is aware she has used a reliable method of belief formation in arriving at B.[149]

The above condition makes reliability a necessary rather than sufficient condition for justified belief; according to the condition, holding a justified belief B requires the believer to have second-order awareness (even if not occurrent) that B reliably formed.[150] This present chapter does not concern the necessity of reliability as a condition for justified belief, which was examined in Chapter 4. Instead, this chapter concerns the necessity of second-order awareness.

A wide class of subjects, for example animals, small children, and infants, may fail to satisfy the second-order awareness condition for justified belief. If they have beliefs that constitute knowledge, and if knowledge requires justified belief, then awareness internalism predicts that these subjects fail to know because they fail to know in virtue of failing to satisfy the awareness condition for justified belief.

The intuition that animals possess knowledge may be illustrated by an empirical example. A study in the scientific literature by John Pilley (2013) describes a female border collie dog named "Chaser". Pilley's research found that Chaser met the criteria for knowing 1022 proper nouns. Chaser could associate nouns with objects

---

[149] See: Chapter 1, § 3
[150] I will assume that the awareness will need to be doxastic in what follows to meet the Subject Perspective Objection of Chapter 3.

such as toys, and respond to many commands; for example, by correctly identifying the hidden toy and retrieving it. Chaser could retain this knowledge in memory. In one section of that paper, Pilley remarks that Chaser "accumulated and maintained this *knowledge* of nouns over a lengthy 32-month period." (p. 230 - my italics).

When considering the awareness condition for justified belief, it is not obvious that a wide class of subjects would satisfy it. This is because, for a belief to be justified, awareness internalism maintains the believer must, at least, have the capacity to believe that her belief results from a reliable method of belief formation. The subjects who plausibly fail to satisfy the awareness condition include animals, infants, and small children: many of these subjects are not, presumably, reflectively sophisticated enough to form meta-beliefs about the reliability of their methods of belief formation.[151]

Laurence BonJour (2009) discusses the objection to awareness internalism I am describing. He describes an intelligent dog, Emma, who has a good working memory and an ability to respond to her environment in novel ways (p. 173). BonJour finds it intuitive that Emma knows, but observes that she does not satisfy an awareness condition for justified belief. In discussing this objection, BonJour indicates that it may be comparatively more intuitive to say that an unsophisticated believer has knowledge rather than justification. This observation is important, because it raises the possibility of a type of knowledge persisting in the absence of second-order awareness. I shall turn to this point later, but for now, I will focus only on intuitions about unsophisticated believers possessing knowledge, since these intuitions are – I suspect – easier to elicit than intuitions about unsophisticated believers holding justified beliefs.

The objection to awareness internalism needs to be considered in more detail. Specifically, if animals know, and if knowledge requires justification, then the awareness condition has a problematic implication; it implies that we are incorrectly attributing knowledge to reflectively unsophisticated subjects. Given how broad the range of cases is, the condition would imply widespread scepticism.[152]

An argument against awareness internalism now arises. The argument contraposes on the implication that reflectively unsophisticated believers lack second-order awareness. The argument may be summarised as follows: since unreflectively sophisticated believers can correctly be credited with knowledge, and since knowledge

---

[151] In considering the above point, we need to keep in mind that the awareness requirement supported is doxastic (it will not be enough, therefore, that the animals merely have a concept or experience of their belief being true).

[152] See Chapter 2, § 2.3 for the stipulation of attributor scepticism.

requires justification, knowledge does not require second-order awareness. More schematically:

1. Knowledge attributions are correct for reflectively unsophisticated subjects who fail to satisfy a second-order awareness condition for justified belief.

2. A belief constitutes knowledge only if it is justified.

    Therefore,

    Second-order awareness is not a requirement for justified belief and knowledge.

There are at least three responses to this dilemma:[153]

A. Accept Premises 1 & 2 (abandon awareness internalism).

B. Accept Premise 2 but reject Premise 1 (unreflective subjects fail to know).

C. Accept Premise 1 but reject Premise 2 (abandon the condition that propositional knowledge depends on justification in all cases).

§1.1. A Strategy

I shall argue that, if reflectively unsophisticated believers can correctly be said to know, strategy (C) is defensible. However, the variant of strategy (**C**) abandons the condition that knowing a proposition always requires having a justified belief in a

---

[153] Notice that the argument assumes the same particularist approach to understanding the conditions for justified belief discussed in Chapter 2 (§ 2.3) and Chapter 3 (§ 2). The particularist approach holds that the correct criteria for justified belief ought to resonate with our intuitions about particular cases of knowledge. Since the awareness condition implies scepticism for a broad class of unsophisticated subject, sceptical implications count against the awareness condition. Conversely, a methodist approach to the conditions for justified belief would maintain that the correct criteria for justified belief should be used to establish whether or not reflectively unsophisticated believers have justified beliefs and knowledge. A methodist approach would maintain that the sceptical implications of awareness internalism are an insufficient basis to reject awareness conditions. I will assume - for argument's sake - that the particularist approach is correct. Although the sceptical implications of awareness internalism are not global here they are still wide-reaching and may still seem counterintuitive enough to count against a particularist approach.

proposition. The strategy will not maintain that no case of knowledge depends on justified belief. Instead, it will hold that not all cases in which a believer has knowledge are cases in which the belief is justified. Instead, it will hold that not all cases in which it is salient to attribute knowledge are cases in which the belief is justified. I will focus primarily on animals, because it is easier to elicit the intuition that they violate the awareness condition, since it is intuitive that many animals have only a rudimentary awareness of their beliefs, if any awareness at all. However, the results may be equally applicable to unsophisticated believers such as infants and small children.

Before doing so, it is important to distinguish a view of animals possessing knowledge, examined in this chapter, from Ernest Sosa's (2007, 2009) distinction between two kinds of knowledge, which he terms "animal knowledge" and "reflective knowledge. Sosa's distinction arises in the context of his virtue-theoretic framework and, using this, Sosa advances a necessary condition for knowledge of any kind: one knows that p only if p is true and one's belief that p results from truth-conducive dispositions. Sosa argues that one cannot know that p unless one's knowledge derives from the manifestation of a cognitive virtue, where a cognitive virtue is a truth-conducive disposition (2007, p. 135). Sosa distinguishes between a belief's truth, its manifestation of some virtue or competence (adroitness) and whether one's belief is true because of one's competence (aptness). A cognitive disposition does not count as a virtue on this account unless it is a truth-conducive disposition.[154] Sosa is using the notion of "animal knowledge" as applying also to human subjects, and it is human knowledge that he is seeking to understand (See 2007, p.98-98). Animal knowledge, on Sosa's account, is of a lesser quality than reflective knowledge. By contrast, the reasons for attributing knowledge to animals, considered in §2, will not appeal to a distinction between two different kinds of knowledge. Instead, we will consider explanatory grounds for attributing knowledge to reflectively unsophisticated believers.

The problem considered in this chapter is how to reconcile awareness internalism with the view that reflectively unsophisticated subjects know propositions. It is important to emphasise, however, that whether animals possess knowledge is contentious. However, what I hope to do is to identify at least *some* reasons for finding it attractive that animals can correctly be said to know. If unsophisticated believers cannot correctly be said to know, then the problem I have

---

[154] Sosa regards animal knowledge as belief that is merely true and *apt*. By contrast, reflective knowledge requires perspectival endorsement of the reliability of one's sources" (2009 p. 136). Sosa characterizes reflective knowledge as governed by the principle that "knowledge is enhanced through justified trust in the reliability of its sources" (ibid., p. 139).

described for awareness internalism in § 1 would not arise. The remainder of this chapter considers whether awareness internalism could in principle accommodate the view that unsophisticated believers can correctly be said to know.

§1.2. **Reflective Sophistication**

There will be varying degrees of reflective sophistication among a broad class of subjects for whom we commonly attribute knowledge. However, I use the expression "reflectively sophisticated" to include subjects only who plausibly have second-order beliefs about the reliability of their first-order beliefs. On this stipulative definition of reflective sophistication, there may be *cognitively sophisticated* subjects who are nevertheless reflectively unsophisticated. An animal may be cognitively sophisticated because it is attuned to its environment and can respond to its environment in complex and novel ways, while being reflectively unsophisticated because it lacks second-order awareness.

A question arises: how do we know whether a believer, human or animal, satisfies the second-order awareness requirement? One attractive proposal is that evidence of second-order awareness is manifest in language. For example, in response to the question, "how do you know?", the answer "I saw it" embodies a second-order stance, precisely because it indicates a belief (even if not occurrent) that our familiar methods of belief formation are generally reliable. Though the notion is still vague, it is sufficiently clear to elicit the intuition that many subjects lack second-order awareness, because they do not belong to a linguistic community in which such questions are routinely asked. Such subjects will not be reflectively sophisticated (in the stipulative sense) even if they are cognitively sophisticated.

A concern about the above approach is that animals *may* have second-order beliefs about their belief-forming procedures, but that we cannot be expected to know they have second-order beliefs because we are – after all - members of a different linguistic community. This concern is legitimate, but it would leave us with no straightforward response, either way, to the question of whether or not animals have second-order awareness. If animals do have second-order awareness, however, then there would be no tension between awareness internalism and the claim knowledge attributions are correct for a broad class of cases. Proceeding on the assumption that animals lack the requisite awareness, we are confronted with the objection that awareness internalism cannot account for animals possessing knowledge.

The next section considers reasons for maintaining that animals do know propositions, and subsequently considers whether awareness internalism can accommodate this.

§ 2. Unsophisticated Knowledge

If animals have knowledge and if knowledge requires belief, it would need to be plausible that the representational states animals have constitute beliefs. However, the notion that an animal can believe that a given state of affairs obtains seems comparatively more controversial than the claim that animals may represent their environment is numerous ways and behave consistently with those representations.

The view that animals have beliefs may be supported by considering the role that belief attribution serves in explaining behaviour. Imagine a Border Collie barks loudly to deter what it represents mentally as an intruder. Not implausibly, the Collie *believes* that there is an intruder in the vicinity. Its belief, coupled with a *desire* to protect its keeper, is one explanation of why it barks. This explanation is not the only one, however. There might in principle be a reductionist explanation of the animal's behaviour; for instance, an explanation of the various neuro-physical states and neural pathways of the dog, which explain why it barked to deter an intruder.[155]

However, as Hilary Kornblith argues in *Knowledge and Its Place in Nature* (2002), it is doubtful that we can characterise an instance of behaviour in purely physical terms.[156] Kornblith observes that it is not intuitive that there is some identical physical state for any two animals which engage in identical behaviour. Kornblith further observes how any explanation is bound to be more complicated and less parsimonious than a belief-desire explanation of animal behaviour.[157] Furthermore, Kornblith argues that given the biological needs of an animal and its ability to have representational states, we have the beginning of a belief-desire

---

[155] See Kornblith on a similar point.

[156] Kornblith's discussion of animal knowledge is placed within his argument for a naturalistic epistemology, which identifies knowledge with reliably true belief, and argues that knowledge is a natural kind which can be investigated by the sciences. In what follows, it is essential to note that Kornblith's view about what knowledge is, and how it derives its normativity, differs importantly to awareness internalism, which he does not endorse. Instead, I want to only appeal to Kornblith's view that animals have knowledge, and I want to appeal also to his argument that there are norms which animal believers satisfy which are a variety of hypothetical imperative. Section 3 of this chapter will further outline the view which emerges from taking these two positions and considering them alongside awareness internalism.

[157] See Kornblith's statements on pp. 37-42.

psychology that explains behaviour, one which parallels the belief-desire psychology we invoke when *explaining* human action.[158]

   If animals (and the non-sophisticated more generally) do indeed have beliefs, then there arises a question of whether, when those beliefs correctly represent states of affairs in the world, the beliefs constitute knowledge. Kornblith supports the claim that animals do possess knowledge, providing their beliefs result from a stable capacity for true belief, a capacity which an animal has evolved to have, and which reliably produces true beliefs. Kornblith writes:

> "Notice that these explanations require more than just the category of true belief. If we are to explain why it is that plovers are able to protect their nests, we must appeal to a capacity to recognise features of the environment, and thus the true beliefs that particular plovers acquire will be the product of a stable capacity for true beliefs. The resulting true beliefs are not accidentally true; they are produced by a cognitive capacity that is attuned to its environment. In a word, the beliefs are reliably produced" (2002, p. 58).

At the level of explaining a specific instance animal behaviour, Kornblith observes we can frequently substitute a knowledge attribution with *a* belief attribution, offering an informative explanation of why any particular animal behaves in a particular manner by appealing to a belief coupled with the relevant desire, rather than knowledge (p. 57).[159] For Kornblith, however, there is a reason for appealing to knowledge rather than belief when explaining animal behaviour. He observes that when we turn to our explanation of how a given *species* behaves in specific ways, then we cannot avoid invoking knowledge, because it is not an accident on any occasion that a species, endowed with a particular cognitive capacity, forms true beliefs about its environment. An animal which is successful in its environment has evolved in this way to be able to successfully navigate its environment (p. 63)[160] The reliable belief-

---

[158] Kornblith's example is "the ravens distract the hawk because they are hungry; they want to steal the hawk's egg; they *believe* that by attempting to take the squirrel away from the hawk, they will thereby be able to take the egg" (2002. p. 38, my emphasis).
[159] To use Kornblith's example, if we want to explain why a particular plover leaves its nest, we need to only appeal to the plover's belief that a predator was approaching" (p57).
[160] Kornblith summarises this as follows: "Notice that these explanations require more than just the category of true belief. If we are to explain why it is that plovers are able to protect their nests, we must appeal to a capacity to recognise features of the environment, and thus the true beliefs that particular plovers acquire will be the product of a stable capacity for true beliefs. The resulting true beliefs are not accidentally true; they are produced by a cognitive capacity that is attuned to its environment. In a word, the beliefs are reliably produced" (2002, p. 58).

forming procedures or capacities are those which an animal has due to natural selection and which allow an animal to navigate its environment and survive successfully.161

To understand the way in which knowledge attributions play an explanatory role, consider the following: when explaining why a given member of a species behaves in a particular way that is atypical of that species, and in a way detrimental to its survival, then we may describe that behaviour by using a knowledge attribution. For example, consider the Hungarian Partridge, a non-migratory bird-species. If a strange circumstance arose where a member of that species attempts to migrate even, though doing so is counter to its survival, then that would signify biological dysfunction. We would not characterise the animal as being attuned to its environment and or knowing that it is heading in a particular direction. Furthermore, - I suspect - we would not characterise the behaviour as "migration". For we would now be wanting to explain why the animal behaved atypically and in a way that is counter to how it has evolved to behave.

This section has so far considered grounds for attributing beliefs animals, and explanatory grounds for attributing knowledge. The proposal now I will examine is that if an animal's cognitive capacities are functioning normally, directed towards a desire to survive, then the beliefs that they form using those capacities, when true, may correctly be said to constitute knowledge.

In the cases discussed so far, it seems that the normal functioning of a cognitive process will be one directed at survival. More generally, the notion that whether one knows requires cognitive capacities that are functioning normally is discussed by Alvin Plantinga (1993b). Plantinga argues that propositional knowledge requires a proper-functioning cognitive system (p. 13). The notion of a properly functioning cognitive system is essential, insofar as true beliefs which result from that function may be beliefs a believer ought to have. However, for the claim that animals possess knowledge to be plausible, we would need to explain the appropriate way in which animals *ought* to have certain beliefs. In addressing this, we may distinguish between two kinds of normativity: hypothetical normativity and categorical normativity. Approximately, a norm N for S's belief B will be a hypothetical norm if whether N obtains for B is contingent on S's desire(s). If epistemic norms are hypothetical norms, then S *ought* to believe that p just in case believing p allows S to satisfy some desire. I explained how in cases of prudential justification, a hypothetical norm applies. The norm, in this case, is characterisable as contingent on a desire for

---

161 This account, Kormblith reminds us, is both reliabilist but naturalistic.

it depends on having the relevant desire and would not hold if no desire were present.162

Analogously, the notion of "ought" for reflectively unsophisticated subjects may be explained in terms of subjects having cognitive capacities that are functioning in the way that they have evolved to. On this view, a proper-functioning cognitive system, working in the way it has evolved to work, may be a cognitive system that is working in the way that it *ought to*.163 The notion of proper-function also applies to biological systems that lack representational mental states, these will be systems that lack experiences or any mental states, but which nevertheless have a proper function. For example, if a heart is functioning normally, then it pumps blood from the right ventricle to the lungs. A failure of the heart muscle to pump blood from the right ventricle to the lungs signifies dysfunction and that the heart is not doing what it *ought* to. 164

We have seen that a hypothetical form of normativity may be applied to biological systems that lack a belief-desire-psychology. For example, the proper-functioning organs of a species are functioning as they *ought* to be functioning. The difficulty with extending this to animals, however, is that when one uses the notion of *ought* in a hypothetical way, this does not always entail that there is some item of propositional knowledge the system possesses. A healthy heart may *know how* to pump blood, but there is not some proposition in the vicinity that it can be credited with knowing, for there is not a belief-desire psychology to invoke. Consequently, to support the view that biological systems with a belief-desire psychology satisfy hypothetical belief norms, one would need to focus on *propositions* an animal believes and one would need to show that such beliefs constitute knowledge.

Let us consider the possibility that there are propositions an animal ought to believe. In considering this, I want to assume – for argument's sake – that animals have a belief-desire psychology (this concession is benign, since the notion of animal knowledge arises as an objection to awareness internalism). Consider, therefore, the following: if a creature has evolved to be able to identify that another species before it

162 See Kornblith (p. 160) for similar statements.
163 See Kornnblith (Chapter 2, esp. Sections 2.6-2.9), who maintains an animal's cognitive capacities are directed at its desire to survive as a member of a species and when those capacities are functioning as they ought, they are directed at that biologically-determined desire.
164 Plantinga draws our attention to how we use "ought" in this way, stating that of a diseased pancreas that it "no longer functions as it ought to" (Plantinga, 1993b, p. 45) The notion of proper functioning is critical to Plantinga's account of warrant, where "warrant" is on Plantinga's view the property, in addition to truth and belief, required for knowledge. Similarly, I suspect the notion of proper-function is also essential to an present explanation of how such animals may be said to know

is a predator, then there is at least some proposition in the vicinity it believes and which it ought to believe, given that it has evolved in such a way that it should be able to identify predators. 165 Furthermore, since an animal's belief that a predator is nearby is one that allows it to satisfy its desire to survive; there is a hypothetical notion of normativity at stake here; it is a belief which the animal ought to hold to the extent that having the belief is instrumental in satisfying the desire. Providing the belief is true and results from cognitive mechanisms it has evolved to have, and which allow it to accurately represent its environment and survive, then it has a true belief that it ought to have.

3. Responding to The Objection

Knowledge attributions for animals occur at the level of explanation. They allow us to explain instances of animal behaviour, in particular how an animal interacts with its environment in a way consistent with its desires. Explaining animal behaviour using knowledge attributions would seem more parsimonious than describing the neuro-physical states or pathways of an animal's brain. As §2 observed, when behaviour results from a normally functioning cognitive procedure, one conferred by the evolutionary process, then when an instance of behaviour results from that procedure and is directed at a desire, a knowledge attribution may be used to explain that instance of behaviour.

The concern, however, is that animal subjects for whom knowledge is commonly attributed, both in everyday contexts and in the scientific literature (e.g. Pilley, 2013), would fail to satisfy an awareness condition for justified belief. In response, I will explain how the conception of knowledge appealed to in explaining animal behaviour is not the same normative conception developed in this thesis. I will describe how animal subjects are not, therefore, counterexamples to the proposals advanced in earlier chapters because the reasons for attributing or denying knowledge to animals are not identical to the reasons we have for attributing or denying knowledge to humans, when humans satisfy the awareness condition for justified belief.

An essential difference between the conception of knowledge that requires awareness and the concept of animals possessing knowledge relates

---

165 In the case of creatures with belief-desire psychology, however, Kornblith thinks that the hypothetical ought holds due to the evolutionary process, and that "Knowledge is of extraordinary instrumental value, for it allows us to achieve our biologically given goals, as well as our more idiosyncratic goals, whatever those goals may be". (2002, p. 160)

to the role of justification when attributing knowledge to animals. To elaborate on this, consider the grounds for attributing knowledge to animals. Regardless of whether it is correct to attribute knowledge to them, we have identified that knowledge attributions are commonplace and provide an explanatory framework. Furthermore, for any particular instance of identical behaviour by two different animals of the same species, where the behaviour is directed towards a desire to survive, a knowledge attribution can explain the behaviour parsimoniously (Kornblith, 2002). Crucially, however, the explanation of animal behaviour does not refer to the concept of epistemic justification but, instead, appeals directly to the contention that they possess knowledge. Consequently, these explanations do not imply that awareness internalism offers the incorrect account of justification, because the *explanations themselves* do not support a competing conception of justification.

It is possible, however, that a reflectively unsophisticated subject may satisfy an externalist conditions for justified belief. For instance, if it is appropriate to credit animals with a belief-desire psychology, then animal subjects may be understood as *believers* who satisfy an externalist standard for justification by having reliably true beliefs about their environment. In which case, animal subjects might be interpreted as a challenge to awareness internalism. However, crucially, the use of knowledge attributions to explain animal behaviour does not indicate that the standard for epistemic justification is an externalist standard. This is because the explanation of animal behaviour makes no essential reference to the concept of epistemic justification. The role that knowledge attributions serve in explaining animal behaviour does not, therefore, indicate that awareness internalism should require us to revise our ordinary standards for justified belief.

To further understand the argument above, we may consider "justification" and in particular its connection to "reasonable" belief and other normative notions, such as responsible belief, before asking whether animal subjects could satisfy a requirement for reasonable or justified belief. If a belief is justified in the normative sense, then it is a reasonable belief to hold. As indicated in Chapter 1 §3, an analysis of justification in terms of reasonable belief would be vulnerable to circularity because "reasonable" belief expresses a normative conception of justification. The necessary conditions for justified belief, therefore, needed to be specified in non-epistemic terms, such as reliability and one's awareness of reliability. However, if the normative conception of justification equally concerns which beliefs are reasonable to hold, then we may ask of animal subjects whether they are *reasonable* in their beliefs. When we

ask this question, there is some resistance to the idea that they hold *reasonable* beliefs.

What is the source of this reluctance? One of the necessary conditions for epistemic justification – being blameless – is trivially satisfied in the case of animals. Specifically, if it is appropriate to attribute a belief-desire psychology to animals, they cannot be regarded as blameworthy in a way that could undermine knowledge, because this would imply a degree of responsibility which they intuitively lack. Kornblith offers the insight that it is more apt to construe animal's knowledge as grounded in the use of a reliable belief-forming procedure. Kornblith concludes that we should think of the knowledge animals have as requiring "proper grounding" rather than justification (p. 91). He observes how the notion of a justified or reasonable belief is not easily applicable to subjects that are not part of our linguistic community and are incapable of being blamed or praised or being held responsible for their beliefs (p. 90). If this is the case, then the situation is that attributions of justification are neither essential for explaining animal behaviour, nor do attributions of justification seem *pre-theoretically* plausible for reflectively unsophisticated subjects; this also includes infants, and perhaps small children.

One potential objection to the above view is since animals possess knowledge, their beliefs must be justified despite violating the awareness condition for justified belief. In particular, justification might be used as a technical locution, denoting the difference between true belief and knowledge.[166] A response along these lines would not, however, be a persuasive objection to awareness internalism. The reason for this is because awareness internalism is an attempt to explicate a normative conception of justified belief connoted with beliefs that are *reasonable* to hold. If is not apt to describe animal subjects as having *reasonable* beliefs, then using justification as a technical locution for animal beliefs, fails to relate to the normative conception of justification. It is possible to assert that animals have justified beliefs, but we would no longer be describing the normative conception of justification, which the thesis has argued requires second-order awareness.

Providing we are clear that awareness internalism is an account of a normative conception of justification, concerning beliefs that are reasonable to hold, then it is clear that the attribution of justification is not necessary for explaining animal behaviour, and it seems plausible that unsophisticated subjects also lack justified beliefs.

---

[166] The difference between true belief and knowledge is what Plantinga (1993a) terms "warrant".

Importantly, there is a difference between the knowledge attributed to animals on the basis of animal behavior directed at the desire to survive, and humans whose beliefs transcend their basest instincts. Humans believe and do things which are not merely directed at the procreation of the species; these beliefs are not always automatic responses to external stimuli. Moreover, we can intentionally act on these beliefs in ways contrary to self-interest. Consider the numerous cognitive activities with which we engage – reflecting, inquiring, revising, scrutinising; these are distinct from the cognitive activities of reflectively unsophisticated believers. Our epistemic practices display a heterogeneity, and our behaviour is not (at least in any obvious sense) reducible to our basest desires. Accordingly, when knowledge is attributed to animals, infants, and small children to explain behaviour, this does not nullify the hypothesis of the thesis, namely that there is a construal of awareness internalism and its relationship to knowledge as a species of non-lucky true belief, which represents a compelling conception of justification

Thesis Summary

In this thesis, I have supported a version of awareness internalism, one which requires that reliability is necessary (but insufficient) for justified belief. I would like, finally, to summarise how the research question has been addressed.

The research question asked whether the specific version of awareness internalism outlined in Chapter 1 could meet a range of intuitions about justification and knowledge. Consider, first, intuitions about justification. One being that justification and knowledge make reflective demands. As Chapter 2 indicated, we can sometimes explain a failure to hold a justified belief in terms of a failure to satisfy a higher-order awareness requirement for justified belief. Equally however, Chapter 2 and Chapter 3 observed that external failures often *explain* a belief's failure to be epistemically justified. Subjects may be blameless and yet epistemically justified; such as the blameless paranoid believer, and the blameless flat-earther living in a scientifically unenlightened community. Chapter 4 observed that, across cases, the explanation of blamelessness often differs - with some beliefs being excusable and others exempt from blame. But in each case, it seems that blamelessness is insufficient for justified belief. Further requirements for justified belief are needed.

In cases of blameless but unjustified belief, we can commonly identify an externalist failure: a failure of reliability or, more generally, a failure of a belief to *be successful* often *explains* a failure to be justified. The paradigmatically externalist accounts of justification explain that verdict but struggle to explain other cases.

By incorporating both a reliability condition and a second-order awareness requirement as conditions for justified belief, the proposal of this thesis has aimed to accommodate both the internalist insight that justification makes reflective demands and the externalist insight that justified belief requires reliably formed belief. If awareness internalism can make justification in part a function of success, then a core motivation for externalist reliabilism, namely that it explains many intuitive cases of justification failure, is also consistent with awareness internalism. By appealing to anti-luck considerations, there is a compelling motivation for awareness internalism: namely because in the absence of second-order awareness, one's beliefs will be epistemically lucky.

I considered the objection that the specification of awareness internalism able to be motivated by a concern with eliminating epistemic luck, generated two versions of scepticism. First, one contention examined was that internalists are targeting ineliminable varieties of epistemic luck (Chapter 2), but we demonstrated deficiencies

in that argument. Second, I examined (in Chapter 3), the argument that to meet anti-luck intuitions, internalists would need a specification of the awareness condition resulting in a sceptical regress. I explored the contention (in Chapter 3) that awareness internalism is trapped between a construal of awareness which either leads to a sceptical regress or offers us no reason to prefer internalism over externalism. I outlined a coping strategy, namely a "non-epistemic doxastic requirement", but outlined the concern that such a requirement would fail to explain intuitive cases of unjustified belief. In responding to that worry, I appealed to the hypothesis that awareness internalism requires awareness of externalist conditions. I argued that awareness internalism may avoid the regress compatibly with the anti-luck motivation for awareness internalism.

A further test for awareness internalism was whether it could explain our intuitions about intuitively justified beliefs held in radical sceptical scenarios. I examined, in Chapter 4, possible worlds in which one uses an unreliable method of belief formation because, for example, one is a disembodied brain-in-a-vat or is deceived by a demon. On first inspection, intuition suggested in radical sceptical worlds, one would continue to have justified beliefs. However, by drawing on Timothy Williamson's insights, it was apparent that the normative picture is more complex than on first impression; it is necessary also to explain an intuition about justification failure in radical sceptical scenarios; by making justification depend on externalist conditions, the intuition can be explained. By considering radical sceptical worlds, it was essential to distinguish between different ways in which a belief can be blameless by either being excusable or exempt. Chapter 4 argued that without distinguishing between different kinds of blamelessness, we would not be able to explain our comparative intuitions. Here, the concern was with a variety of modal luck: that it is counter-intuitive, in the actual world, that our putatively true beliefs are justified, but in demon worlds they have the same normative status as beliefs that are intuitively unjustified in the actual world.

The result of this inquiry was that awareness internalism as construed in Chapter 1 (§ 3), has the resources to explain a wide class of our epistemic intuitions without implying radical scepticism. The conception of knowledge as a variety of non-lucky true belief and the view that awareness internalism effectively targets epistemic luck, required examination of the varieties of epistemic luck and the conditions required to eliminate epistemic luck. I would hope that Chapter 5 has provided much more detail on the different conditions available, and the grounds we may have for supporting one condition over another.

A framework for competing intuitions

Awareness internalism, providing it requires awareness of externalist conditions, reconciles a class of intuitions that may otherwise seem opposed. For example, since that reliability conditions explain cases of justification failure (e.g. examples of cultural isolationism), if we were to identify justified belief with reliably-formed belief, then we would have no straightforward way of explaining counterexamples to externalism – specifically, counterexamples that challenge the *sufficiency* of reliability as a condition for justified belief (Chapter 2). Conversely, if we understand counterexamples to externalism as supporting a notion of internalism that does not require reliability, then we are left with no straightforward explanation of cases in which justification is undermined by reliability failure. The version of awareness internalism defended in this thesis offers a way of reconciling common intuitions, because it allows that each of these intuitions is making a legitimate insight about the requirements for justified belief.

I argued that awareness internalism is most defensible if the counterexamples it uses do not involve an outright rejection of externalist intuitions about the counterexamples (Chapter 1, § 4). I sought a non-dogmatic response to externalist intuitions about counterexamples, so opposing intuitions may both, at least partially, be correct. Here, I appealed to Hetherington's notion of minimal and non-minimal ways in which a belief can constitute knowledge. By doing so, I argued that awareness internalism does not need to involve a dogmatic assertion about the correctness of intuitions supporting awareness internalism. Instead, there is room for more than one way in which one's belief may be epistemically justified, even if the most *normatively satisfactory* knowledge requires second-order awareness.

Chapter 6 identified a challenge to awareness internalism, namely that reflectively unsophisticated subjects may know propositions despite not satisfying an awareness condition for justified belief. Knowledge attributions can explain and predict behavior and, for animals, how it is directed at survival. The knowledge invoked to explain behavior was not one which required the concept of epistemic justification in attributing knowledge. It did not pose a special problem for the concept of human knowledge which require justification and second-order awareness.

The connection between epistemic luck and justified belief and knowledge was used to motivate internalist and externalist intuitions. The application of anti-luck epistemology allowed us to see that there were anti-luck motivations for the version of awareness internalism identified in Chapter 1. We saw, furthermore, that appealing to epistemic luck to motivate internalism did not require us to think of internal

justification as rooted in fulfilling an epistemic duty or as consisting merely in blameless believing.

The view that emerges is an internalist view of justification – specifically, a view of awareness internalism which makes reliability a necessary condition for justified belief. However, awareness internalism sits alongside a qualified form of externalism about knowledge, which allows that some types of knowledge do not need to be internally justified. I argued that this explains a wide class of our intuitions without resulting in scepticism.

The Import for scepticism

Awareness internalism is only attractive if it does not lead to radical scepticism. I considered two versions of scepticism affecting the condition – an under-determination argument, which purported to show how reflective luck is ineliminable (Chapter 2) and the regress argument which purported to show that awareness internalism imposes unsatisfiable demands on justified belief. Importantly, as described in Chapter 2, (§ 2.2) and Chapter 3 (§ 2), both of the sceptical arguments facing awareness internalism rested on particularist assumptions about the necessary conditions for justified belief and knowledge. We saw how sceptical consequences would be avoided with a methodist approach to the conditions for justified belief and knowledge. However, the response I adopted was to provide a rationale for particularism and to try to accommodate sceptical objections. I did this, firstly, by questioning the assumption that internalists are *merely* targeting reflective luck (Chapter 2), and secondly, by understanding the awareness condition as not requiring an infinitely long chain of justified beliefs. Chapter 3 explained how the latter may be accomplished while preserving the anti-luck motivation for awareness internalism.

However, awareness internalism is not immune to scepticism. In Chapter 3 (§ 4), I identified a form of meta-epistemological scepticism. Specifically, the awareness condition provided no assurance that *in fact* our beliefs are epistemically justified. The meta-epistemological has been applied, by Fumerton, to externalist theories of justification, which he highlights do not provide philosophical assurance that the antecedents of their conditionals for justified belief are *in fact* satisfied by our beliefs. I explained how the same objection may be levelled against awareness internalism. In this respect, meta-epistemological scepticism did not provide a reason to prefer externalism over awareness internalism. However, it did – I think - reveal a limitation of the condition for justified belief. For if one wants to *prove* that a belief in a proposition is justified, then awareness internalism falls short of providing that assurance.

The above limitation creates a concern that awareness internalism, therefore, fails to offer a philosophically satisfactory account of justified belief. Indeed, Richard Fumerton has argued that the failure of externalist accounts of justification to provide us with assurance that the antecedents of their conditionals are satisfied, suggests they are failing to explicate a concept of justification of philosophical interest (Fumerton, 1995). Similarly, if awareness internalism does not offer assurance that the necessary reliability condition for justified belief is satisfied, then Fumerton's objection to externalism leaves us with philosophically uninteresting concept of justification which is also an objection also to the type of awareness internalism I have defended.

Fumerton's objection has an important implication for how I would want to present the version of awareness internalism supported in this thesis. In particular, Fumerton's objection means we should clarify that awareness internalism has been supported on explanatory rather than demonstrative grounds. I have attempted to explain our third-person intuitions about various cases of belief formation; how they may be reconciled with opposing intuitions about cases of belief formation and how a version of awareness internalism can meet those intuitions, while also avoiding scepticism. However, the scepticism I avoided was the scepticism which arises if one specifies conditions for justified belief that cannot be satisfied by finite creatures; this is distinct from the meta-epistemological scepticism, arising when one fails to demonstrate, or provide assurance, that one has justified beliefs. Awareness internalism avoids scepticism only insofar as it does not entail it; that is to say, awareness internalism is compatible with the possibility of justified belief and knowledge, but falls short of *demonstrating* that we do, in fact, have justified beliefs and knowledge. Awareness internalism thus offers an explanatory rather than a demonstrative response to scepticism.

Bibliography

Alson, W.P. (2005) *Beyond Justification: Dimensions of Epistemic Evaluation* (Cornell: Cornell University Press).

Armstrong, D.M. (1973). *Belief, Truth and Knowledge* (Cambridge: Cambridge University Press).

Audi, R. (2001) "An Internalist Theory of Normative Grounds", in *Philosophical Topics*, Vol. 29, No. 1/2, pp. 19-46.

Becker, K. (2007). *Epistemology Modalized*. (Oxford: Routledge).

Bergmann, M. (2006). *Justification Without Awareness* (Oxford: Oxford University Press).

BonJour, L. (1985). *The Structure of Empirical Knowledge*. (Abbreviated as "*SEK*" throughout the thesis). (Harvard: Harvard University Press).

BonJour, L. (1987). "Nozick, Externalism, and Skepticism" in "*The Possibility of Knowledge: Nozick and His Critics* (Ed. S. Luper-Foy), pp. 297-314.

BonJour, L. (2003). *Epistemic Justification: internalism vs externalism, foundations vs virtues*. (Oxford: Blackwell).

Chisholm, R.M. (1982). "*The Foundations of Knowing*" (Brighton, UK: The Harvester Press Limited).

Cohen, S. (1984). "Justification and truth", *Philosophical Studies: An International Journal for Philosophy in the Ananltic Tradition*, Vol. 46, No. 3, pp. 279-295.

Cohen, S. (2002). "Basic Knowledge and the Problem of Easy Knowledge" *Philosophy and Phenomenological Research*, Vol. 65, No. 2, pp. 309-329.

DeRose, K. (1995). "Solving the Skeptical Problem", *The Philosophical Review*, Vol. 104, No. 1, pp. 1-52.

DeRose, K. (2004). "Sosa, Safety, Sensitivity, and Skeptical Hypotheses" in John Greco (Ed), *Ernest Sosa: And His Critics*, pp. 22-41. (Oxford: Blackwell).

Dretske, F. (1970). "Epistemic Operators", *The Journal of Philosophy*, Vol. 67, No. 24. pp. 1007-1023.

Engel, M. (1992a). "Is Epistemic Luck Compatible with Knowledge", *The Southern Journal of Philosophy*, 30(2), pp. 59 – 72.

Engel, M. (1992b). "Personal and Doxastic Justification in Epistemology", *Philosophical Studies*, No. 67. Pp. 133-150.

Feldman, R; Conee, E. (1985). "Evidentialism", *Philosohpical Studies*, Volume 48 (Vol. 15), pp. 15-34.

Feldman, R. (2000). "The Ethics of Belief", *Philosophy and Phenomenological Research*, Vol. 60, No. 3. Pp. 667-695).

Feldman, R. (2001); Conee, Earl. "Internalism Defended", *American Philosophical Quarterly*, Vol. 38, No. 1. pp. 1-18.

Fumerton, R. (1995). "*Meta-Epistemology and Skepticism*" (Rowman and Littlefield).

Gettier, E. (1963). "Is Justified True Belief Knowledge?", *Analysis*, Vol. 23, No. 6, pp. 121-123.

Greco, J. (2003) "Virtue and Luck, Epistemic and Otherwise" *Metaphilosophy*, 34(3) pp. 353- 366.

Greco, J. (2012). Better Safe than Sensitive. In K. Becker and T. Black (Eds.), *The Sensitivity Principle in Epistemology* (pp. 193-206). (Cambridge: Cambridge University Press).

Grefte, J. (2017). "Epistemic Justification and epistemic luck", *Synthese*, Vol. 159, No. 9, pp. 3821-3836.

Goldman, A. (1979). "What is Justified Belief" in Justification and Knowledge, George S. Pappas (Eds.,) (London: Reidel), pp. 1-12.

Goldman, A. (1988). "Strong and Weak Justification", *Philosophical Perspectives,* Vol. 2, pp. 51-69.

Harman, G. (1973). *Thought*. (Princeton: Princeton University Press).

Hetherington, S. (2001). *Good Knowledge, Bad Knowledge: On two dogmas in epistemology*. (Oxford: Oxford University Press).

Kim, J. (1984). Concepts of Supervenience", *Philosophy and Phenomenological Research*, Vol. 45, No. 2, pp. 153-176.

Klagge, J.C. (1988). "Supervenience: Ontological and Ascriptive" *Australasian Journal of Philosophy",* (66:4), pp. 461-47

Kornblith, H. (2002). *"Knowledge and its Place in Nature"* (Oxford: Oxford University Press).

Lewis, D. (1973a). *Counterfactuals*, (Oxford: Oxford University Press).

Lewis, D. (1973b). "Counterfactuals and Comparative Possibility", *Journal of Philosophical Logic*, 2(4). pp. 418-446.

Lewis, D. (1996). "Elusive Knowledge", *Australasian Journal of Philosophy* (74:4), pp. 549-567.

Littlejohn, C. (2009). "The Externalist's Demon". *Canadian Journal of Philosophy*, Vol. 32, Number 3. Pp. 399-434

Moller, N; Hansson, S; Peterson, M (2006). "Safety is more than the antonym of risk", *Journal of Applied Philosophy*, Vol. 23, No. 4, pp. 419-432.

Moore, G.E. (1918). "Some Judgments of Perception", *Proceedings of the Aristotelian Society*, 19, pp. 1-28.

Moore, G.E. (1925). "A Defence of Commonsense" in Thomas Baldwin (ed), *G.E. Moore: Selected Writings*, pp. 106-133, (London: Routledge).

Moore, G.E. (1939). 'Proof of an External World', in Thomas Baldwin (ed), 2013, *G.E. Moore: Selected Writings,* pp. 146-171, (London: Routledge).

Neil, C. (2019). "Safety, domination, and differential support", *Synthese*. https://doi.org/10.1007/s11229-019-02091-2.

Nozick, R. (1981). *Philosophical Explanations*. (Oxford: Oxford University Press).

Pilley, J. (2013). "Border collie comprehends sentences containing a prepositional object, verb, and direct object" in *Learning and Motivation* (44), pp. 229-240.

Plantinga, A. (1993a). *Warrant: the current debate* (Oxford: Oxford University Press).

Plantinga, A. (1993b). *Warrant and Proper Function*. (Oxford: Oxford University Press).

Pritchard, D. (2002). "Resurrecting the Moorean Response to the Sceptic", *Philosophical Studies* 10(3), pp. 283-307.

Pritchard, D (2005). *Epistemic Luck* (Oxford: Oxford University Press).

Pritchard, D. (2008). "Sensitivity, Safety, and Anti-Luck Epistemology", in John Greco (ed) *The Oxford Handbook of Scepticism*. pp. 438-453.

Pritchard, D. (2012). "In Defence of Modest Anti-Luck Epistemology" in Kelly Becker and Tim Black (Ed), *The Sensitivity Principle in Epistemology*, pp. 173-192.

Pritchard, D. (2015). "Anti-Luck Epistemology and the Gettier Problem", *Philosophical Studies*, 172, pp. 93-111.

Pritchard, D. (2016). "Epistemic Risk", *Journal of Philosophy*, Vol. 113, Issue 11, pp. 550-571.

Putnam, H. (1973). "Meaning and Reference", *The Journal of Philosophy*, Vol. 70 (No. 19), pp. 699-711.

Putnam, H. (1981). *Reason, Truth, and History*. (Cambridge: Cambridge University Press).

Pollock, J. (1974). *Knowledge and Justification* (Princeton: Princeton University Press).

Radford, C..(1966). "Knowledge by Examples", *Analysis*, Vol. 27, No. 1, pp. 1-11

Russell, B. (1948). *Human Knowledge: Its Scope and Limits*. (London: George Allen and Unwin).

Sosa, E. (1999). "How to Defeat Opposition to Moore", Nous, 33 (s13), pp. 141-153.

Sosa, E. (2007). *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume 1*. (Oxford: Oxford University Press).

Sosa, E. (2009). *Reflective Knowledge: Apt Belief and Reflective Knowledge, Volume 2* (Oxford: Oxford University Press).

Unger, P. (1968). "An Analysis of Factual Knowledge", *The Journal of Philosophy*, 65(6), pp. 157-170.

Van Cleve, J. (1985). "Epistemic Supervenience and the Circle of Belief", *The Monist*, pp. 90-104.

Williams, B. (1970). "Deciding to Believe" in *Problems of the Self: Philosophical Papers 1956-1972*. Pp. 136-151 (Cambridge: Cambridge University Press).

Williams, M. (1996). Unnatural Doubts (Princeton: Princeton University Press).

Williamson. T. (Forthcoming) "Justifications, Excuses, and Sceptical Scenarios".http://media.philosophy.ox.ac.uk/assets/pdf_file/0018/35343/Luxembourg.pdf

Williamson, T. (2000). *Knowledge and its Limits*. (Oxford: Oxford University Press)

Williamson, T. (2007). "On Being Justified in One's Head", in *Rationality and the Good: Critical Essays on the Ethics and Epistemology of Robert Audi*, (Ed. M. Timmons, J. Greco, A. Mele). Oxford: Oxford University Press.

Wright, C. (2002) "Anti-Sceptics: Simple and Subtle", *Philosophy and Phenomenological Research*, Vol. 65, No. 2, pp. 330-348.

Zagzebski, L. (1994). "The Inescapability of Gettier Problems", *The Philosophical Quarterly*, 44(174), pp. 65-73.

Zalabardo, J. (2017). "Safety, Sensitivity and Differential Support", *Synthese* https://doi.org/10.1007/s11229-017-1645-z.