



# THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

**Externalist Epistemology and the Constitution of Cognitive Abilities**

Evan Butts

PhD by Research in Philosophy

The University of Edinburgh

2011

## Table of Contents

<b>Declaration</b> .....	<b>iii</b>
<b>Acknowledgements</b> .....	<b>iv</b>
<b>Abstract</b> .....	<b>vi</b>
<b>Foreword</b> .....	<b>vii</b>
<b>Chapter 1</b> .....	<b>1</b>
<i>Internalism and Externalism in Philosophy of Cognitive Science</i>	
<b>Chapter 2</b> .....	<b>26</b>
<i>Internalism and Externalism in Epistemology</i>	
<b>Chapter 3</b> .....	<b>49</b>
<i>Ability in Philosophy of Cognitive Science and Epistemology</i>	
<b>Chapter 4</b> .....	<b>75</b>
<i>From Epistemic Externalism to Cognitive Externalism</i>	
<b>Chapter 5</b> .....	<b>102</b>
<i>Mentalism is Not Epistemic Ur-Internalism</i>	
<b>Chapter 6</b> .....	<b>135</b>
<i>From Epistemic Anti-Individualism to Cognitive Externalism</i>	
<b>Afterword</b> .....	<b>160</b>
<b>References</b> .....	<b>165</b>

## **Declaration**

I, Evan Thomas Butts, declare the following. This thesis, being submitted for examination in pursuit of a PhD by Research in philosophy, has been wholly composed by me. The work contained herein is my own, and has not been submitted in pursuit of any other academic degree or professional qualification.

Signature:

Date:

## Acknowledgements

A PhD thesis requires patience and ability not only of the author, but also of the author's family, friends and colleagues. Here I list those people who have been singularly helpful to and patient with me, especially as deadlines began to loom.

First, I owe a great debt to my advisors, Duncan Pritchard, Jesper Kallestrup, Mark Sprevak and Andy Clark. Each of them has given more generously of their time than required by their role. They have not only contributed immeasurably to the progression of the thesis itself, but they have also done much to prepare me for the profession into which I soon hope to enter.

Next, I thank my parents, who did not despair of their only son when he embarked upon a career in academic philosophy. The list of things for which I owe them thanks cannot be inscribed here, or else there would be no space left for the arguments I aim to give. I thank my father, Donald, for his wisdom and experience concerning the plight of the burgeoning academic, as well as his no-nonsense encouragement, "Write, boy! Write!" I thank my mother for her lifelong efforts to reform my natural tendency towards what she once called "tortured, Byzantine prose" into the relatively limpid style I now strive to employ.

The faculty and postgraduate community of the University of Edinburgh have been friendly and helpful throughout my stay with them. A few people (beyond my previously mentioned advisors) deserve special note. Allan Hazlett has provided much lively and entertaining discussion on a broad range of topics, not the least of which was my discussion of epistemic internalism. Orestis Palermos and I share almost identical philosophical interests, and have as such had lots of useful discussion on a variety of matters pertaining to my thesis, especially the notion of ability in epistemology. Finally, Joey Pollock dauntlessly encouraged me—even when I told her not to. Moreover, she has listened to the breadth of the arguments from my thesis, providing useful feedback and pointing me in the direction of extremely helpful papers.

A philosopher who wishes to remain sane and (relatively) normal needs friends who are not themselves philosophers. So last—but not least—I thank the members of the Edinburgh University Triathlon club. Their contribution to my continued sanity has been twofold. First, they have been good friends. Second, the rigorous physical training I have enjoyed as a member of the club has helped give structure to my days and has kept me healthy and energized during what might have otherwise been an unhealthy and indolent time. A few people merit particular mention. I would like to thank Declan Wilson for patiently coaching my swimming technique and for excellent banter. Thanks go to Finlay McAndrew for deepening my love and appreciation of running, and his scientific advice on health and nutrition. Final thanks go to my frequent training companions and good friends, Kate Cora, Lyndsey Cruickshanks, Roisín Fahey and David Steele.

Thanks to all those above for keeping me on the rails.

## **Abstract**

Cognitive abilities have been invoked to do much work in externalist epistemology. An ability condition (sometimes in conjunction with a separate, anti-luck condition) is seen to be key in satisfying direction-of-fit and modal stability intuitions which attach to the accrual of positive epistemic status to doxastic attitudes. While the notion of ability has been given some extensive treatment in the literature (especially John Greco, Alan Millar and Ernest Sosa), the implications for these abilities being particularly cognitive ones has been given less attention. To rectify this oversight, I examine the debate over the nature of cognition from philosophy of cognitive science, paying particular attention to the debate between defenders of internalist theories (Fred Adams, Kenneth Aizawa and Rob Rupert) and externalist theories (so-called “extended mind” positions). Armed with substantive accounts of cognition, I argue that the epistemological externalist’s obligation to repudiate epistemological internalism forces her to adopt some sort of externalist account of cognition.

## Foreword

‘Internalism’ and ‘externalism’ are terms which have been applied to sets of opposing positions in numerous philosophical debates. For instance, in ethics and moral philosophy there is a distinction between internalism and externalism about reasons for action (sometimes called reasons internalism-externalism, and sometimes motivational internalism-externalism). Briefly, an internalist about reasons for action claims that one has a reason to perform some action only if one has some subjective motivational reason to do so. (E.g., I have reason to save the proverbial [and ubiquitous!] drowning child if—and only if—I desire to do so, or if it would be good from my perspective to do so, etc.) In contrast, the externalist about reasons for action says that one may be motivated to perform some action for reasons not a part of one’s subjective motivational state. (E.g., I have reason to save the proverbial drowning child if it is [objectively] good or right to do so.) We will not discuss this internalism-externalism debate here.

In philosophy of mind, there is the notorious debate about mental content, which has sometimes gone under the heading of internalism versus externalism about mental content. (Of course, this debate has also been cashed-out in terms of individualism versus anti-individualism and so-called “broad” versus “narrow” content.) Here, the issue has been how to individuate the contents of subjects’ thoughts, often construed as the attitudes



subjects bear towards propositions (e.g., bearing the attitude of belief towards the proposition *the sun is shining*). Proponents of content internalism claim that the factors which individuate content are “inside” that to which we predicate the content in question. There is not full agreement as to what exactly this means, but it has been variously claimed that ‘inside’ in this context means either “in the head” (*pace* the famous Putnam slogan), “internal to the individual”, or “part of the mind of the individual”. Content externalists deny that the factors which individuate mental contents are so circumscribed. In the course of this work, I will talk a little about content internalism-externalism—but it is not the focus of my arguments. For the most part, content internalism-externalism will serve as a point of comparison and contrast for the other internalism-externalism disputes upon which I will focus.

Equally notorious—if not more so—is the internalism versus externalism debate in epistemology. Strictly speaking, the internalism-externalism clash arises in many epistemological disputes, nearly one-for-one with the different epistemic statuses which epistemologists posit. One may defend internalism (or externalism) with regard to knowledge, justification, rationality and evidence, to name a few. Moreover, it is usually argued that defending internalism (or externalism) about one epistemic status does not thereby commit you to a corresponding claim about another epistemic status. While there is significant agreement about what ‘internal’ means in the epistemological context (in contrast to the debate concerning mental content), opinion is not unanimous in this regard. There is a small (but vocal) minority who insist upon a sense of ‘internal’ which they claim to be distinct from that accepted by the majority of theorists. The majority of theorists take epistemic internality to be characterised by some sort of first-person or reflective access to certain facts. The minority view, on the other hand, wants to assimilate epistemic internality to (their take on) the sense of ‘internal’ at play in the mental content debates. The dispute between the former and the latter—so-called “accessibilism” and “mentalism”, respectively—will be important for the purposes of this work. Both sides will first be briefly set out in Chapter 2. The main argument of this work (found in Chapter 4) will advert to both positions, but not the actual dispute between them. It is in Chapter 5 where we will

take up the issue of whether accessibilism or mentalism identifies the basic, primary notion of internality in epistemology.

Finally, disputes about cognitive and mental processes have been carried out in terms of internalism and externalism. In this debate, perhaps the most well-known formulation of the externalist position often goes under the heading “the extended mind thesis”, due to Andy Clark and David Chalmers (1998). Subsequently, varieties of internalism-externalism dichotomies have been coined in an attempt to capture the debate. For example, there have been the following: vehicle internalism-externalism (due to Susan Hurley 1998a), locational internalism externalism (due to Rob Wilson 2004), cognitive internalism-externalism (due to Richard Menary 2007) and psychological internalism-externalism<sup>1</sup>. In the debate on cognition, the sense of ‘internal’ at least has a clear outer boundary at the skin of individual agents<sup>2</sup>. Whether the boundary of the internal is “farther in” (i.e., the skull rather than the skin) is an in-house debate between some opponents of the extended mind thesis and others. In any case, internalism-externalism about cognition will be the other debate of primary interest to us. Chapter 1 will comprise a sampling and general overview of the different positions regarding cognition. When we come to the main argument in Chapter 4, the relation between internalism-externalism concerning cognition and epistemic internalism-externalism will be explored.

There is no immediately apparent reason why any of these debates should have anything to do with each other. After all, simply because ‘internalism’ and ‘externalism’ are used to label the opposing sides of various debates should give us no cause to look for similarities or relations among said debates. Moreover, upon seeing the disparate senses of ‘internal’ and ‘external’ deployed in the various debates, we might have reasons to think the debates were not similar at all. That being said, there have been interesting issues broached concerning the ramifications of content externalism in epistemology. Specifically, much has been said about what content externalism (i.e., the idea that mental contents are individuated by factors “outside” the mind or subject to which those contents are attributed) implies for

---

<sup>1</sup>Note that this distinction is also used in debate concerning mental content.

the view that agents have some sort of privileged epistemic access to the contents of their thoughts<sup>3</sup>. That is just to say, internalism-externalism debates in one philosophical sub-discipline may have consequences in another sub-discipline. So, it is certainly not beyond imagining that an internalism-externalism debate in epistemology might bear upon an internalism-externalism debate in philosophy of cognitive science. In fact, it is particularly the relation between these two internalism-externalism debates that I wish to explore.

Attention has been given to the interface of epistemology with cognitive science and psychology. Suggestions as to what the relation is among these disciplines have ranged from the aggressive naturalism of Quine (1969) to the more moderate position seen in Goldman (1986). Further, there has been recent work specifically concerning the relation between externalism about cognition and various forms of epistemic externalism. Particularly, there has been an exchange over how well virtue epistemology fits with the externalism about cognition. (See, e.g., Kelp [forthcoming] and Vaesan [forthcoming].) These debates seek to adjudicate whether externalism about cognition tells in favour of one version of externalist epistemology over another. (Vaesan claims virtue epistemology does not fit well with externalism about cognition. Kelp provides a defence of virtue epistemology.) My argument will differ from the ones found in this debate. I will not attempt to argue for particular versions of epistemic externalism over others on the basis of externalism about cognition. As such, the arguments in said debate will play no role in my own argument. Instead, the purpose of this work is to explore more general implications for externalist epistemology in the face of the internalism-externalism debate in philosophy of cognitive science. These more general implications have not, to my knowledge, been explored elsewhere.

Let me get to the point then. In what follows, I will argue that there is a connection between a certain sort of epistemic externalism (*viz.*, those who rely on what is known as “the ability intuition”) and externalism in philosophy of cognitive science. Specifically, I intend to make a case for a commitment on the part of epistemic externalists (of the relevant type) to some form of externalism about cognition. I take this to be an interesting point

---

<sup>2</sup> Note, however, that some proponents of so-called “embodied cognition” theses may consider their positions to be of the extended-mind type with respect to what they take to be the traditional (and outmoded) notion of cognition being limited to the brain.

because some externalist epistemologists explicitly deny such a commitment, or at least suggest that their epistemic externalism is neutral with regard to the debate in philosophy of cognitive science. However, even if no epistemic externalist had anything to say about the matter, insofar as externalism about cognition is a radical and controversial thesis, a commitment to externalism in philosophy of cognitive science may be seen as an unwelcome burden to externalist epistemology.

Here is the plan for my argument. In the first two chapters, I will give an overview of both internalist and externalist positions in philosophy of cognitive science and epistemology. This information will be crucial to isolating the senses of ‘internal’ and ‘external’ used in the respective debates so that we can see in what way these positions interact. Further, in Chapter 2 I will specify exactly which forms of epistemic externalism are the targets of my argument by discussing the ability intuition in epistemology. Specifically, the targets of my main argument will be externalist theories of justification which invoke an ability condition. (As we will see, it is *cognitive* abilities in particular to which the epistemic externalist appeals.) Finally, both chapters will delineate the conditions under which the relevant brand of externalism would be true by pinning down what precisely the respective externalist must deny with her repudiation of her respective opponents.

As I have already briefly mentioned, the specific brands of epistemic externalism with which we will be concerned are those which rely on an ability condition. As such, in Chapter 3 I take up the task of discussing the notion of ability itself. The account of ability I examine is independent from debates in both philosophy of cognitive science and epistemology as it is concerned with the conditions in which it is appropriate to attribute abilities to agents in general. In addition, I will also explore how the general notion of ability relates to the way ability (or some closely-related notion) is used in philosophy of cognitive science and epistemology. This task is crucial because it is the epistemic externalist’s appeal to cognitive abilities through which I make my argument to a commitment on their part to externalism about cognition. At the end of Chapter 3, I will make a brief attempt at constructing an argument directly from the notion of ability actually used by epistemic

---

<sup>3</sup> For example, see Brown (2004).

externalists to externalism about cognition. However, I will subsequently demonstrate that such an argument is not in the offing.

In Chapter 4, I will consider a second and more serious argument. Having shown in the previous chapter that an argument directly from the notion of ability actually used by epistemic externalists does not commit them to externalism in philosophy of cognitive science, I will change tack. The argument in Chapter 4 will proceed via two considerations. First, I will restate exactly what the epistemic externalist must deny by her rejection of epistemic internalism. Second, I will explore in what way the epistemic externalist's denial of epistemic internalism affects her appeal to cognitive abilities. I argue that her commitment to reject epistemic internalism forces the epistemic externalist to adopt an externalist position about cognitive abilities. More exactly, I argue that it is the epistemic externalist's commitment to rejecting mentalism which engenders her commitment to externalism in philosophy of cognitive science.

I will then discuss a few possible objections to this argument, before identifying a more significant objection. During the course of responding to these objections, I will further specify and clarify what epistemic externalist positions are affected by my argument. The latter, more significant objection will be the focus of Chapter 5.

In Chapter 5, I consider the claim that one of the forms of epistemic internalism—*viz.*, mentalism—which I suggest the epistemic externalist must deny, is not actually a form of epistemic internalism at all. As such, the immediately obvious reasons for claiming that the epistemic externalist must deny mentalism evaporate. However, if that is the case, then the basis for my argument in Chapter 4 also appears to evaporate. Thus, my attempt to show that epistemic externalist has a commitment to some form of externalism about cognition will fail. I will argue that mentalism is not, in fact, a form of epistemic internalism. As such, I provide grounds to reject the argument of Chapter 4.

However, in Chapter 6, I will demonstrate that the epistemic externalist should reject mentalism anyway, and that the commitment to externalism about cognition on the epistemic externalist's part does indeed ensue on the basis of a third and final argument. Because the commitment is still engendered by the denial of mentalism, the new argument will be quite close to that of Chapter 4. This last argument will be based on the distinction

made by Sanford Goldberg (2007) between epistemic individualism and anti-individualism. Thus, the aim of this work will be achieved.

By way of conclusion, I will move on to explore the consequences of the commitment I will have established. In particular, I will discuss in what ways such a commitment might be problematic for the epistemic externalist. Ultimately, however, I will claim that there are moves the epistemic externalist can make (consistent with her epistemic externalism) which render a commitment to externalism about cognition unproblematic. Importantly, I will demonstrate that these moves are independently motivated by more purely epistemological considerations. Thus, making these moves will not be merely a desperate attempt to dissolve any cognitive dissonance on the part of the epistemic externalist.

Hence, I will conclude that the epistemic externalist who appeals to cognitive abilities commits herself to externalism about cognition. However, she should not be too worried about such a commitment. Cleaving to externalism about cognition will force the epistemic externalist to give up nothing that she does not already have independent reason to relinquish.



## **Chapter 1**

### **Internalism and Externalism in Philosophy of Cognitive Science**

#### *Introduction*

What are the factors which constitute and explain our mental and cognitive abilities, processes, and states, etc.? This is the motivating question in the internalism-externalism debate in philosophy of mind and cognitive science. Importantly, the question is not directly about where the mind is. Instead, the goal is to identify which factors are needed to give a satisfactory explanation of phenomena like belief, vision, memory and so forth. Only after the fact should we note whether the objects and processes appealed to are either all within the brain; whether they are all within the brain and body; or whether they are in the brain, body and the material and/or social world. It is not that where these factors are is not interesting or important—it is just that, in principle, the location of the factors should not matter as to whether or not they do explanatory work.

The claim that the factors to which we will need to appeal extend beyond the boundaries of body-delineated individuals comes in many varieties, and goes under various names. We will examine some of these varieties in due course, and talk about them in their own terms. For simplicity and clarity, I will use the umbrella term “cognitive externalism” to



talk about any theory which claims that factors outside of bodily individuals are part of cognitive processes, states, etc. Opposing cognitive externalist theories are those which claim that all of the factors constitutive of cognition are within bodily individuals. Some circumscribe the spread of cognitive processes more narrowly than do others, claiming that we need not look beyond the brain. In any case, whether the opponents of cognitive externalism locate the factors constitutive of cognition solely within the brain or not, they are united in opposing the notion that any factors outside of the boundaries of the body do any cognitively constitutive work. Again, for simplicity and clarity, I will refer to all theories opposed to cognitive externalism as “cognitive internalist” theories<sup>4</sup>.

Before we become absorbed by cataloguing these two sorts of positions, let me stress something by way of reiteration. The spatial distribution of factors putatively constituting cognitive phenomena is not of first importance. What is primary is which sets of factors give the best explanation of various cognitive phenomena. After the relative explanatory merits of different suggested sets of factors are settled, we can look and see whether the most promising set spreads outside of the bodily individual. That is to say, the possibility that the best explanation of cognitive phenomena will advert to factors beyond the body is not ruled out from the start. Cognitive externalism is possible, and not only logically so; it might now be the case—or could in the future become the case—in the actual world. Throughout, we will assume that all parties to the debate accept this rather weak assumption. Indeed, the two positions opposed to cognitive externalism upon which we will focus—Ken Adams & Fred Aizawa (2008) and Rob Rupert (2009)—agree that cognitive externalism is possible. The real debate about cognitive externalism concerns whether it obtains in the actual world: whether it is the best explanation for the cognitive and mental phenomena we do observe. It is not the goal of this chapter to take a stance as to whether cognitive externalism obtains in the actual world. All we will establish is what would need to be the case for cognitive externalism to describe our world.

Before we can do this, however, we should take a close look at what the various parties to this debate claim. First, we will look at the cognitive internalists. To keep things manageable, I will limit the investigation to the opponents of cognitive externalism already

---

<sup>4</sup> The terms ‘cognitive internalism’ and ‘cognitive externalism’ are taken from Menary (2007).

mentioned—Adams & Aizawa and Rupert. There are substantive disagreements between the positions espoused in their respective works, but they agree that radical positions which locate factors explanatorily relevant to our mental and cognitive capacities outside of the individual organism are wrong. We will not ignore the disagreements between Adams & Aizawa and Rupert, however. These disagreements will make it more difficult to formulate an explicit statement of what needs to be the case for cognitive externalism to be confirmed, but they will also make the resulting statement more useful.

§1: *Adams & Aizawa's Cognitive Internalism*<sup>5</sup>

Adams & Aizawa christen their version of cognitive internalism “intracranialism” to highlight which objects and processes they believe suffice to explain mental and cognitive phenomena. Specifically, Adams & Aizawa claim that the factors which underpin cognition are solely within the brain. While they think, strictly speaking, that only intracranial factors are relevant, Adams & Aizawa explicitly note that it is consistent with their position to include extracranial nervous tissue (18). Hence, for purposes of our discussion, ‘brain’ should be understood to have the following non-standard extension: all anatomical portions of the body usually collected under the term ‘brain’ plus the nervous system—wherever in the body neurological tissue may be found. Because ‘brain’ has this non-standard extension in our discussion, the literal meaning of ‘extracranial’ is somewhat misleading. Hence, we should also adopt a non-standard extension of ‘extracranial’ for our purposes. Instead of meaning “outside the cranium”, ‘extracranial’ will mean “outside the brain and nervous system”. Adams & Aizawa’s intracranialist cognitive internalism, interpreted in this way, is the claim that no factors outside of the brain and nervous system are constitutive of mental and cognitive phenomena. I suggest these verbal gymnastics to maximize the charitability of my interpretation of Adams & Aizawa, and to avoid an unnecessary proliferation of terminology. The terminological terrain of the various internalism-externalism debates is already too complicated.

---

<sup>5</sup> All citations of Adams & Aizawa are from their 2008 book, *The Bounds of Cognition*. However, also see Adams & Aizawa (2001) for a paper-length treatment of the same issues.

The intracranialist thesis is sustained by two restrictions. First, Adams & Aizawa maintain that genuine cognition involves only operations over so-called “original” or “non-derived” (sometimes, “underived”) content. Determining the exact nature of such content is a complicated issue, and one which we will not explore here<sup>6</sup>. It will suffice for our purposes to get a general grasp of what Adams & Aizawa are after. Second, Adams & Aizawa argue for a particular interpretation of the scientific practice for individuating process types, which general principles they think apply as well in cognitive psychology. These practices, they claim, do not support multiple realization intuitions for mental and cognitive processes. Specifically, Adams & Aizawa think that there is no support to be found for the arguments, made by some content externalists, that there is an explanatorily useful notion of cognition which is broader than the categories found in standard cognitive psychology (§4.3). The combination of these two restrictions gives the full explanation of what Adams & Aizawa take to characterize cognition as it in fact occurs: certain kinds of processing over non-derived contents.

Again, I stress that the intracranialist position is not itself a conceptual restriction on where explanatorily relevant factors for cognition can be found. Adams & Aizawa do not obviously beg the question against cognitive externalism. Instead, they argue for the two constraints mentioned above and claim that it follows from the constraints that, in fact, there are no extracranial factors (where this term has the non-standard extension mentioned earlier) which are constitutive of mental and cognitive phenomenon. Thus, the restrictions will not feature in our ultimate specification of what it is the cognitive externalist denies when she rejects cognitive internalism. As such, our exploration of the two restrictions will be brief.

The distinction between derived and non-derived content is stated by Adams & Aizawa as follows.

Roughly speaking, the idea is that derived content arises from the way in which items are handled or treated by intentional agents. For the most part, things with derived content are assigned that content by intentional agents who already have thoughts with meaning. Underived content arises from

---

<sup>6</sup>The notion of non-derived content has been criticized extensively. E.g., see Clark (2006).

conditions that do not require independent or prior existence of other content, representations, or intentional agents (32).

The non-derived content criterion is meant to tell against heterodox theories which cite social factors like language (and other symbol-utilizing systems) as those which extend cognition beyond the brain of an individual.

The ubiquitous toy example in this case is the accountant doing arithmetic with a pad and pencil. It is claimed by some cognitive externalists that the cognitive phenomenon (i.e., solving a sum, quotient, product, etc.) in the particular case is best explained as comprising not only any internal calculation the accountant performs, but also the external calculation she performs on the pad. In the case of a large sum (e.g.,  $25,967 + 33,925$ ), the internal calculations are the additions of the individual columns (e.g., adding 5 to 7 in the fifth column). The external parts of the procedure are any carry-over digits from the internal calculations of the sum of a column, which are written above the next column to be added (e.g., carrying the 1 from the addition of 5 and 7 in the fifth column over to the fourth column). Again, this back-and-forth, in-and-out procedure is alleged to be the proper explanation of the accountant's performance.

Adams & Aizawa will claim, however, that the numbers and symbols written on the pad—and hence any operations performed over them—will not be genuinely cognitive because those numbers and symbols derive their meaning from a more basic source of meaning—*viz.*, the accountant. Her number-thoughts already had content, which she invested in the marks on the page. Hence, the content of those marks is derived from the contents of the accountant's thoughts. Adams & Aizawa nod towards the sorts of accounts which might explain the nature and provenance of non-derived content (Stitch & Warfield 1994; Searle 1980; 1984). However, we will not be concerned here with such accounts.

We might be concerned, though, with why non-derived content should be considered a “mark of the cognitive” (i.e., a feature which sets cognitive phenomena apart from non-cognitive phenomena). Adams & Aizawa here appeal to the representational nature of thought. They take themselves to be justified in invoking the notion of representation because of its prevalence in cognitive psychology (32-33). That these

representations are underived is supposed to be the best way to explain how a creature could have thoughts in the absence of any other creature having thoughts.

Part of what motivates the view that cognitive processes involve underived [emphasis removed] representations is that such content appears to be required to explain lone thinkers. Some hundreds of millions of years ago, the brain of some primitive fish evolved in such a way as to incorporate a fundamentally new type of state. That primitive fish's brain contained a thought or belief . . . And, whatever that first content was, it could not have been had in virtue of the contentful states of another organism, since *ex hypothesi* there were no such states (33).

Again, whether or not the story Adams & Aizawa tell is plausible is not the point here. Now armed with an understanding of what motivates the non-derived content restraint, we shall move on to Adams & Aizawa's second constraint.

Once again, here is a quote from Adams & Aizawa to provide a basis for discussing their criteria.

Given, however, that cognitive processes are not spread throughout the whole of space-time, one would expect that processes as distinctive as cognitive processes would supervene on correspondingly distinctive lower-level processes. Roughly speaking, lower level processes should be as distinctive as the higher-level processes they realize (69).

This is one intuition driving Adams & Aizawa's argument for their second restriction on what should count as cognitive. Further, they take the bulk of cognitive psychological practice—and indeed, scientific practice generally—to support this intuition. For the most part, claim Adams & Aizawa, science individuates phenomena by underlying mechanisms. Where personal-level observation might lump phenomena together because of similarities in observed effects, rigorous science makes distinctions based on the mechanism which causes that effect in different instances. Among the numerous examples Adams & Aizawa adduce are physiological features of animals. While both sharks and dolphins both have similar streamlined fin and body structures, these similar physiological features arose from different evolutionary lines and at different times in evolutionary history (59). Thus, evolutionary

biologists consider these physiological features to be distinct despite having identical functions. Applying this to the study of cognition, Adams & Aizawa want us to conclude that something is not cognitive merely because it achieves a certain result, but because it achieves that result in a certain way—or perhaps because it has a certain phylogeny.

In the case of cognitive psychology, Adams & Aizawa adduce specific examples to support their intuition. One of the sets of studies to which they appeal concerns memory. These examples are meant to illustrate the tendency of “...cognitive psychologists [to] strive to individuate cognitive processes by reference to underlying laws or mechanisms” (63). Briefly, Adams & Aizawa claim that memory studies have suggested certain effects, like primacy and recency, which are best described by appealing to laws of memory or specific memory mechanisms. Importantly, such effects would not be exhibited in the case of extended systems comprising individuals and artefacts.

Consider the primacy and recency effects exhibited in studies on memory (Anderson 1995, as cited in Adams & Aizawa). When presented with a list of words, subjects tend to have superior recall of the first few items on the list (primacy) and the last few items on the list (recency). Such effects would not manifest in an “agent” comprising an individual and her notebook, which notebook is supposed to be an external version of her normal biological memory. Such an “agent”, when presented with the same list of words as an agent lacking a notebook, can write them in her notebook. The agent with the notebook will “recall” all items on the list equally well. However, Adams & Aizawa claim that we should not say that the notebook-possessing subject has a better memory than the subject who lacks the notebook does. Why not?

Adams & Aizawa argue that there is no reason to expect a unified science of cognition which counts both the individual-notebook conglomeration and the un-augmented individual as cognitive. After all, what is the process in common between biological memory and notebook “memory”? More generally, “[what] could the processes found in [any] of these combinations [of individuals and artefacts] have in common that make them cognitive processes?” (73). Setting aside any problems regarding derived versus non-derived content, Adams & Aizawa express scepticism that any more coarse-grained, purely functional construal of cognitive abilities will stand up under empirical scrutiny. Hence, on the grounds

that science tends to individuate phenomena by reference to underlying processes, Adams & Aizawa are pessimistic that putative examples of extended cognition will ultimately be classed as genuinely cognitive<sup>7</sup>.

In summary, Adams & Aizawa take themselves to have given two criteria which a phenomenon must meet to be counted as genuinely cognitive. These criteria do not rule out beforehand that there could be any instances of extracranial cognition. The claim is that, given these criteria, no extracranial phenomena alleged by cognitive externalists to be cognitive are genuinely cognitive.

### §2: *Rob Rupert's Cognitive Internalism*<sup>8</sup>

Rob Rupert, in a review of Adams & Aizawa's book, objects to their method of defending cognitive internalism. His objection also demonstrates the starting point of his own contribution to the debate.

Rather than focusing on the distinction between genuinely cognitive portions of the cognitive system and those that are not, we should try to differentiate the theoretically important system, from the standpoint of successful cognitive psychology, from further causal contributors to the production of intelligent behaviour. On a functional approach, identifying the privileged system and its theoretically relevant properties would seem—at a first pass, anyway—to be all there is to identifying cognitive states and processes (2010).

The privileged system, according to Rupert, is the individual (usually human) organism. Hence, the category of objects and entities to which a theory of cognition can appeal is broader on Rupert's account than on Adams & Aizawa's. But Rupert still claims that his account will predict results which resist the cognitive externalist picture. Particularly, Rupert focuses on criticisms of cognitive externalism which challenge its adequacy/superiority as an empirical thesis about cognition, claiming that the adoption of cognitive externalism: a) does

---

<sup>7</sup> This argument has been criticized in Sprevak (2007).

<sup>8</sup> Unless otherwise noted, citations of Rupert are from his 2009 book, *Cognitive Systems and the Extended Mind*.

not bear any empirical fruit that cognitive internalist positions cannot also yield; and b) does undermine many useful results obtained under the cognitive internalist's framework.

What are Rupert's own criteria, which he claims will deliver to us cognitive internalism? As his criticism of Adams & Aizawa suggests, the notion of a cognitive system is prior to that of a cognitive state. But what is the so-called "privileged system", and how are we supposed to identify it? Ultimately, Rupert concludes that the organism (what I call the 'bodily individual') is the privileged system. This claim is the result (and not an assumption) of Rupert's way to determine the scope of a system by measuring the level of integration amongst various factors and resources, both internal and external to the body. The way integration is measured on this account is somewhat complicated. I reproduce here a large excerpt for easy reference.

Thus, for a given subject at a given time, there exists a set of mechanisms, capacities, external resources, etc. each of which has contributed distinctively to that subject's cognitive processing on at least one occasion. For each such type of mechanism, relative to each kind of cognitive phenomenon that it helps to produce, there is a conditional probability (determined by relative frequency, if by nothing else) of its use relative to each of the other mechanisms, abilities, etc. in the set, as well as a conditional probability of its use relative to each subset thereof. These conditional probabilities can be rank ordered. Assume that mechanisms A, B, and C have contributed distinctively to cognitive processing in the subject in question at least once (but not necessarily to the same token process). Now assume that, given the history of the co-contribution of A, B, and C,  $P(C \mid A \& B) = 0.7$ ; then the set  $\{A, B, C\}$  goes on the list as a 0.7 (relative to some particular kind of cognitive outcome) and will likely come in ahead of many other sets (for a particular subject at a particular time). The same set might also appear at a different place in the ordering because, for instance,  $P(A \mid B \& C)$  does not equal 0.7 relative to the same kind of cognitive outcome. Moreover, given the variety of kinds of cognitive outcome, the same sets appear on the list many more times, most likely with many different associated probabilities. Next, consider the likelihood of a natural cutoff between the higher probabilities and lower ones, a gap that separates highly interdependent mechanisms from those that are less so. (In the absence of a significant gap in the ordering, 0.5 would seem to mark the relevant cutoff point.) Now count the number of times each type of mechanism appears on the list of sets with higher conditional probabilities (i.e., those sets above the significant gap on the list); these frequencies themselves exhibit a rank ordering, and a natural cutoff (another



significant gap) separates those mechanisms that appear frequently on the list—that is, are highly interdependent and heavily co-employed—from those that appear rarely. This indicates which mechanisms are parts of the integrated set to be identified with the cognitive system and which are, in contrast, resources used by the cognitive system (42 – 43).

Basically, if some factor shows up frequently enough in groupings of factors which themselves frequently enough produce cognitive outcomes, then that factor is part of an integrated cognitive system. That is, such a factor is part of what underpins a cognitive phenomenon, rather than merely some input to that phenomenon, or merely some condition within which that phenomenon occurs. What counts as a cognitive outcome is not given explicitly, however. Rupert relies on us to already have a grasp upon the sorts of phenomena cognitive science is interested in explaining. Whatever these are, they are the cognitive outcomes with respect to which we measure various factors' contributions.

Additionally, Rupert notes that a condition for “screening off” is necessary. Some factors will score highly which no party to the cognitive internalism-externalism debate will want to count as part of cognitive processes or systems. Rupert's example is the sun, which will figure highly in the outcomes of visual perception. Clearly, however, the sun is not part of any earthly creature's cognitive system. But how do we screen-off such factors in a non *ad hoc* fashion? The suggestion by Rupert is to screen-off any factor which “causally affects only one or a very small number of other mechanisms, and achieves a high integration score simply on account of that effect” (43). Thus, the sun is not allowed to count as part of an integrated cognitive system because it only contributes via its effect on the photoreceptive cells of organisms. Likewise, this is how Rupert's position is supposed to exclude most of the external factors cognitive externalists will want to include as part of cognitive systems or processes: such factors do all their work via narrow effects on organisms' sensory and perceptual apparatus.

### §3: *Cognitive Externalism*

Myriad and sundry are the cognitive externalist positions. However, not all of them are motivated by the same considerations. Indeed some may be in tension with each other

(Clark 2008a). There is neither reason, nor hope, to review all of the ways in which one might be a cognitive externalist here. We will look at those cognitive externalist positions which are most amenable to reconstruction in the sober metaphysical language of ‘realization’ and ‘constitution’. The different cognitive externalist positions we will examine (and most cognitive externalist positions in general) are driven by considerations of the intimate interplay and entwinement of internal and external factors in the production of cognitive phenomena. These considerations are approached in numerous ways. There are positions like that found in (one strand of) Clark & Chalmers (1998) which extend intuitions from functionalism in the philosophy of mind. Another strand in Clark & Chalmers is one which argues for cognitive externalism on the basis of intimate, dynamic coupling of internal and external factors. Somewhat similar to the dynamic coupling thesis are both the enactivist (Noë 2004) and integrationist theses (Menary 2007), which advert to something more than mere coupling—such as active engagement with the environment, or skilled and norm-governed use of the environment (respectively). Finally, there is a more straightforward route—similar to the extended functionalism motivation—which argues directly to cognitive externalism by revising the notion of realization (Wilson 2004). The proponents of all of these criteria take themselves to be able to demonstrate that cognition actually does extend beyond the bodily individual. We will look at these accounts, briefly, in turn.

### §3.1: *Extended Functionalism*

Functionalism is a well-known thesis in philosophy of mind (e.g., Putnam 1975a; Fodor 1968). Briefly, the idea behind the position is that a mental state is characterized by its causal profile, or function. That is, all there is to know about a mental state is what inputs it accepts, what outputs it yields, and the sort of connection it has to other mental states. It is not important whether a function is realized in this-or-that material. Hence, mental states could be had by creatures whose neural architecture radically differs from ours in both constitution and arrangement. Immediately, it should be apparent how this could be friendly ground for the cognitive externalist. If there is nothing special about brains—or even organic material more generally—with respect to be the basis of mental functions, then these functions could be realized outside of brains.

Clark & Chalmers (1998) give two examples to motivate cognitive externalism. In this section, we will focus on their second example, the now famous Otto's notebook case (12 – 13). In the next section—when exploring the dynamic coupling arguments for cognitive externalism—we will focus on the *Tetris* example. Strictly, these two cases motivate distinct conclusions. The case of Otto's notebook is supposed to motivate a cognitive externalist account of *mental states*, and is treated as something distinct from that which is established by the *Tetris* example (8)—*viz.*, cognitive externalism about *cognitive processes*. I will not make much of this difference. The strategies in the various examples can be adapted to different sorts of states and processes if the mental-cognitive divide strikes one as significant and problematic. If we consider the Otto case to be about *memory* extending beyond the bodily individual, this seems to be an argument for cognitive externalism about cognitive processes.

We are asked to imagine that an Alzheimer's sufferer, Otto, being aware of his memory defect, has instituted an intricate system for recording information in a notebook. Therein, Otto writes down facts which he thinks he will want to make use of at a later time, knowing that his biological memory will not be able to retain these facts. In particular, we are told that Otto learns of a new exhibition beginning soon at the local art museum. Upon learning about the exhibit, Otto records in his notebook the address of the museum (among other logistical details concerning the exhibit, I imagine). Subsequently, Otto forms the intention to go to the museum. (Assume that sufficient time has passed so that Otto's biological memory no longer holds any trace concerning the location of the museum.) He consults his notebook in order to find the address and then navigates to the museum.

Clark & Chalmers invite us to ask what we need to appeal to in order to explain Otto's trip to the museum. By assuming the widely held belief-desire model of psychology, Clark & Chalmers take it that Otto's trip to the museum is initiated by his desire to go there, combined with his belief that it is in such-and-such a location (say, 1280 Peachtree St.). We have been given no reason to think that Otto's desires are anywhere other than where we would expect them to be (i.e., somewhere in his brain). Nor have we been given any reason to think that Otto's *occurrent* belief concerning the museum's location is anywhere other than where we would expect beliefs to be (again, somewhere in Otto's brain). Instead, Clark &

Chalmers say that Otto has a *dispositional* belief concerning the museum's location, formed by the act of writing in his notebook when he first looked up where the museum was. This is supposed to be like the dispositional belief that an individual with a normally-functioning biological memory (Clark & Chalmers call her Inga) has when she learns that the museum is at 1280 Peachtree St. and stores it in her biological memory. We are urged to avoid the temptation to distinguish between Otto and Inga according to the following consideration. Inga, when she forms the desire to go to the museum, automatically recalls its location from her biological memory and then proceeds on her journey. Otto, on the other hand, forms the desire to go to the museum and then must think to himself, "I wrote the location of that in my notebook", proceed to look up the location of the museum, and then make his trip. The intermediate step "remember where I recorded this" is otiose, claim Clark & Chalmers, and based on nothing but the question-begging prejudice that factors outside of the brain/body are fundamentally ineligible for inclusion in genuine mental or cognitive processes. After all, we could likewise attribute such an intermediate step in Inga's retrieval, wherein in she thinks to herself, "I stashed the location of the museum in my biological memory". In fact, one might think that this is actually what happens. Or at least, the same sorts of problems we can envision with Otto's putative notebook-memory could easily have analogues in the case of biological memory. Would the possibility of such things cause us to doubt that dispositional beliefs in biological memory are not part of an agent's mind? Clark & Chalmers think not.

Thus, given that in both the Otto and Inga cases we need an occurrent belief concerning the location of the museum to pair with the desire to go there—and given that we further need a dispositional belief to explain the occurrent one—we are asked to think where the relevant information resides in each case. For Inga, the answer is obvious and (relatively) uninteresting. In Otto's case, however, the claim is that the relevant information is nowhere but in the notebook. Subtracting the recorded location of the museum in the notebook, we are left either with an unexplained trip to the museum by Otto, or with a failed trip. But both unexplained behaviour and precipitous declines in (counterfactual) behavioural competence are undesirable results, which results we can avoid if only we allow Otto's notebook to be part of his mind. To link this back into the functionalist framework

mentioned at the beginning of the section, the information in Otto's notebook can serve as the filler of an important causal role in the belief-desire theoretical explanation of Otto's museum trip. On the assumption that the notebook is a constant fixture in Otto's life, this situation will be repeated over again many times. That is, the role-filling capacity of the notebook will be systematically important in explanation of Otto's behaviour. Thus, Clark & Chalmers conclude that we ought to consider Otto's notebook as partly constitutive of some of his mental states. But then, Otto's mind is partly constituted by factors external to his body. Hence, cognitive externalism is true.

Clark & Chalmers stress that this sort of externalism is distinct from the more familiar thesis of content externalism. On a content externalist account, external factors (e.g., the use facts of some concept in a linguistic community) partly determine the content of thoughts. Thus, given a pair of subjects who are physical duplicates of each other, if one is embedded in a community where 'arthritis' denotes a disease of the joints and the other is embedded in a community where 'arthritis' denotes a disease of the joints *and* bones, the two think different thoughts when they think something which they would formulate as "That person has arthritis" (Burge 1979). Alternatively, take the same duplicate pair and imagine that one is in an environment in which there are trees and the other is in an environment in which there are no trees. The former can have tree-thoughts and the latter cannot (Putnam 1975b).

However, Clark & Chalmers note that the external factors appealed to in this sort of case do their work passively and historically. The subject in the community which uses 'arthritis' to denote a joint ailment thinks about joint ailments when she has thoughts of the form "That person has arthritis" merely in virtue of belonging to that community. There is no external factor with which she must be interacting at a particular time in order to have an arthritis-thought at that time. However, for the externalism that Clark & Chalmers are advocating, a more active engagement with external resources is in the offing. Hence, content externalist style theses are given the label "passive externalism", while the cognitive externalism which Clark & Chalmers argue for is given the label "active externalism". The difference between the two is best explained by looking at the second strand of argument Clark & Chalmers use to motivate cognitive externalism. Let us move on to that.

### §3.2: *Dynamic Coupling*

The second strand of argument in Clark & Chalmers, one from dynamic coupling, can be used either independently from the extended functionalist intuitions, or in concert with them. Numerous authors deploy this argument, or something close to it. Nevertheless, I will stick primarily to talking about the argument as it is found in Clark & Chalmers, noting when it becomes necessary different developments of the argument.

In the same paper where we find the Otto's notebook example, we also find the *Tetris* example. For those not into video games from the early 1990's, *Tetris* is a puzzle game. Blocks (or, "zoids") made up of various arrangements of four cells move from the top of a screen downward, accumulating at the bottom of the play area. The object for the player is to rotate and place these blocks so that rows of the play area are completely filled with cells, at which point the cells will be cleared from the play area, giving the player more room to continue. If the any cell from a block cannot be fit into the play area, the player loses and must start again.

Clark & Chalmers describe studies (Kirsch & Maligo 1994) using *Tetris* as a performance task. Under one condition, participants mentally rotate (i.e., rotate them in imagination) blocks to find the desired orientation before performing the rotation on the screen. In the second condition, subjects can perform the rotation of the screen at will to find the desired orientation. The interesting result is that players using on-screen rotation to find optimal orientations perform significantly better than those using only mental rotation. Nor is this a passive, one-way information transfer. While the subject does not perform the rotation of the blocks mentally before doing so on-screen, she does still need to take in and process the information from on-screen rotations. This information will then be evaluated to determine whether further rotations are necessary, or whether the block can be placed. Clark & Chalmers (as well as Clark 2008b) take this to demonstrate that cognitive effort can be distributed between internal and external resources in an active, reciprocal manner. That is, the information values of elements of the video game and elements in the visual and spatial processes of the agent are closely linked and sensitive to changes across short intervals of

time. Such situations give rise to dynamic coupling (or, “continuous reciprocal causation” (Clark 1997; 2008b). Clark alludes to a more mathematics-inspired formulation.

Two components are treated as a coupled system in a specific technical sense; that is, the equation describing the evolution of each component contains a term that factors in the other system’s current state (technically, the state variables of the first system are also the parameters of the second, and vice versa) (2008b: 24)<sup>9</sup>.

The existence of such a relationship between elements internal to the body and external to the body is supposed to undermine the grounds for thinking that the external resource is merely input for the internal resource<sup>10</sup>.

This can be used to augment the extended functionalist picture outlined by the Otto’s notebook example. For instance, Clark & Chalmers suggest that the existence of a dynamic coupling relationship between an internal resource and an external one is a prerequisite for the external resource to be eligible to as the filler of the causal role of a mental state. These conditions are stated in Clark & Chalmers (17), and frequently restated and refined by Clark. Here is a recent formulation.

- 1) That the resource be reliably available and typically invoked
- 2) That any information retrieved be more-or-less reliably endorsed. It should usually be subject to critical scrutiny . . . It should be deemed about as trustworthy as [the deliverance of a biological counterpart resource]
- 3) That the information in the resource should be easily accessible as and when required (Clark 2010: 46).

Meeting these conditions ensures that the contribution of the resource in question is relatively stable and systematic. The conditions are developed to answer reasonable concerns

---

<sup>9</sup> Even more technical, here is Adams & Aizawa (2008). “[If] the rates of change of variables  $x$  and  $y$  with respect to time are given by  $dx/dt = x + y$  and  $dy/dt = y + 2x$ , then the variables are coupled” (108).

<sup>10</sup> One might claim, more radically, that these sorts of considerations undermine the distinction between internal and external portions of cognitive/mental processes. Timothy Williamson (2000) seems to have something like this in mind when he argues for the non-decomposability of mental states into internal and external portions. We will stick with the less radical reading here.

about the theory allowing for cognitive processes to extend absurdly far outside the bodily individual, or concerns that one-off occurrences will lead to fleeting extensions of cognition. Both of these concerns would plausibly raise difficulties for the argument that a given external resource could fill the causal role it is meant to. These conditions also explain the difference between passive and active externalism mentioned in the previous section. While there is an historical element to the conditions, they are meant to determine when an *instance* of an individual exploiting an external resource counts as an extension of a cognitive process. Recall also that this is not supposed to be a one-way affair: the external resource must affect the individual's behaviour and be affected by it. In the *Tetris* case, this was exemplified by interplay between on-screen rotation and evaluation of block positioning. Only when all of this is the case does cognition extend beyond the individual. However, these conditions do not obtain between an individual and the objects in her environment which determine mental content. A subject need not be in a situation of continuous reciprocal causation with *any particular* external object meeting the three conditions above in order to have thoughts with a given content. It is not a particular tree which a subject interacts with that allows her to entertain tree-thoughts, but the (existence of) trees in her environment generally. However, it would be absurd to claim that every tree is part of a subject's every tree-thought.

One need not be committed to functionalism to avail one's self of this sort of argument for cognitive externalism. It should be clear that this is the case, so I will not argue for it. In any case, the dynamical systems and coupling arguments for cognitive externalism are widely applied, though not always in the context of functionalist assumptions (which are explicitly invoked in Clark & Chalmers). The details of particular appeals<sup>11</sup> are not important here, but the gist is the same as what we get from Clark and Clark & Chalmers: if elements are in intimate enough interaction with each other, then their location internal to or external to the body is irrelevant.

There are two distinct intuitions evinced in these sorts of arguments. The first is a subtraction intuition. It is supposed to be apparent that, in the cases where external factors

---

<sup>11</sup> See van Gelder (1995) for the paradigmatic dynamical systems argument to cognitive externalism. For a slightly different coupling argument, see Hagueland (1998).



are in a dynamic coupling relationship with internal factors, subtracting the external element would drastically affect the cognitive performance of a relevant subject. Take away Otto's notebook, and he cannot get to the museum (and will presumably fail in numerous similar navigational tasks, at least). Or, in the *Tetris* case, when the on-screen information is treated merely as input for mental processes of rotation, finding the correct orientation is much slower and more effortful. The second intuition is one about explanations. If we ignore or downgrade the contribution of internal factors, the cognitive behaviour we observe will be at best poorly explained and at worst mysterious. Adverting to external factors as parts of the cognitive processes or systems themselves (as opposed to merely background conditions, input or whatever) gives us a fuller and richer explanatory picture. So the story goes.

### §3.3: *Alternative Developments of the Coupling Intuition*

Other sorts of theory take a slightly different tack to the end of fleshing-out how the internal and external are closely related in producing cognitive phenomena. Alva Noë (2004; 2009) develops what he calls an “enactive” approach to perceptual cognitive activity<sup>12</sup>. The label is meant to emphasize that perception is not a passive phenomenon, but an active one. Here is a quote emblematic of the enactive approach.

If perception is in part constituted by our *possession and exercise of bodily skills*—as I argue in this book—then it may also depend on our possession of the sort of bodies that can encompass those skills [emphasis added] (25).

The possession of bodily skills is constituted by knowledge of what Noë calls “sensorimotor contingencies” or “sensorimotor dependencies”; that is, knowledge of how sensory input will change in response to certain movements or manipulations. Hence, visually perceiving a mug is partly constituted by a subject knowing how the mug's appearance would change if she should move to a different viewing angle or move the mug in some fashion (e.g., rotating it). Not only is perception partly constituted by the knowledge which comprises these skills, but it is also partly constituted by the exercise of those skills. Perception is partly the actions a perceiver undertakes, such as saccading (moving the eyes quickly across a

scene), tilting her head, rotating an object, etc. Hence, perception includes not only parts of the body outside the brain, but parts of the world outside of the body. An agent's manipulations of a mug—or her movements oriented towards the mug—are constitutive of her perceiving the mug.

The enactivist theory develops the intuition that internal and external factors are intimately related by claiming that cognition is essentially active, involving the exercise of skills in the environment. Cognisers are seen to be essentially embodied and embedded. Enactivism appears to embrace neither the functionalism which motivates Clark & Chalmers, nor the specific dynamical systems route to cognitive externalism. The account found in Noë argues directly for the constitutive dependence of cognition on factors external to the body.

Another alternative is found in Richard Menary (2007). Menary's so-called "integrationism" takes the dynamic coupling mentioned in Clark & Chalmers as merely the starting point of a cognitive externalist account. Moreover, Menary takes it that these manipulations are not merely causal, but also normative. Specifically, an important set of manipulations—*viz.*, the manipulations of external representations—is norm-guided.

There are right and wrong ways to manipulate an external representation to reach a desired end and we learn how to manipulate the representations correctly. An account of cognitive norms has not sufficiently been developed, because the initial focus of [cognitive externalists] has been on the causal and constitutive aspects of manipulation (138).

Menary takes manipulations of external representations to be crucial to paradigmatic, high-level cognitive tasks like logical and mathematical reasoning. Nor are these external representations supposed to be seen merely as off-loading cognitive effort from internal to external resources. Learning to manipulate these external resources transforms the set of cognitive tasks of which an agent is capable, claims Menary; what subjects can do with only internal resources and what they can do after learning to manipulate external resources is

---

<sup>12</sup> For more on the enactivist approach see, e.g., Rowlands 2009 and Thompson 2007.

different as a matter of kind, rather than merely a matter of degree<sup>13</sup>. Further, the contribution of internal and external resources is supposed to be symmetrical, with neither being more important than the other<sup>14</sup>.

Certainly, justice has not been done to either the enactivist or the integrationist approaches to cognitive externalism. No doubt, my characterization of these positions as mere alternative developments to the same motivations driving extended functionalist and dynamic coupling arguments would be found objectionable (especially by Menary). However, it is sufficiently accurate for our purposes to characterize both enactivism and integrationism as fleshing-out the intuition that internal and external factors are related to each other in a sufficiently intimate way (with respect to the production of cognitive phenomena) such that they should be treated as constituents of cognition, rather than merely causal factors.

#### §3.4: *Wide Realization*

Finally, we look at an argument which operates directly via the concept of realisation. Rob Wilson (2004)<sup>15</sup> gives a critique of the traditional account of realisation, and then puts forth his own notion which he claims allows for cognitive externalism. The traditional account compromises two theses—the constitutivity thesis and the sufficiency thesis<sup>16</sup>.

*(Sufficiency)* Realizers are metaphysically sufficient for the properties or states they realize (*op cit* 103).

*(Constitutivity)* Realizers of states and properties are exhaustively physically constituted by the intrinsic, physical states of the individual whose states or properties they are (*op cit* 104).

---

<sup>13</sup> See Dennett (1996) for a similar claim.

<sup>14</sup> Menary takes this to distinguish his view from other views, such as those found in Clark & Chalmers (1998) and Clark (2008b) which seem to give more importance to internal resources and off-loading of work from those resources to external ones.

<sup>15</sup> All references to Wilson are from his 2004 book, *Boundaries of the Mind: The Individual in the Fragile Sciences*.

<sup>16</sup> Wilson attributes views on realization characterized by these two theses to the likes of Jaegwon Kim 1993, among others.

The sufficiency thesis holds that the presence of the realiser for a state or property is all that need obtain for the relevant state or property to obtain. Thus, if the realiser of a toaster is a certain arrangement of electrically-conductive wires and coils, housed in a shell along with a spring-loaded shelf, then having all these elements so arranged will be all there is to having a toaster. More to the point, if there were a world in which there were the aforementioned elements—properly arranged—*and nothing else*, there would still be a toaster. A (metaphysically) lonely toaster is still a toaster.

The constitutivity thesis claims that nothing outside of the individual to whom we are predicating some state or property is part of the realiser of that state or property. Think about Fred's burned toast. If we want to know what the realiser of the property *being burned* is, when predicated of the piece of toast on Fred's plate, then we should look no further than the physical stuff which composes the offending toast. No facts about Fred, the plate upon which the toast rests, weather conditions on Mt. Washington, or the linguistic use of 'burned' in Fred's community (etc.) have anything to do with whether the relevant piece of toast is burned.

So far, so good. However, Wilson notes that the traditional account deploys another distinction as well between core and total realizations.

*(Core Realizer)* A state of the specific part of [the subject] that is readily identifiable as playing a crucial causal role in the producing or sustaining [the relevant state or property]

*(Total Realizer)* A state of [the subject], containing any given core realization as a proper part, that is metaphysically sufficient for [the relevant state or property] (108)

Returning to our toaster, the core realiser of being a toaster is that which is responsible for toasting bread—*viz.*, the arrangement of electrically conductive wires and coils. As for the total realiser, it is that which suffices for the existence of a toaster; which we saw earlier were the core realisers (i.e., wires and coils), together with the shell and spring-loaded shelf.

With all these notions in hand, Wilson makes an argument to the effect that the combination of the constitutivity and sufficiency theses cannot cover all cases of realization.

According to Wilson, one can either stipulate the constitutivity thesis or find cases of properties the realizations of which violate the sufficiency thesis (Chp. 5 §7); or stipulate the sufficiency thesis and find cases of properties the realizations of which violate the constitutivity thesis (Chp. 5, §8). It is the latter category which is of interest to the cognitive externalist. Properties which have realisers violating the constitutivity thesis are ones whose realizations are partly *outside* of the individual to whom the property in question is attributed. Wilson dubs these “wide realizations”. He notes that there is a further distinction within this category as well, between those realisations which are wide and those which are what he calls “radically wide realizations”.

*(Wide Realization)* A total realization [of some property or state] whose noncore part is not located entirely within [the individual] who has [the state or property] (112)

*(Radically Wide Realization)* A wide realization whose core part is not located entirely within [the individual] who has [the property or state] (116)

Wilson points to the familiar content externalist cases to exemplify wide realisation. In the case of propositional attitudes (e.g., water-thoughts), water itself is part of the total realiser of a water-thought, according to Wilson. A subject’s (Fred) water-thought has as a core realisation what goes on neurologically with Fred. However, this is not sufficient for Fred’s thought to be a water-thought: that he is related to actual water (in some way) must also be the case.

It seems that for Wilson a cognitive state with a wide realisation would be sufficient to demonstrate cognitive externalism. However, if content externalist states are the examples of this, we should doubt Wilson. Recall that Clark & Chalmers introduced the distinction between active and passive externalism. The external factors adverted to in content externalist explanations of mental content are not properly related to individuals to demonstrate cognitive externalism. We do not want a theory which suggests that all water is part of every water-thought Fred has.

Nevertheless, we can still use Wilson's apparatus to delineate a situation consonant with cognitive externalism construed as active externalism. In an instance of radically wide realisation, the core part of the realization of some state is partially outside the individual to whom the state is attributed. The core part of a realisation is the part which plays a "crucial causal role in producing or sustaining" (108) some property. Hence, a case of radically wide realisation would be one in which the factors responsible for fulfilling the crucial casual role are not entirely within the individual to whom the realized state or property is attributed. This is just what is supposed to occur in the Otto's notebook, for example. The information in Otto's notebook is supposed to do some of the crucial causal work assigned to a belief (i.e., represent the state of the world). Obviously, the pages of the notebook and the information they hold are outside of Otto, the individual to whom we are attributing a belief about the location of a museum. This is thus a case of radically wide realisation according to Wilson, and a case of cognitive externalism construed as active externalism according to Clark & Chalmers. Thinking of cognitive externalism in these combined terms allows us to see what sort of metaphysical picture might underlie it, and further allude to some of its motivations and explanatory benefits (i.e., the benefits and motivations we've discussed in the preceding sections).

### §3.5: *Summary*

The various positions which have been briefly described here each attempt to give form to the intuition that the obtaining of certain sorts of relations among factors internal to the body and factors external to the body engender states-of-affairs in which cognitive states or processes are partly constituted by both sorts of factor. In particular, many cognitive externalist accounts rely on illuminating the close and mutual interaction and dependence of internal and external factors with regard to the production of cognitive phenomena. Whatever the going account of this interaction or dependence is, cognitive externalists all aim to establish that the constituents of cognition are not found only within the body and brain.

*§4: Confirmation Criterion*

I take it that the dialectic structure of the debate between cognitive internalism and cognitive externalism justifies considering the former as the default position and the latter as a challenger to the orthodoxy in philosophy of cognitive science<sup>17</sup>. As such, cognitive externalism bears the burden of proof. But what would the cognitive externalist have to do demonstrate to discharge that burden? That is, what are the confirmation criteria for cognitive externalism?

Both Adams & Aizawa's and Rupert's positions maintain that the best account of cognitive phenomena is in terms of states internal to individual subjects, which subjects are delineated along the boundaries of the body. These two positions differ with respect to which internal factors are relevant to cognition, but agree that only internal factors are relevant. As we have seen, neither position is concerned with arguing against the possibility of confirming instances of cognitive externalism. It is agreed by cognitive internalists that theirs is an empirical hypothesis which allows even for the epistemic possibility of cognitive externalism. However, the cognitive internalist's claim is that, as a matter of fact, there are no confirming instances of cognitive externalism. Insofar as any cognitive externalist theories are deemed coherent, they are not taken to give an adequate or superior explanation of any observed cognitive phenomena. As such, cognitive externalism would stand confirmed under the following criteria.

*(Cognitive Externalism Confirmation Condition)* There is at least one actual instance of a cognitive phenomenon the constitutive machinery of which:

a) Extends beyond the brain (and nervous system) of an individual

And

---

<sup>17</sup> This claim is controversial. All I mean to say is that cognitive internalism is historically the predominant picture of cognition, and one which has apparently led to significant results during its reign. I take it that this is enough to sustain the claim concerning the dialectic of the argument between cognitive internalists and externalists.

b) Extends at least beyond the non-neural portions of an individual which directly subserve, or interface with, an individual's brain (and nervous system)<sup>18</sup>

Condition (a) is provided by Adams & Aizawa's theory. Given that they count as relevantly internal only what is intracranial (where this has the non-standard meaning I defined earlier), no cognitive phenomenon will count towards the confirmation of cognitive externalism unless it has constitutive machinery outside the brain and central nervous system of a given individual. Condition (b) is provided by Rupert's theory. While Rupert's theory is nominally organism-centric, it is clear that he would be opposed to considering the activity of muscles (for instance) cognitive (2009: 44). Thus, a close reading of Rupert indicates that a confirming instance of cognitive externalism is one in which the constitutive machinery of some cognitive phenomenon extends *at least* beyond certain portions of the individual, but more likely beyond the individual herself<sup>19</sup>. Hence, the formulation of condition (b) given is meant to do justice to the close reading of Rupert's theory. Rupert claims that his account of integration will yield (more or less) the same results found by orthodox cognitive psychology. These results do not include extra-organismic factors *or* certain intra-organismic factors (again, like muscles).

The weakest rejection of cognitive internalism is quite weak indeed. However, even if no extant cognitive externalist theory actually made its case based on the weakest rejection of cognitive internalism, we would still want to focus on the weak rejection. In order to illuminate any connections among various externalist positions within philosophical sub-

---

<sup>18</sup> Strictly speaking, both Adams & Aizawa's and Rupert's theories can accommodate actual instances of cognition which are partly constituted by external factors. What they will insist on denying is that there are sufficient (or sufficiently related) instances for any interesting generalizations to be made about them. I am grateful to Mark Sprevak for pointing this out to me. I omit this qualification for the sake of simplicity in my conditions. However, nothing in my subsequent arguments will hang on this omission.

<sup>19</sup> It is clear that any cognitive phenomenon the constitutive machinery of which outstrips the boundaries of the individual will also be a phenomenon that outstrips the boundaries of certain physical subsystems of that individual. However, it is also clear that a phenomenon which outstrips the boundaries of certain physical subsystems of an individual need not outstrip the boundaries of the whole individual.



disciplines, we must properly label all theories "externalist" which merit the distinction. But the only way to do that is to formulate the weakest, most noncommittal versions of these theories. We have done this for externalism in the philosophy of cognitive science. Now we move on to epistemology.

## Chapter 2

### Internalism and Externalism in Epistemology

#### *Introduction*

Now I turn to externalism-internalism in epistemology. As in the previous chapter on cognitive externalism-internalism, I will proceed as follows. First, I will discuss what ‘internal’ and ‘external’ mean in the context of epistemology. We will see that this is a matter of some dispute amongst epistemologists, with two primary accounts of what “internal” means in epistemology on offer. Both of these positions—accessibilism and mentalism—will be examined; but I will not come down in favour of one or the other here.

Characterising epistemic internalism will put us in a position to give an outline of epistemic externalism as a negative thesis—but that will not yet be sufficient for my purpose. Hence, I will also outline epistemic externalism’s positive thesis (actually, only one version of many possible positive epistemic externalist theses). The basic germ of the positive thesis is reliabilism. Of course, basic reliabilism faces well-known problems. I will trace the proposed fixes for these problems to the more sophisticated, contemporary versions of reliabilism. The most important aspect of these improved reliabilist theories for my purposes will be the invocation of cognitive abilities.

Having laid out epistemic internalism and epistemic externalism, I will be in a position to formulate confirmation criteria for epistemic externalism, just as I did for cognitive externalism in the last chapter. This will complete the groundwork for the main arguments of this work.

### *§1: Epistemic Internalism*

In the literature, there are two primary accounts of what ‘internal’ means in epistemology. We can construe these two accounts as providing different answers to the question, “What factors are relevantly internal in epistemic theorizing?” On its own, this question is underspecified. After all, there are numerous different phenomena—or, “epistemic statuses” as they shall henceforth be called—towards which an epistemologist might take an internalist stance (e.g., justification, knowledge, understanding, well-groundedness or rationality). Here, we shall only be concerned with justification.

The two accounts we will discuss—accessibilism and mentalism<sup>20</sup>—both attempt to formulate what I call epistemic “ur-internalism” about epistemic justification (henceforth, simply “ur-internalism”). That is, each account aims to articulate the basic and unelaborated principles which any theorist of justification must endorse for her position to be an internalist one. A delineation of the basic core of epistemic internalism will also furnish us with what the epistemic externalist is obliged to reject if her position is to be an externalist one.

Despite the differences we will examine, there is much in common between the versions of ur-internalism. Both rule out reliabilist accounts of justification and more generally any theory which allows contingent environmental and social factors to possess any justificatory potency. Further, each allows a distinction between propositional justification and doxastic justification<sup>21</sup>. We will have little need to be concerned with this last distinction, but I mention it to clear out any potential noise in my comparison of accessibilism to mentalism, which I want to revolve solely around what these positions respectively consider

---

<sup>20</sup> Both ‘accessibilism’ and ‘mentalism’ are coined in Conee & Feldman (2004).

to be the relevantly internal factors in justification. In what follows, whenever I talk about a subject's justified beliefs, I intend to cover only those beliefs which the subject actually holds (i.e., those which are may or may not be doxastically justified).

Before exploring the two versions of ur-internalism, the general animus of internalism about justification should be made clear. The thought that motivates internalists about justification is the idea that all of the facts relevant to justification should be *within* an agent's perspective (BonJour 1992; Conee & Feldman 2004; Vahid 2005; Goldberg 2007). Hamid Vahid briefly characterizes this motivation.

The perspectival dimension is characterized in terms of the phenomenology of the cogniser's epistemic situation, how things appear from his or her perspective and the epistemically relevant factors available to the agent . . . equally relevant to the perspectival dimension of justification is how responsibly an agent behaves vis-à-vis the evidence available to her when forming a belief, that is, whether her beliefs are based on the total evidence she possesses (2011: 145).

As Vahid further notes, 'availability' naturally lends itself to some notion of cognitive access. This is clearly what an accessibilist ur-internalism would use as its taking-off point. However, Vahid (2005) also says, "As noted earlier, what [the access constraint] highlights is the availability of reasons for a belief to the agent holding that belief . . . The perspective requirement says nothing about the "mode" of such access" (10). Vahid's point is that an access requirement need have nothing to do with reflection—this is an extra imposed by the accessibilist ur-internalism. The gathering of environmental and behavioural data is also access, after a fashion<sup>22</sup>. For my purposes, however, I will treat access as the internalist does.

Of course, "within an agent's perspective" might equally be taken to mean simply a part of that agent's mental life. Here, 'mental life' and 'perspective' are treated as equivalent terms. And of course, the parts of an agent's mental life are just the mental phenomena

---

<sup>21</sup> A fact provides propositional justification if it justifies a possible [as in, non-actual] belief for a subject. A fact provides doxastic justification, on the other hand, if it justifies an actual belief for a subject.

<sup>22</sup> See also John Gibbons' (2006) "Access Externalism".

which belong to her<sup>23</sup>. This is obviously the taking-off point of a mentalist ur-internalism. It is important to note that it is specifically, and only, the relevant subject's mental life which is at issue for the internalist of this sort. The mental lives of other subjects may vary freely without affecting the justificatory status of the subject under consideration, according to the mentalist.

### §1.2: *Accessibilism*

Accessibilism, as the name suggests, holds that internalist theories of justification are necessarily characterised by some sort of access requirements. Jim Pryor gives an accessibilist ur-internalism which he calls “simple internalism”.

*Simple Internalism* (SI) Whether one is justified in believing P supervenes on facts which one is in a position to know by reflection alone (2001: 104).

This principle claims that if a fact is to be able to justify a belief (possible or actual) for some subject, she must be able—in principle—to be aware of that fact (but not necessarily the further fact that the first fact does justificatory work) by reflection alone. That is, the agent must be able to become aware of justifying facts without further recourse to behavioural or environmental data. Hence, for Fred to justifiably believe that the toast is burning it must be the case that Fred is able to reflectively grasp some factor which does in fact justify his belief. In the case of burning toast, Fred must be able to access by reflection his olfactory sensations of carbonizing bread, his visual sensations of smoke rising from the toaster, his belief that the toaster was on a high setting when he put the bread in many minutes ago, or something similar. Again, Fred need not also be able to reflectively determine that these facts justify his belief—it is enough that the justification relationship holds between the facts to which Fred has reflective access and the relevant proposition.

Pryor takes simple internalism to be “the most minimal internalist position” (*ibid*). Hence, I interpret simple internalism as the formulation of an accessibilist ur-internalism.

---

<sup>23</sup> This is in contrast to accounts like Goldberg (2007), which claim that the cognitive/mental lives of *other agents* may be relevant to a given agent's epistemic status.

There are two general ways to strengthen the basic accessibilist account. Either the accessibilist can place stronger requirements on the nature of subjects' access to the facts justifying their beliefs, or she can introduce requirements to the effect that subjects must have reflective access to their justificatory status as well. Adding the latter sort of requirement yields what Pryor calls "access internalism".

*(Access Internalism)* One always has 'special access' to one's justificatory status (2004: 105).

Some take access internalism to be the minimal commitment of the epistemic internalist. For instance, Vahid (2011) says that eschewing the "higher-order" requirement of access internalism leaves any resulting theory "too weak to be regarded as a species of internalism" (148). We can see such a position as treating 'being justified' on a par with 'having reasons', or perhaps 'having a well-grounded belief'. This would be to seize upon the importance of being able to defend and responsibly hold beliefs, which we briefly noted earlier as one of the general background motivations for epistemic internalism. In short, insisting upon access internalism as the minimal commitment of the epistemic internalist is to locate a deontological thesis (i.e., one concerned with prescriptions and proscriptions about what agents should do, epistemically speaking) at the heart of epistemic internalism (Alston 1986; Goldman 1999; Plantinga 1993). However, there may be compelling reasons to separate justification, reasons, and good grounds from each other. First, the notion that a deontological motivation is at the heart of internalism has been frequently criticised (e.g., Conee & Feldman 2004). Second, even if there is a sense of 'justification' in epistemology which is deontological—and even if this is the primary sense of 'justification' in epistemology—it seems plausible that there is another sense of 'justification'—a theoretically interesting one—which answers to Pryor's simple internalism<sup>24</sup>. In any event, I will assume that this is the case.

---

<sup>24</sup> Earl Conee and Richard Feldman independently argue that some sense of 'justification' can be isolated from both 'reasons' and 'good grounds'. See Conee (in Conee & Feldman 2004: 50) and Feldman (2005: 175).

There are a variety of ways that the accessibilist might strengthen SI without requiring that subjects have any sort of access to their justificatory status. The accessibilist might require that subjects *actually* reflect upon the facts which justify their beliefs, rather than merely requiring that they be able to do so *in principle*. Alternatively, she might specify how many of the facts which justify subjects' beliefs the subjects must reflect upon. She might do both of these things. Hence, the strongest iteration of SI would require that subjects must, in order for some belief to be justified, actually reflect upon every fact which justifies the relevant belief. On the opposite end of the spectrum, the weakest version of SI would require only that subjects must, in order for some belief to be justified, be able in principle to reflect upon some fact which justifies the relevant belief (Brown 2007: 16; Fumerton 2007: 36). There are numerous variations in between the two ends of the spectrum. In short, on an accessibilist ur-internalism, 'internal' means (in principle) open to access by reflection.

Our purposes require that we isolate the weakest position epistemological externalism is obliged to reject. Taking accessibilism as ur-internalism, the externalist must reject that reflective access to the factors which justify beliefs is a necessary condition upon the justification of beliefs. The weakest manifestation of accessibilism requires only that a subject, in order to be justified in having a belief, must in principle be able to access (by reflection alone) at least one fact which justifies that belief (though she need not actually perform any such reflection). An externalist must deny this. However, she is not thereby committed to denying that subjects are ever in position to reflectively access the justifiers of their beliefs. She does not even deny that subjects generally will be in a position to have, in principle, reflective access to the justifiers of their beliefs. Instead, the externalist denies this weakest manifestation of accessibilism because she takes there to be at least one possible case of a justified belief wherein the subject whose belief is in question cannot, even in principle, access by reflection alone any of the facts which justify that belief.

### §1.3: *Mentalism*

The alternative version of ur-internalism is called "mentalism". Whereas accessibilism identifies an access relation between a subject and some facts as what is relevantly internal about justification, mentalism claims that *mental states* are the relevant internal factors. Earl

Conee and Richard Feldman have given an extensive development and defence of internalism as mentalism<sup>25</sup>. They characterize mentalism with the following two theses.

(*S*) The justificatory status of a person's doxastic attitudes strongly supervenes<sup>26</sup> on the person's occurrent and dispositional mental states, events and conditions.

(*M*) If any two possible individuals are exactly alike mentally, then they are alike justificatorially, e.g., the same beliefs are justified for them to the same extent (Conee & Feldman 2004: 56)

In contrast to accessibilism, mentalism makes no claims about the relevant subject's ability to access the facts which justify her beliefs (potential or actual)—not even in principle<sup>27</sup>. In

---

<sup>25</sup> For a similar position, see Wedgewood (2002). We will not discuss Wedgewood's view in this work. I take it to be sufficiently similar to Conee & Feldman's position such that any arguments I make concerning mentalism apply, *mutatis mutandis*, to Wedgewood's position. See Wedgewood (2010: 230) for a characterisation of epistemic internalism by Wedgewood which I take to support my parity claim.

<sup>26</sup> In a footnote, Conee (2007, 55) claims that a weak supervenience condition will fail to exclude reliabilist elements and that such a result is anathema to any internalism worth the designation. Specifically, he means that such a condition will allow two subjects ( $S_1$  and  $S_2$ ) in distinct possible worlds ( $w_1$  and  $w_2$ ) to differ with respect to the propositions which they are justified in believing *without also* differing with respect to their mental states. (The same consideration would apply even if we chose to talk about facts of reflective access rather than mental states.) If  $S_1$  in  $w_1$  has the same mental states as  $S_2$  in  $w_2$ , but not the same justificatory status, then it is possible that there is room for external variations to make a justificatory difference. But Conee claims that the internalist should not brook even the possibility of justificatory potency for external variations. Hence, weak supervenience claims are not sufficient for the epistemic internalist, by Conee's lights.

<sup>27</sup> It is difficult to pin down a single interpretation of Conee and/or Feldman on this point. For instance, sometimes we get statements like the following.

But the internalism defended here does not imply that justified beliefs are backed by [reasons which a subject takes to be supportive reasons]. All that is required is *cognitive access* to evidence which in fact epistemically supports the belief [emphasis added] (Conee, in Conee & Feldman 2004: 50)

This looks like an exact reproduction of Pryor's simple internalism. However, we also get statements which appear to pull in another direction.

The fundamental internalist idea is that justification is internal to us. It seems extra to imply that knowledge of our innards have any particular psychology. It also seems extra to imply that knowledge of our innards is definitely available to us. It seems that an internalist could coherently deny these things (Conee 2007: 56).

This second statement is aimed explicitly at Pryor's simple internalism. Both statements are about justification. Further, both statements are about what is *minimally* required for justification. However, the former clearly promotes an accessibilist ur-internalism whereas the latter seems to repudiate just



fact, the absence of such a requirement is one of the vaunted virtues of mentalism<sup>28</sup>. Hence, the notion of what is relevantly internal is not the same as it is for the accessibilist. If being amenable to reflection does not mark what is internal about justification, then what is internal about it? According to Conee & Feldman, the answer is simple and intuitive. Justification supervenes on the mental, and the mental is internal.

Internalism, on this view, is a thesis holding that certain epistemic matters supervene on certain internal states. The internal states in question are experiences, apparent memories, other beliefs, and the like. In general, they are mental states of the believer. Internalism implies that necessarily if two people are internally alike, then they are epistemically alike (Feldman 2004: 144).

Here, ‘internal’ is meant in the sense used in the content externalism-internalism debate: “The relevant notion of what is internal is primarily spatial—the interiors of two people are the same in their intrinsic qualities” (Conee & Feldman 2004: 82). Of course, not just any state which counts as internal on this criterion will be relevant to justification. The idea is that among the phenomena exhaustively constituted by factors internal to individuals are mental phenomena. Nor do Conee & Feldman wish to delineate a specific subcategory of mental states as the providers of justification—that much is clear in (S). Simply put, they take justification to supervene on the mental, broadly construed. Importantly, eschewing any access requirements as necessary suggests that the mere presence of certain mental states can provide justification for subjects<sup>29</sup>. Hence, Fred’s belief that the toast is burning is justified in

---

such an account. One possible way to make these statements coherent is to recast mentalism as a repudiation of access internalism rather than accessibilism. However, this would seem to involve accusing Conee of grossly misunderstanding Pryor’s simple internalism. On the assumptions that Conee does not misunderstand Pryor—and that Conee & Feldman’s mentalism is supposed to be a thesis both distinct from and superior to an accessibilist ur-internalism—I will assume that the direction indicated by the latter statement is the one we should follow with respect to interpreting Conee & Feldman.

<sup>28</sup> See the second quotation in footnote 27.

<sup>29</sup> I infer this conclusion from the fact that mentalism is supposed to be distinct from accessibilism. Consider this quote from Feldman.

One possibility for a *strengthened* account of internalism turns on the idea that there are internal (or mental states) to which believers lack access and that it is contrary to the spirit of internalism to allow epistemic status to supervene on such states [emphasis added] (150).

virtue of various relevant sensory and perceptual states, beliefs etc., even when Fred has not reflected (or cannot reflect) upon these facts. Justification depends on only mental factors, according to the basic formulation of mentalism.

Note that mentalist ur-internalism does not exclude theories of justification which have access requirements from counting as internalist theses. Instead, it claims that a theory may be of the internalist variety without having such a restriction, so long as it takes mental phenomena to be those which confer justification to beliefs. (It does not make reflective access to justifiers a mandatory, necessary criterion.) Hence, adding a reflective access requirement to the basic tenet of mentalism (*vis-à-vis*, the supervenience claim, [S]) is akin to altering the basic formulation of accessibilism with a requirement that subjects have reflective access to their own justificatory status (rather than merely having reflective access to the justifiers themselves). Adding an access requirement would have the effect of paring down the range of mental phenomena upon which justification is allowed to supervene. Likewise, the mentalist could restrict the domain of mental phenomena she puts into the supervenience base of justification in other ways. For instance, she might appeal only to beliefs (*à la* traditional coherentism), only to mental states (rather than events and conditions as well), or only to occurrent mental phenomena (rather than to dispositional ones as well). Of course, adding any of these further restrictions would require arguments in favour of the restrictions.

However, for now I will put aside the question of which specific versions of internalism are sanctioned by mentalist ur-internalism. It will suffice to focus on the characteristic theses—(S) and (M)—set out by Conee & Feldman. By themselves, these theses claim that (possibly) all and (definitely) only mental phenomena are in the supervenience base for justification. Of course, to deny mentalism, the externalist need not deny that mental phenomena are relevant to justification. Such a move would be patently

---

That an account which demands that the supervenience base for justification be restricted to accessible mental states is a strengthening of internalism indicates that mentalism (as ur-internalism) holds that mental states can have justificatory potency *without being accessible at all*. This appears to amount to the claim that non-accessible mental states do their justificatory work just by being “in” the subject’s mind. Subsequently, Feldman doubts that adding constraints to the effect that only accessible mental state makes much of a difference. He claims that the distinction between accessible and non-accessible mental states is “too obscure to resolve” (ibid).

absurd and also out-of-step with both the letter and spirit of externalism (as we will see). Likewise, the externalist still need not deny that any particular sort of mental phenomenon (e.g., reflective access) is ever relevant to justification (we saw this in the externalist repudiation of accessibilism). However, unlike the externalist's rejection of the accessibilist's basic claim, here the externalist is not merely objecting to the necessity attached to the internalist requirement. After all, the externalist will think there must be something mental necessarily at work in any instance of justification—specifically, a belief (or similar mental state). Hence, the externalist's gripe with the basic formulation of mentalism is that *only* mental states are allowed justificatory potency (Vahid 2011: 145). (Or at least that is the initial appearance. We will examine this further in a later chapter.) In addition to the mental phenomena themselves, the externalist appeals to the causal histories and/or reliability (actual or counterfactual) of the mental phenomena<sup>30</sup>. For example, in determining whether Fred's beliefs about burning toast are justified, the mentalist will ask about Fred's sensory states and beliefs, querying whether he has perceptual states (like visual impressions of smoke issuing from a toaster), among other things, the mere presence of which would lend support to a belief that the toast is burning. The externalist will want to know whether Fred has such states as well, but will also want to know how reliably such states correlate with the presence of actual burning toast.

### *§2: Epistemic Externalism*

Characterised negatively, epistemological externalism is the denial that there are internalist epistemic conditions which must necessarily be met in every instance of justification. As we have seen, this involves denying that every instance of justification entails either that the subject whose belief is at issue is in principle able to reflectively access her grounds for that belief, or that the justification of the subject's belief is exhausted by what is going on in her mental life. On the positive side, externalist theses attempt to delineate what necessary external, epistemic condition(s) must be met in instances of justification. There have been

---

<sup>30</sup> It will become an issue later whether the causal histories and facts underlying the reliability statistics of mental phenomena are partly constitutive of the mental phenomena. We will put this question aside for the time being.

various rival externalist theses, the winnowing of which I will give no account here. For the sake of simplicity, I will talk about reliabilism.<sup>31</sup> As it is usually given as an account of knowledge, that is how we will look at it initially. Reliabilism can be applied to justification as well. Once we have set out the facets of reliabilist theory as it concerns knowledge, we will adjust to talk about justification.

Generic reliabilism states that knowledge is true belief arrived at by a reliable belief-forming process. Here a process is reliable in the relevant sense if it gets the right result more often than not. Thus, we can say of Fred that he knows that the toast is burning if: 1) he has a belief that the toast is burning; 2) the toast is, in fact, burning; and 3) he has the belief that the toast is burning because of a process which more often than not correctly ascertains whether or not toast is burning.<sup>32</sup>

However, we quickly run into problems here. First, the reliable belief-forming process criterion is too indiscriminate with respect to the cases for which it licenses attribution of positive epistemic status. Say that Fred believes that the toast is burning and some daemon—wishing to preserve Fred from the embarrassment of forming false beliefs—causes the toast to burn. Do we still wish to say of Fred that he knows that the toast is burning? It is counterintuitive to attribute knowledge to agents when the direction-of-fit between agents' beliefs and the world runs from the world to the beliefs. Appropriate attributions of knowledge occur when a subject's beliefs come to match the world—not the other way around. Even if this does not seem clearly counterintuitive, counter-examples which leverage the same background intuition are more convincing. Simply imagine that Fred's belief and the state of the universe coincide serendipitously: Fred whimsically tosses darts at a board to pick what he believes and happens to form true beliefs more often than not. It is more clearly counterintuitive to attribute knowledge to Fred in this case because he is merely *lucky*. However, if an agent's belief is only luckily true, it does not seem that they

---

<sup>31</sup> I do not think this tack poses any problems: it is not unreasonable to state that all viable forms of externalism ultimately come to the same point to which I track reliabilism.

<sup>32</sup> Note that this is more specific already than generic reliabilism in that the reliability of the relevant process is indexed to a certain task; *viz.*, ascertaining the toastiness of toast.

deserve epistemic credit for it<sup>33</sup>. The anti-luck intuition leveraged here is not unique to reliabilism, or indeed epistemic externalism. The basic lesson of Gettier cases<sup>34</sup> is that beliefs must be—whatever else they are—true in a non-lucky way if they are to count as knowledge.

In both the daemon and dartboard cases, Fred's belief comes out true through no relevant effort of his own. However, it seems that knowledge attributions are only appropriate when the relevant subject's true belief is in some way down to him. That is, Fred must have come to believe the truth through the exercise of some ability of his<sup>35</sup>. By restricting the class of reliable belief-forming processes from merely any which get it right more often than not, to the abilities of the subject about whose knowledge status we wish to inquire, we have banned cases such as the dart tossing and meddling daemon ones from counting as knowledge<sup>36</sup>. As John Greco (1999) has it, it is the reliability of agents we are concerned with—not processes *per se*. This appeal to abilities can be seen to answer the anti-luck intuition mentioned above or instead to answer to a second (and independent)

---

<sup>33</sup> Note that this is different from it being lucky that the agent has the belief that she does. Think about Fred again. Suppose that in between throwing darts, he is watching out for a particular person to enter the room so he can give her some money he owes her. Just before he makes a toss, he looks to the door and sees the woman he is waiting for enter the room, subsequently forming the (reliable and true) belief that she has arrived. However, because he was about to make a toss, he could have easily not been looking at the door—in a sense, it was *lucky* that he looked when he did. Hence, it is lucky that Fred has the true belief that he does have. However, this is not the sort of luck with which we are concerned. In the daemon and dartboard cases above, what is at issue is the fact that Fred's beliefs are *luckily true*, rather than luckily possessed at all. Beliefs which are luckily true are subject to what Pritchard (2005) calls “veritic luck”; which is a type of luck that undermines the epistemic status of any belief characterized by it.

<sup>34</sup> Briefly, a Gettier case is one in which a subject has a belief which is justified and true, and which is true not in virtue of being justified but rather by accident. (See Gettier 1963 for the original cases.) Zagzebski (1996: 288) helpfully notes that these cases will arise for any account of knowledge (internalist or externalist) in which “there is a small degree of independence between this other element [i.e., true belief] and the truth . . .” The fake-barn cases to which I will often refer in this work are Gettier-style cases with respect to reliabilism.

<sup>35</sup> I ignore a pure modal or anti-luck strategy for solving this problem because I take it to have been demonstrated that such positions, when properly elaborated to avoid counterexamples, collapse into positions which invoke an ability intuition. I will not discuss this claim here, but direct the reader to Pritchard et al (2010: Chp. 3, §3) and Pritchard (forthcoming) discussion of the matter.

<sup>36</sup> This discussion misses some of the subtleties of the in-house debate regarding what exactly constitutes getting-it-right-more-often-than-not (Conee & Feldman 1998). This is not the place to debate such matters.

intuition—*viz.*, the ability intuition (Pritchard forthcoming). For now, we will assume that an appeal to abilities answers the anti-luck intuition, and will not countenance a separate ability intuition. We will re-evaluate this assumption later.

How do we translate this to talk about justification? After all, some philosophers have argued that the internalism-externalism debate should not be thought of as concerning the nature of justification, but the necessity of justification for knowledge. For instance, Feldman (2004: 148) claims that treating externalism as a thesis about justification gets things wrong. He draws this conclusion by considering two things. First, he notes that it is uncontroversial that justified beliefs can be false. Second, Feldman claims that it would be hard to make sense of this fact about justified beliefs on causal epistemic theories. Here, the sort of causal theory he has in mind is one which stipulates that a belief's epistemic status arises in virtue of some causal connection between the state-of-affairs which the belief concerns and the belief itself. Hence, if Fred's belief that the toast is burning is to have a positive epistemic status (on the externalist theory Feldman envisions), that belief must be in some way caused by the fact that the toast is burning. For instance, if gouts of flame and smoke are erupting from his toaster, the flames and smoke must cause Fred to believe that the toast is burning by affecting his visual and/or olfactory experience.

But, says Feldman, such facts either do or do not cause Fred's belief. If these facts do cause Fred's belief, then the belief is true and has a positive epistemic status. However, if the flames and smoke do not cause Fred's belief, then the belief lacks a positive epistemic status. Feldman's point is that, on the sort of externalist theory he envisions, beliefs only have the relevant positive epistemic status when true. But on the assumption that a belief can have the particular positive epistemic status of being justified, the positive epistemic status that the externalist tries to capture cannot be justification. Thus, the internalism-externalism debate in epistemology is not one about the nature of justification.

However, there is a way to frame an internalism-externalism debate about justification. Moreover, it can be done with a causal version of epistemic externalism, *pace* Feldman. Think about the dartboard case. Again, Fred forms some belief—say, that it is raining outside—by tossing a dart. If the dart hits a red section, Fred will believe that it is raining; if the dart hits a black section, Fred will believe that it is not raining. (The board has only red sections and

black sections.) Fred tosses a dart, it lands in a red section, and he duly begins to believe that it's raining outside. Suppose that it is in fact not raining outside. Another person, Jethro, asks Fred if it is raining outside. Fred says it is. "Why do you believe that?" Jethro asks. Fred replies, "Because the dart I threw landed in a red section of the board". No doubt, Jethro will be perplexed. Fred's belief that it is raining is intuitively unjustified. Note this is not because it is actually not raining: justified beliefs can be false, as has been noted (unlike knowledgeable ones, as has also been noted). The problem is that there is something suspect about the way Fred forms his belief. Specifically, a dart toss is not a good reason to believe things about the weather. Why not? Because the results of dart tosses are not good indicators of weather conditions. That is, dart tossing is not a *reliable* way to form weather beliefs, where 'reliable' has the sense indicated in the above discussion about cases of knowledge. We can say a belief is justified, then, when it is the outcome of a reliable process<sup>37</sup>. More specifically, we can say that a belief is justified when a subject has that belief *because* of the exercise of a reliable process, where 'because' is given a causal-explanatory reading (Greco 2007; Goldberg 2007). If that belief is also true (and true because of the exercise of the reliable process), then it will be knowledge; if not, it will merely be justified. Nevertheless, this is still a positive epistemic status. Hence, if Fred forms the belief that it is raining by looking outside, that belief was caused in him by the use of his visual faculties (which we will assume are reliable). Thus, *contra* Feldman, the externalist deploying a causal criteria need not claim that it is the state-of-affairs with which a belief is concerned (e.g., that it is raining outside) that causes a belief, but rather the exercise of a reliable process. Further, because a reliable process is not an infallible one, a belief may be caused by the exercise of such a process and yet be false. The externalist can use causal connections to talk about positive epistemic statuses of beliefs which obtain even when the beliefs in question are false. That is, the externalist can give her a non-internalist account of the nature of justification.

---

<sup>37</sup> A reminder that we are here interested in doxastic justification, rather than propositional justification. That is, we are concerned with the epistemic statuses of subject's actual beliefs, rather than the statuses of beliefs subjects might have, but do not actually have.

Returning to the reliabilist about knowledge, she is not out of the woods yet. Even having restricted the class of reliable process to which he may advert, our epistemologist still cannot give a satisfactory account of knowledge. If we look at the definition of ‘ability’ offered by Greco, we can see why.

S has an ability [A, to achieve a result, R] A(R/C) relative to environment E [in conditions, C, of E] =<sub>def</sub> Across the set of relevantly close possible worlds W where S is in C and in E, S has a high rate of success achieving R (2007: 66).

Here is an example. Fred can walk (ability) from his house to his grandmother’s house (result, R) on solid, friction-holding surfaces in normal gravity (environment, E), provided that he is not being mauled by wolves, that there are no earthquakes occurring, etc. (condition, C). The remaining difficulty is that by indexing abilities to environments, conditions and results, we make abilities abundant. That is, both which (and how many) abilities a subject has—and specifically which ability she used to achieve a given result—will differ with respect to how environments and conditions are specified. Think about Fred’s belief concerning his toast. We can add details to the case to fill out all the variables in Greco’s account. On the occasion in question, Fred is clear-headed, bright-eyed and bushy-tailed. His kitchen is well-lit by the morning sun and he is carefully observing the toaster in between answering crossword clues. Fred is the subject, and we are concerned to know whether he has the ability to form true beliefs about the burned/non-burned status of his toast in a well-lit environment while in a clear-headed, attentive and largely undistracted condition. It seems Fred can readily form true beliefs about his toast in the situation. But what exactly is the ability he is deploying? If we add to our description that it is Tuesday, is the relevant ability *forming true beliefs about toast on Tuesdays*? Or, perhaps the ability is *forming true beliefs about toast in the morning*. If the abilities are relevant to the environment and conditions, then we can create abilities of a finer grain by providing finer-grained conditions.

Problematically, this leaves us with no reasons in favour of attributing Fred’s belief to one ability over another—even in the case that some of the specifications differ with respect to their relative reliability. Some specifications may even be patently unreliable. This is a



version of what is known as the Generality Problem (Conee & Feldman 1998). If this is the case, however, the theory leaves us unsatisfied. Upon deciding that Fred knows that the toast is burning, if we ask, “But in virtue of what does he know that the toast is burning? Which of his myriad abilities did he use?” we will be met with a long list of answers and no non-arbitrary way to pick one from the list. Moreover, the level of reliability will vary as a function of which ability is picked. Hence, whether a belief is the product of a reliable process or not will depend on which ability is claimed to have produced it. However, this will affect both whether the belief is a case of knowledge and whether it is justified, as those epistemic statuses both depend on a belief being produced by a reliable process. Thus, the externalist needs to further restrict the class of abilities to which she can appeal when making attributions or knowledge and justification. How might she do this?

We might answer this question by appealing to the class of cognitive or intellectual abilities or characteristics. These are the abilities to discriminate, gather, process and remember various pieces of information. This category includes our senses, reflective abilities and various reasoning skills, among other things. It is important to note that cognitive/intellectual abilities are not solely the province of the epistemological externalist. Internalists can (and do<sup>38</sup>) also appeal to these sorts of abilities. Thus (foreshadowing the main argument of this work) insofar as externalists must deny internalism, they are at pains to ensure that the characterization they give of cognitive/intellectual abilities is distinct from the way internalists describe such abilities. But why should invoking particularly cognitive abilities go anywhere towards solving the generality problem? After all, perception has been adverted to in all the previous examples, which admittedly fell prey to this concern about generality. Labelling them “cognitive” will not change any of that.

I take it that the advantage to this appeal comes from the idea that there are branches of science which study cognitive abilities, and which can hone in on the grain of these abilities. The combined yield of (evolutionary) biology, (cognitive and perceptual) psychology, neuroscience and the like plausibly inform us as to many features of cognition. For instance, it might be that we have separate abilities to spatially locate objects by vision and to identify objects by vision (Ungerleider & Mishkin 1982; Goodale & Milner 1992)—

each underpinned by different functional neurological areas<sup>39</sup>. On the other hand, it is unlikely that any of these branches of science will reveal that we have different abilities on Tuesdays relative to other days of the week. Moreover, as we will discuss in later chapters, it might be discovered that our cognitive abilities are partly constituted by portions of our external environment. In any case, the appeal to specifically cognitive abilities deals with the generality problem by isolating the fineness of grain of abilities. There may still be an element of shiftiness, as Greco (2009) points out. Different sciences have distinct explanatory goals which may affect the grain at which they describe abilities. This feature of abilities is taken to be inherited from the interest-relative nature of causal explanation, of which explanation by ability is a sub-category on Greco's account. However, this remaining shifty element is not problematic, as the different abilities specified in different contexts will be so specified in a principled manner rather than an arbitrary one. That is, the interests and purposes of those theorists evaluating the situation will determine the correct grain of ability. This is taken to solve the generality problem because the question, "Which ability did Fred use?" is no longer an open question with only arbitrary answers.

However, first note that even having gone to all this trouble, the externalist cannot rest, for there are examples of abilities of the agent which are apparently stable, reliable, cognitive processes which intuitively do not confer justification on the beliefs they produce. The brain lesion case is such an example. Alvin has a brain lesion which causes him to form true beliefs about complex mathematical propositions. This lesion is reliable because it always prompts him to believe correctly, it is stable because it is always there, and it is apparently cognitive as it seems to function like an extremely fast reasoning capacity. (Also, the specificity of the domain of propositions upon which it causes him to form beliefs makes it look like a good candidate for a cognitive ability.) However, the lesion does not function in concert with the rest of Alvin's abilities. As Pritchard (2010) puts it, the lesion is not *integrated* with Alvin's other abilities; a fact which intuitively bars the beliefs produced by the lesion from accruing justification. So, the abilities the externalist must appeal to for their account

---

<sup>38</sup> Linda Zagzebski (1996) makes an epistemic internalist appeal to abilities.

must be stable, reliable, *integrated* cognitive abilities. Moreover, these abilities must be cashed-out in a particularly externalist way. What this particular construal of abilities is will be the subject of the subsequent two chapters.

### §3: *Confirmation Criteria*

The dialectical structure of the epistemic internalism-externalism debate is fraught. Historically, internalism seems to be the default position<sup>40</sup>. Internalism about knowledge (where this is the position that there are *no* external factors relevant to knowledge) is not a position that anyone holds, if indeed anyone ever did defend it<sup>41</sup>. A more reasonable interpretation of an internalist position concerning knowledge is as follows: whatever factor must be added to true belief to account for knowledge, that factor can be given an internalist treatment. Hence, a theorist who thinks that the justified true belief account of knowledge is correct counts as an internalist about knowledge if she is an internalist about justification. Even if the more extreme form of knowledge-internalism is not—and never was—defended by anyone, this latter form of knowledge-internalism does have defenders (e.g., Chisholm 1989; Feldman 2004). Thus, the issue of internalism versus externalism about justification in which we are interested is doubly interesting: i.e., in its own right and as it pertains to internalism versus externalism about knowledge.

With regard to justification itself, we have looked at two types of internalist theory—*viz.*, accessibilism and mentalism. Considering what these theories promulgate, we can formulate confirmation criteria for epistemic externalism, negatively construed, just as we formulated confirmation criteria for cognitive externalism in the previous chapter.

---

<sup>39</sup>I am not suggesting that the state-of-the-art empirical data support this hypothesis: recent studies cast doubt on its viability (e.g., Franz et al. 2000). The point is just to demonstrate what sort of distinctions amongst abilities might be evinced by empirical research.

<sup>40</sup>This point will be discussed further in Chapter 5.

<sup>41</sup>John McDowell's position in "Knowledge and the Internal" (1995) *might* be construed as a defense of radical internalism. However, it is more likely that his position should be interpreted in the fashion of Williamson (2000): i.e., that there is something flawed about attempting to separate internal from external factors.

(*Negative Epistemic Externalism Confirmation Condition*) There is at least one possible instance of justified belief for some subject, a justificatorily potent basis for which is:

- a) Not accessible to the relevant subject, in principle, by reflection alone
- And
- b) Not part of the relevant subject's mental life (construed as internal).

I take it that the instance need only be a possible one because both accessibilism and mentalism are necessary theses about the nature of justification. That is, I take both sorts of position to endorse a *strong supervenience* claim. A strong supervenience claim is one which dictates that two individuals, in two distinct possible worlds, are duplicates with respect to the supervenient property if (and only if) they are duplicates with respect to the subvenient property. For the accessibilist, this means that if Fred (a denizen of the actual world) and bizarro-Fred (a denizen of bizarro-world) are duplicates in terms of the things to which they have reflective access, then they are also duplicates with respect to the justification they have for their beliefs. Pryor does not state that SI is a strong supervenience thesis. However, I take it that it is safe to assume SI is such a thesis for the same reasons that Conee explicitly states (S) must be a strong supervenience thesis<sup>42</sup>.

Condition (a) is derived from the accessibilism. As the accessibilist claims that in *every* instance of epistemic justification subjects can access (by reflection alone) those facts which justify their beliefs, the negative epistemic externalist need only find one case in which a subject is justified but lacks such access (or has such access, but is justified in part due to factors not amenable to reflective access as well).

Condition (b) is derived from mentalism. The mentalist claims that only subjects' mental states are relevant to justification. Specifically, it is claimed that this is an internalist thesis because justification supervenes on the mental, which is itself internal. The specification that the mental is construed internally does not make it explicitly into the

formal articulation of mentalism in the theses (S) and (M) mentioned earlier, because it is simply assumed to be the case. I flag the internal construal of the mental in the confirmation criteria because it will become important later. In our discussion of mentalism, it was pointed out that to deny the mentalist position is to deny that justification is exhausted by the mental: i.e., the epistemic externalist says there are non-mental factors relevant to justification. However, making explicit the construal of the mental as external appears to open a new avenue for the epistemic externalist. Perhaps, in denying the mentalist, the epistemic externalist can agree that justification is exhausted by the mental while at the same time denying that the mental is internal. This will involve accepting all of the formal apparatus of mentalism (i.e., [S] and [M]), but repudiating the implicit assumption I have identified. This issue will be further explored in Chapters 4 and 5.

The condition is rendered as a conjunctive one: the epistemic externalist must deny both accessibilism and mentalism. However, if the way Conee & Feldman describe mentalism is accurate, then it includes accessibilism—in which case, denying mentalism entails denying accessibilism. It might seem that it is unnecessary, then, to be as long-winded about the confirmation criteria for negative epistemic externalism. However, there is a rationale for formulating the criteria this way. In Chapter 5, we will examine reasons to doubt whether mentalism is actually an epistemic internalist-friendly theory. Should such reasons prove good, then the epistemic externalist will not be obliged to deny mentalism. If that were the case—and if the confirmation criteria were formulated only with the denial of mentalism as their object—it might appear that the confirmation conditions for negative epistemic externalism were empty. However, this would not be the case: the epistemic externalist would still need to deny what the accessibilist claims. That is, even if the epistemic externalist were not at pains to deny that a subject's justificatory status supervenes only on her mental life (broadly construed), the epistemic externalist would still be obliged to deny that a subject's justificatory status supervened only on the facts to which she (in principle) has reflective access. Hence, I render the confirmation criteria as a conjunction of the denial of accessibilism and mentalist.

---

<sup>42</sup> See footnote 26 of this work.

However, the confirmation of negative externalism is not sufficient for our purposes. The argument I will make in chapter four will concern the commitments of the positive account of epistemic externalism that we have explored in this chapter—*viz.*, externalist theories which posit an ability condition on positive epistemic statuses. While negative epistemic externalism is clearly weaker than epistemic internalism with regard to the number of cases required to confirm it, positive epistemic externalism is not (clearly) weaker in this respect. Indeed, epistemologists putting forth positive versions of externalism seem to take up their internalist counterparts' ambitions to describe all instances of justification. Frequently, externalists will give their conditions for justification in the form of biconditionals (i.e., a belief is justified if and only if some condition, C, obtains), which suggests that they mean to encompass every possible instance of epistemic justification with their account. Of course, it would be too quick to conclude anything merely on this basis. After all, it has often been opined that theorists are all talking about different things when using 'justification'. The extent of this equivocation has led many to abandon the term all together (e.g., Plantinga 1993; Zagzebski 1996<sup>43</sup>), or simply stipulate that they are talking about one notion of justification rather than another (e.g., Burge 2003<sup>44</sup>). Pluralism about justification ensues. However, if this is the case, then perhaps mentalists, accessibilists and epistemic externalists are all concerned with distinct phenomena when they speak of justification. But then, the ambitions of these respective theorists would be to capture all instances of the particular thing in which they are interested. Further, these respective ambitions need not be at odds.

There is a way to put the debate such that externalists and internalists are arguing over the same epistemic status, construed in the same way. John Greco (2005) notes that there are two general modes for evaluating beliefs: an objective mode and a subjective mode. In the first mode we evaluate how subjects' beliefs fit with the facts—how the world is. In the

---

<sup>43</sup> Zagzebski actually advocates abandoning justification as a primary target of epistemic analysis, rather than bemoaning the problems its analysis has caused.

<sup>44</sup> It is slightly inaccurate to say that Burge opts for one view of justification over another. Instead, he takes the general term for epistemic support to be "warrant", of which there are two kinds: *viz.*, so-called "entitlement" and "justification". Burge reserves the former for externalist theses, and leaves the latter to the internalist.

second mode we evaluate how subjects' beliefs fit with their evidence—how subjects take the world to be. One might think that the externalist justification is justification in the objective mode, while internalist justification is justification in the subjective mode. The two sorts of theorist can go about their business independent of each other. Not so, claims Greco: no internalist theory is true of any epistemic status. Even subjective appropriateness is better explained in an externalist fashion. Whether Greco's arguments to this effect succeed is not important here. What is important is that there is a way of construing the debate over epistemic justification in which the internalist and the externalist are genuinely at odds, rather than using the same term to describe different epistemic statuses. In this context, the use of biconditionals in formulating conditions for justification does indicate an ambition to encompass all cases. Both the positive externalist and the internalist about justification intend to give an account of all possible instances of epistemic justification. Thus, the confirmation criteria should be put as follows.

*(Positive Epistemic Externalism Confirmation Criteria)* All possible instances of epistemic justification are instances of beliefs arrived at via the exercise of reliable, cognitive abilities<sup>45</sup>.

The primary question which will engage us for the remaining chapters will be, "What do the confirmation criteria for cognitive externalism and positive epistemic externalism have to do with each other? Does the confirmation of one speak to the confirmation of the other?" Answering this question will take a lot of work. First, we will need a better grasp of the notion of ability as it will play an important role in the arguments of chapter four. Chapter 3 will be devoted to exploring this notion.

---

<sup>45</sup> Even should this be an inapt description of the ambitions of epistemic externalists in Greco's mould, my later arguments will not be undone. As the confirmation criteria for cognitive externalism require only actual instances of extended cognitive processes, I could make my argument concerning the overlap of these confirmation criteria with those of positive epistemic externalism even if the latter were a contingent thesis. That is, even if positive externalism describes only actual instances of justification, my further arguments would not suffer.

## Chapter 3

### Ability in Philosophy of Cognitive Science and Epistemology

#### *Introduction*

Ability is a key notion in both the epistemology and philosophy of cognitive science literature. In epistemology, it is generally thought that epistemic success must be attributable to the agent through some ability of theirs for any positive epistemic status to accrue to their doxastic attitudes. But which abilities are relevant to epistemic success, and when do agents have these abilities?

In philosophy of cognitive science, one of the primary questions is what are the constitutive elements of our cognitive abilities: are they exclusively factors internal to brains and/or bodily individuals; or are there some constitutive elements out in the world, external to brains/bodily individuals and/or the social structures, in which these individuals participate?

It would be helpful to have a grasp on the notion of ability in order to give satisfactory answers to these questions. There does exist a philosophical literature on the analysis of ability, though it has progressed independent of the two debates with which we have been concerned. (The debate has primarily unfolded in the context of philosophy of action—



more specifically concerning the compatibility of free will and determinism.) There are definitions of ability bruited in the epistemological literature as well, but as ability is not the primary focus of this literature, authors understandably do not dwell on it overmuch. Below, I plan to dwell on ability a bit in order to draw out whether philosophers of cognition and epistemologists mean to pick out the same sorts of thing with ‘ability’; and indeed, whether internalists and externalists within sub disciplines mean to pick out the same sorts of thing.

### *§1: Ability Attribution*

In a discussion of J.L. Austin’s (1956) ruminations on ability, A.M. Honoré (1964) introduces a distinction between two senses of ‘can’—*viz.*, a particular sense and a general sense. The former sense, says Honoré, implies success in the relevant situation, and the denial of a can-claim (a can’t-claim) in the particular sense implies failure.

‘[Can] he sink this putt?’, where the ‘can’ is particular, is like ‘will he sink this putt?’ . . . Conversely, ‘he can’t sink it’ virtually means ‘he won’t sink it’. . . [Further] suppose someone has succeeded in doing something. It is perfectly correct to use ‘could’ or ‘was able to’ to express this. Thus, ‘I could[/can] see you in the undergrowth’ is properly said only when I have succeeded in seeing you (464).

On the other hand is the general sense of ‘can’. Honoré contrasts the general sense with the particular sense first by noting that the former involves types of action, whereas the latter involves particular actions (tokenings of an action type). As opposed to predicting or implying success on a given occasion, the general sense of ‘can’ is used to indicate a “general competence, ability or skill” on the part of the relevant agent.

If a typist claims that she can type sixty words a minute, she is asserting that normally or on average she types at that rate, when she wants to, not merely that she has once managed by a fluke to get sixty words in a minute (465).

Unlike the particular sense of ‘can’, it is appropriate to use the general sense even when someone fails on a particular occasion. Thus, a golfer generally-can sink long putts even if she has just failed to sink such a putt. Just as a single failure does not imply that someone

generally-cannot do something, a single success does not imply that someone generally-can do something. Hence, it may be that a golfer generally-cannot sink long putts even if she has just sunk one such putt. This will appear problematic if we phrase the claim as, “The golfer sunk a long putt, but she cannot/is not able to sink long putts”. But Honoré’s point is that this is only problematic until we realise that there are separate uses of ‘can’ with different conditions under which it is appropriate to assert them. Further, these separate uses indicate different things about the agent in question. Focusing on the general sense of ‘can’, Honoré notes the importance of normality.

But if no qualification is added, it is implied that the ‘can’ statement refers to the agent’s normal or standard performance . . . at least in the cases where ‘can’ is unemphatic. If, on the other hand, emphasis is placed on ‘can’, as in, ‘he *can* make a good after dinner speech’, the implication is that he does not usually make a good speech, either because he does not try or because on the few occasions on which he speaks well, this is attributable to luck, that is, to some rare circumstance or combination of circumstances (465).

So, saying of an agent that she generally-can do something implies that normally she succeeds at doing that thing. The second part of the above quote, along with the latter part of the quote concerning the typist above, highlight the role of luck and special circumstances in making ‘can’ claims about agents. In both the dinner speech and typist examples, Honoré notes that it is correct to claim that the respective agents generally-can succeed at the relevant action only if their success is not lucky. In the dinner speech case, the notion of luck is fleshed-out as one particularly concerning special circumstances. Should this ineloquent dinner guest make an articulate speech, it will not be because he has an ability to orate well, but because a snippet of obscure poetry came to his mind or because a French scientist hiding in the bushes was beaming thoughts into his head (for instance). Moreover, Honoré seems to mean that the general sense of ‘can’ implies not only performance normally issuing in success, but also performance normally issuing in success under normal circumstances.

Thus, if a golfer can (general) hole six-foot putts, he can (general) hole this six-foot putt, provided there is nothing special in the circumstances of the case to prevent him from doing so on this occasion. Conversely, a golfer who

can't (general) hole six-foot putts can't (general) hole this six-foot putt unless there is something in the special circumstances to enable him to do so (466).

Again, actual success does not imply that an agent generally-can do something, nor does actual failure imply that an agent generally-can't do something. In either instance, luck can be the mitigating factor. Nor does possible success or failure in a particular instance imply that an agent generally-can or generally-can't do something.

'He could have, if he'd kept his feet together' is true if it is the case that he had the necessary ability and that, if he had kept his feet together, there would have been nothing to prevent him, which does not entail that the golfer would have actually sunk the putt, even if he had kept his feet together (467).

As Honoré notes, all of the talk in terms of a general sense of 'can' may be put in terms of ability talk. The question now is how well this ability talk meshes with the notion of ability at play in philosophy of cognitive science and epistemology.

### *§2: Ability in Philosophy of Cognitive Science*

It is somewhat difficult to extract a general notion of ability from the proceedings of philosophy of cognitive science. Most of what is to be gleaned comes from the discussions of modularity (e.g., Fodor 1988) and levels (e.g., Marr 1982). Further, we have in Chomsky (1965) the distinction between competence and performance. In each of these discussions, what we have called abilities or general uses of 'can' are seen as an individual's idealised capacities. That is, the goal of cognitive scientific inquiry is to determine what cognitive outcomes an agent generally-could achieve across a range of possible sets of input. There is much literature on what human idealised capacities amount to, but nothing in the way of consensus. Indeed, with respect to this issue Rupert (2009) merely notes that, "It is a difficult problem to individuate kinds of mechanism, ability, capacity, etc. This is a general problem, faced by all parties to the debate" (42). In what follows, we will look for a notion of cognitive ability which both cognitive internalists and externalists might agree upon.

We can get somewhere by looking more closely at what Rob Rupert says against the cognitive externalism of Rob Wilson (2004). Rupert makes the following claim.

Cognitive science should explain how [a] subject comes into that relation [with the world] by adverting to that organism's local persisting cognitive capacities that can manifest themselves in a variety of ways in a variety of circumstances (2009: 83)

This is claimed in opposition to what Rupert takes Rob Wilson to claim. We saw in Chapter 1 that Wilson's route to cognitive externalism is via the notion of realisation. Wilson takes it that various actions can be accomplished by doing something which is identical with respect to purely bodily description—*viz.*, signing a slip of paper (2004: 116). Fred might express his disgust for the actions of some political or corporate body by signing a petition, agree to trade some of his time and expertise for a salary and benefits by signing a contract, or indicate his preferences for dispensing with his belongings *post-mortem* by signing a will. In Wilson's terms, these actions are widely realized: i.e., they are phenomena “[with a total realization] whose noncore part is not located entirely within [the individual to whom we attribute the phenomena]” (*op cit* 112). These actions are widely realized because they would not be the very actions they were without the inclusion of factors external to the individual to whom we attribute the action. That is, what allows us to differentiate one instance of Fred signing a paper from another are the social and legal conditions surrounding the signing.

In addition, Wilson claims that the phenomena in question are *radically* widely realised: i.e., they are phenomena “[with] a wide realization whose core part is not entirely located within [the individual to whom we attribute the phenomena]” (116). In the case of signing a paper, the wide part of the core realisation is the individual signatory's interaction with pieces of the world—*viz.*, the pen and piece of paper.

However, according to Rupert, this entails an unnecessary proliferation of abilities. And while these (pseudo) abilities might be interesting for some purposes, Rupert claims they are not interesting from the perspective of studying cognition. What is interesting from that standpoint is the sense in which some organism succeeds in doing one type of thing on numerous occasions, the conditions of which occasions may vary. In the disparate signings,

Fred manifests one “cognitive-motor capacity” (2009: 83). Or, as we might say, Fred exercises one-and-the-same ability on each of numerous occasions rather than a different ability on each occasion. However, this discussion of Wilson’s account of realisation is undertaken to show that cognitive abilities are not underpinned by mechanisms which outstretch the organism to which those abilities are attributed. There is not yet much to suggest an account of ability itself. Rupert’s talk of “cognitive capacities that can manifest themselves in a variety of ways in a variety of circumstances” sounds a lot like normally succeeding in normal circumstances. However, such an observation demonstrates rather little.

Perhaps we should look to Rupert’s favoured way of demarcating systems—via the notion of integration. We looked briefly at his method of measuring integration in Chapter 1<sup>46</sup>. In short, the factors which will turn out to be part of an integrated cognitive system are those which appear frequently in sets of factors that have sufficiently high conditional probability scores (at least  $P \geq 0.5$ ) with respect to various, pre-theoretically determined cognitive outcomes. These are the factors which will underpin an organism’s cognitive abilities.

Now suppose that there is some cognitive outcome for which no set of factors reaches the threshold conditional probability (for a given subject) Rupert mentions. In this case, it would seem that Rupert’s account delivers the verdict that none of the factors in question underpin an ability to achieve the relevant outcome in the subject. These factors may of course be part of the integrated cognitive system of the organism in question because of their participation in sets with sufficiently high conditional probability with respect to other cognitive outcomes, but that is not the point here. Instead, the point is that it seems we can conclude that a subject only has an ability to achieve some cognitive outcome if there is at least one set of factors which reaches the threshold of significance Rupert posits. Otherwise, though the relevant organism may achieve the outcome many times, there will be nothing stable enough to pin the organism’s successes on to call it an ability. It will be as if the

---

<sup>46</sup> Please refer back to Chapter 1 for the extensive quote from Rupert outlining the technical details of his account. For the purposes of this chapter, only a basic grasp of the details is requisite.

organism is always in special circumstances when successfully achieving this cognitive outcome—always getting lucky.

Perhaps a more concrete example can drive this idea home. Suppose we want to know whether Nic can predict the next day's weather. We note that he is more often than not successful in giving correct predictions. However, further investigation demonstrates that Nic *never* uses the same method to formulate his predictions. Not only are the external aids he uses different on every occasion (haruspicy, advanced satellite imagery, tarot cards, observing animal behaviour, spirit communication, etc.), but that each external aid manages to engage *different* portions of his brain and body on the various occasions. Does Nic have an ability to predict the next day's weather? It seems not. Even though he is highly successful, it seems that his success is explained away in each case by special circumstances, or luck. But if this example captures what Rupert's more formal apparatus would judge, then it seems we can reconstruct a notion of ability from Rupert's account of which factors underpin abilities. Moreover, what we have reconstructed is consistent with what we got from Honoré in the previous section, at least with regard to the constraint that a subject must normally succeed in normal circumstances. And clearly, actual success on a particular occasion will not imply the possession of ability. Further, assuming that Rupert does not require that an organism succeed in achieving a cognitive outcome every time it attempts to achieve it (which seems a reasonable assumption, in general), we can further conclude that our Rupert-inspired account of ability will not count actual failure on a particular occasion to imply lack of ability. It is less clear that we can make any remotely plausible speculations concerning the judgments this reconstruction would license regarding possible success or failure. However, I will simply assume that Rupert would require some sort of stability across situations relevantly similar to actual occasions.

Of course, Rupert suspects that his method of measuring integration will deliver the result that all of the factors which are part of an organism's integrated cognitive system are within that organism. After all, the fact that the organism moves through an environment means that the external factors with which it is related will come and go, as will their contributions to its cognitive outcomes. On the other hand, the internal factors will always be with the organism, as will their contributions to the organism's cognition. However, as he

takes it that this is not a forgone conclusion of his model, it might turn out that (either currently, or in the future) some of the factors which are part of some organism's integrated cognitive system are external to that organism. Thus, we might take our reconstruction of an account of ability from Rupert to be one amenable to both the cognitive internalist (like Rupert) and the cognitive externalist. However, even if we were not comfortable making that claim, I think it is possible to reconstruct a similar account of ability from cognitive externalist accounts of the factors which underpin ability.

As a brief example, consider dynamic coupling arguments for cognitive externalism, of which some strands in Clark & Chalmers (1998) are an example<sup>47</sup>. Recall that there are a number of conditions stipulating in what cases external props could properly count as part of a cognitive process. These were that the resource be readily available when it would be useful, frequently relied upon when it would be useful, and that the deliverances of the resource be automatically endorsed (Clark & Chalmers 1998: 17; Clark 2010: 46). Further, given a reasonable understanding of Clark & Chalmers' so-called Parity Principle<sup>48</sup>, we should apply the same criteria for internal resources. Should an internal resource lay fallow, be difficult to utilize or be treated suspiciously, we should look askance at its inclusion as part of a cognitive process. As these criteria were adduced to address concerns about the "portability and reliability" (*op cit* 10 – 12) of external resources, we can see that this sort of argument for cognitive externalism does not differ much from Rupert's argument for cognitive internalism. Both arguments are concerned to demonstrate something about the intimacy of organisms' interactions with external factors and what affect those factors have on the organisms' cognition. However, they come to different conclusions. In any case, I will take the similarity of method between the two to be a *prima facie* case that we can extract a relevantly similar account of ability possession from the respective accounts of ability constitution. The cognitive internalist and the cognitive externalist will agree that a subject has the ability to retrieve items from memory, but will disagree on the sorts of things which

---

<sup>47</sup> For further development of this strand of Clark & Chalmers' argument, see Clark (2008; 2010).

<sup>48</sup> Parity Principle: "If, as we confront some task, part of the world functions as some process which, were it done in the head, we would have no hesitation in recognizing it as a cognitive process, then that part of the world (so we claim) is part of the cognitive process" (Clark & Chalmers 1998: 8).

(can) constitute that ability. The cognitive internalist takes cognitive abilities to be narrowly constituted, while the cognitive externalist takes some abilities to be widely constituted.

### *§3: Ability and Epistemology*

In the case of one prominent example of epistemic externalism, the offered notion of ability is more-or-less in lockstep with Honoré's analysis of ability. Recall that John Greco gives the following definition of ability.

S has an ability [A, to achieve a result, R] A(R/C) relative to environment E [in conditions, C, of E] =<sub>def</sub> Across the set of relevantly close possible worlds W where S is in C and in E, S has a high rate of success achieving R (2007: 66)

“S has a high rate of success achieving R” is just to say that S normally succeeds at R.

Requiring that an agent, in order to possess an ability, must have a high rate of success in achieving some outcome only in certain environments and conditions indicates that the successes and failures which matter are those occurring under normal circumstances.

Consider running. If we want to determine whether Colin has the ability to run 42 kilometres in under three hours, we won't care if he fails to do so when suffering from a fever or in gale force winds (on the negative side), or if he succeeds in doing so when under the effect of performance-enhancing drugs or situated on a long moving sidewalk (on the positive side)<sup>49</sup>. Colin's ability will be determined by his performance when in unsupplemented good health on normal terrain under reasonable atmospheric conditions, etc. This meshes with Honoré's contention that attributions and claims of ability exclude success and failure under special circumstances.

Finally, both Greco and Honoré stipulate that it is not merely actual success and failure which are relevant to ability, but also possible success or failure. In Greco's definition, this is made obvious by the reference to relevantly close possible worlds. While Honoré does

---

<sup>49</sup> In his discussion of ability, only performance-impairing environments and conditions are considered. Discounting enhancing factors is consistent with his account, however.



not talk in terms of possible worlds, I have already noted what he says to indicate possible success and failure are relevant to ability attributions.

A brief aside: Note that when we fully appreciate what it is to possess an ability in general, the notion of a reliable ability appears strange. If one possesses an ability to achieve some result when (and only when) one normally (under normal circumstances) achieves said result, what would it mean to possess further a reliable ability to achieve that result? Think of Colin's case. To be attributed the ability to run 42 kilometres in under three hours he must not only have actually succeeded in doing so on most of the occasions when he attempted to do so, but it must also be true that he would have succeeded on most occasions on which he might have tried. In what way would Colin's record need to be augmented for him to be *reliably* able to run 42 kilometres in under three hours? There is no space to significantly improve either Colin's actual or counterfactual running record without making it flawless; but it is obvious that perfection is not the intended target of 'reliably able'.

This seems to make a fetish out of the habitual invocation of reliable abilities by epistemic externalists who adopt an ability condition. While there is clearly a difference between a process which reliably issues in a desired outcome, and a process which does not so issue, that is because a process is not individuated by its success statistics. Here is an example of a process: rejecting applicants to a job who disagree with any position you hold (call it the "parochial process"). This process has an input (a job application) and an output (rejection/non-rejection). The parochial process is only reliable or unreliable with respect to a certain outcome. For instance, it reliably produces like-minded cohorts in a department; and it unreliably produces diverse cohorts in a department. Compare this with the *ability* to produce like-minded cohorts in a department. The goal (*viz.*, producing like-minded cohorts) is built into the description of the ability; hence, the reliability of the ability could not change as a function of goals. The reliability of an ability could change within a range as a function of the ratio of success to attempts. However, given the account of ability on offer, this ratio could not drop below whatever proportion of successes to attempts is indicated by "a high rate of success" or else no ability would be possessed. The same applies to counterfactual evaluations of success.

I speculate (and stress that this is highly speculative) that the continued use of ‘reliable’ to describe abilities in the context of accounts of ability which seem to render talk of reliability redundant is simply a holdover from process reliabilism. The epistemic status belief formed on the basis of some process (individuated by the actions and mechanisms it involves) can be questioned on the grounds of the reliability of that process with respect to the end of forming true beliefs (or whatever). A particular process (e.g., throwing darts at a dart board) may have an abysmally low rate of success (actual or counterfactual) with respect to forming true beliefs (or whatever). An ability to form true beliefs with regard to a certain topic (e.g., the colour of facing surfaces) cannot, by definition, have a low rate of success. Reliabilists have (for the most part) switched from talk of processes to talk of abilities, but have not divested themselves of all the trappings of process-talk. I mention this briefly now, as it will come up again in the next chapter. For the purposes of this chapter, it is but an observation of something curious.

Other epistemic externalists account of ability all seem to satisfy roughly the same criteria as well (we will subsequently discuss Sosa [2007], Millar [2009], Millar [in Pritchard et al. 2010] and Kallestrup & Pritchard [forthcoming]). To have a certain ability, a subject must usually achieve the result towards which that ability is geared, given that the environment and conditions in which she is attempting to achieve the relevant result are not patently hostile to such attempts. That the subject must only *usually* achieve the result—and not necessarily succeed in any given instance—demonstrates that these accounts all have the general sense of ‘can’ as their target.

There is a variable among these various accounts which is of interest here. Millar notes a difference between accounts in which abilities are *broadly* individuated from ones in which they are *narrowly* individuated (Pritchard et al. 2010). The issue of broadness versus narrowness is broached with respect to agents treated as bodily individuals: i.e., human beings as we usually pick them out. An account construes abilities narrowly if the bases for abilities are *totally within* bodily individuals, where environments and conditions only determine whether abilities are successful exercised. Thus, an account construes abilities broadly if the bases for abilities are *partially outside* bodily individuals, where the ability is

possessed only in environments in which it can be successfully exercised. Ernest Sosa's notion of ability is straightforwardly narrow. (As is Greco's<sup>50</sup>.)

[A] competence is a disposition, one with its basis resident in the competent agent, one that would in appropriately normal conditions ensure (or make highly likely) the success of any relevant performance issued by it (2007: 29).

This is a narrow construal of abilities because it renders them as dispositions which, even if they do not actually issue in success, would do so if embedded in appropriate environments. Hostile or otherwise unsuitable environments simply block these dispositions from successfully manifesting, rather than preventing them from being possessed. Hence, a subject has the ability to determine whether toast is burning by smell if she is disposed to form the judgment that toast is burning in response to certain (proximal) olfactory stimuli—which stimuli, in friendly environments, indicate the burning of toast. The subject's disposition to form such judgments will make more likely the success of her epistemic performances (forming beliefs about the state of the world) in friendly worlds, where success here is construed as the production of a true belief. Importantly, for Sosa *the very same ability* (*viz.*, forming a belief whether by smell) is deployed both in the scenario where the relevant (proximal) olfactory stimuli do reliably correlate with burning toast and in the scenario where such stimuli do not so correlate.

Alan Millar considers abilities in the broad sense. “The assumption here is that you cannot tell that something is G from its visual appearance—the way it looks—unless that appearance is distinctive of things that are G” (Pritchard et al. 2010: 130). In Millar's account, a way of looking is a physical feature of things in the world, rather than some phenomenological or psychological feature of a subject (*op cit* 125). Thus, whether some appearance is distinctive of objects with a certain property depends on the correlation

---

<sup>50</sup> Greco and Sosa's accounts are sufficiently similar for remaining purposes. The general considerations which apply to Sosa will also apply to Greco. I note here that I appreciate that there are significant differences between Greco's and Sosa's views with regard to the nature of the relation holding between the exercise of abilities and positive epistemic statuses. However, this difference should not be relevant in what follows.

between having the relevant property and having the relevant appearance. Millar deals with fake-barn cases using his construal of appearances.

When Barney judges falsely in fake-barn territory he fails to exercise an ability to tell of certain structures that they are barns from the way they look. Indeed, he does not have the ability to tell structures around there that they are barns from the way they look. Of course, when he is there he does something like that he also does back home [in real-barn territory]—judge of some structures that look like barns that they are barns—and in doing so he will sometimes judge correctly. But that does not amount to his being able to tell of the structures that they are barns from the way they look (*op cit* 126 - 27).

Fake-barns in fake-barn county have the same appearance as real-barns do. It is built into the case that they cannot be told apart by looking (from the roadside). Thus, according to Millar, *barn-appearance* is not sufficiently correlated with *being a barn*. More exactly, in a county where half the objects which present barn-appearance are not real-barns, such appearance is obviously not distinctive of real-barns. Moreover, if there are only real-barns in a county, barn-appearance obviously is distinctive of real-barns. As the proportion of fake-barns to real ones increases, barn-appearance becomes less distinctive of real-barn. Up to a point, according to Millar, an appearance shared between a real object and a superficial replica of that object remains distinctive of the real object despite the presence of replicas. There will be borderline cases where the actual real-to-replica ratio makes it indeterminate whether an appearance is distinctive of the real objects, and hence indeterminate whether a subject can have an ability to tell of some object that is a such-and-such from its appearing to be so.

Millar's account is clearly one on which abilities are broadly construed<sup>51</sup>. However, Millar's account of abilities differs from Sosa's in another way. Millar is concerned with abilities the success of which is gaining knowledge, rather than the formation of true beliefs (on a particular basis)<sup>52</sup>, as in Sosa's case. Does the difference between true belief and

---

<sup>51</sup> This is not a surprise: Millar says so himself after making the distinction, which I have already noted I am borrowing from him.

<sup>52</sup> "On a particular basis" will be omitted henceforth, but talk of Sosa's account of ability should be understood to implicitly include the phrase.

knowledge account for the difference between the two authors with respect to construing abilities as narrow or broad? Millar suggests otherwise.

When discussing Sosa's account, Millar (2009) expresses puzzlement regarding Sosa's explanation of why an agent manifests an ability in the kaleidoscope case. Briefly, the kaleidoscope case is one in which an agent is viewing a wall which presents a certain colour appearance. Whether the presented appearance matches the actual colour of the wall is at the whim of a trickster, who has the ability to make the wall appear to be one colour when it is actually another. Sosa makes the following claim.

The kaleidoscope perceiver does seem to exercise an ability in its appropriate conditions. These include his open-eyed alertness, the well-lit medium, the proximity and size of the unoccluded surface, etc. Plausibly, he gets it right, in believing the seen surface to be red, through the exercise of a perceptual competence in its appropriate conditions. Yet the jokester could as easily give him a bad combination (red light + white surface) as the good actual combination (white light + red surface) (2007: 99 - 100).

Millar notes that for Sosa, success is the formation of a true belief. Nevertheless, Millar contends that it is odd to claim that in the case mentioned above any ability-generating positive epistemic attitudes is manifested. Because of the trickster, poised as she is to make white walls appear red, a wall's appearing red is not sufficiently correlated with a wall's being red in the case at hand (2009: 228). But if this is the case, it seems that an agent in the kaleidoscope situation could easily make a false judgment about the colour of the walls. This false judgment is a failure irrespective of whether an account is concerned with abilities to form true beliefs or abilities to gain knowledge. (Both require a judgment to be true.) That the agent actually succeeds is unimportant, according to Millar.

A way to see what Millar is after is to reify the possibilities that the trickster's poise represents in Sosa's version of the case. Imagine a series of rooms, in some of which the trickster sets up the bad combination (red light + white surface), while setting up the good combination (white light + red surface) in others. On Millar's account, whether a subject in this situation has the ability to determine that the walls are red from their appearing so will depend on the ratio of good rooms to bad ones, because whether red-appearances are

distinctive of red walls depends on this ratio as well. For it to be the case that the subject could easily make a false judgment, it needs to be the case that in a sufficient proportion of cases she will judge that a wall is red from its appearing to be so when it appears to be red but is not. In turn, for the previous situation to obtain, a sufficient proportion of the rooms need to be bad ones (red light + white surface). Now we have replication of the fake-barn county example and Millar's point about the possession of ability can be made just the same.

All this was in the service of demonstrating that Sosa's and Millar's differences regarding whether they construe abilities narrowly or broadly does not stem from the ability-result pairs with which they are, respectively, concerned. If Millar's account is good, it applies equally to abilities to form true beliefs as it does abilities to gain knowledge. Given that the difference between Millar and Sosa does not depend on the ability-result pairs in which they are interested, is there reason to favour a Sosa-like, narrow construal of abilities over a Millar-like, broad one? Recall that from Millar's position, Sosa gets the both the possession and manifestation conditions of ability wrong.

How might we construe Millar's puzzlement in a principled way? Does Sosa run afoul of the independent, plausible analysis of ability given by Honoré? At first glance, it might seem that Sosa confuses the particular and general senses of 'can' when claiming that the subject in the kaleidoscope case manifests an ability. In Sosa's version, we say of the subject in the kaleidoscope case that she *can* form a true belief about the colour of the wall, because she *does*. This is taken to be a manifestation of ability by Sosa because, despite the trickster lurking in the background, the ambient conditions appear favourable for determining what colour the wall is by looking, and because the subject does form "[an] object-level belief [correctly attributable] to his object-level visual competence" (2007: 100). However, in the multiple-room version I suggested, the subject forms a true belief *only when in a good room*, because only in these rooms are the ambient conditions favourable. In these rooms, the subject *can* form a true belief because she *does*. The question is, in the multiple room setup, does being in one of the good rooms count as a special circumstance which would discount any success from counting towards the attribution of an ability to the subject? Recall Honoré's dinner speech example. He notes that, while the speaker can give a good speech (contrary to reasonable expectation based upon the speaker's previous performances) in

some special (combination of) circumstance(s), this is not the sense of ‘can’ which relates to ability descriptions.

Millar and Honoré would agree that actual success (in isolation) is unimportant with respect to whether an agent possesses a particular ability. Given that the kaleidoscope subject actually succeeds but—according to Millar—could easily fail, perhaps we can translate Millar’s gripe with Sosa to be that the latter confuses the general and particular senses of can. However, we also noted that Honoré took possible failure (in isolation) to be unimportant with respect to whether or not a subject possesses a particular ability. This derives from considering the golfer about to attempt a six-foot putt. *If* the golfer generally-can hole such putts (i.e., he is able to hole such putts), then that he might possibly fail does not demonstrate that he possesses no ability to hole such putts. However, whether the subject in the kaleidoscope room has the relevant ability is exactly what is at issue. Sosa assumes the kaleidoscope subject has the relevant ability and Millar denies this claim. Are we left, then, simply with Millar’s bemusement?

Is there a difference between the single-room and multiple-room setups which licenses Sosa’s claim that, in the single-room version, the subject possesses the relevant ability? Or, is there reason to think that being in a good room in the multiple-room setup does not count as a special circumstance? An obvious difference is that the possibility of the trickster affecting the bad combination (red light + white surface) in the single-room version is a *mere* possibility, one which obtains nowhere in the actual world. In this case, the modally close possibility is that the subject is (counterfactually) in a room which appears exactly like the good room in the actual world, but in which the trickster has (counterfactually) affected the bad combination. On the other hand, in the multiple-room scenario the possibility of the trickster affecting the bad combination is an *actualized* possibility, one which does obtain in the actual world. In this case, the modally close possibility is that the subject is (counterfactually) in one of the (actual) bad rooms instead of the (actual) good rooms. In each case, the relevant close possible situation is one in which the subject is in unfavourable conditions and fails to form a true belief (on the same basis that the true belief is formed in the actual case). But, in each case, what would have to change in order that the situation becomes one in which the subject forms a false belief (again, on the same basis upon which

the true belief is formed in the actual case) is different. In the single-room case, what the trickster actually does must be different. In the multiple-room case, where the subject actually is must be different. Having said this, however, it is difficult to tell what to make of it. There are no obvious reasons to take the difference I have identified between the two scenarios as significant without begging the question against either a narrow or broad view of abilities.

### §3.1: *Abilities and Epistemic Twin Earth*

A change of scenery is in order. Jesper Kallestrup and Duncan Pritchard's (manuscript) "Virtue Epistemology and Epistemic Twin Earth"<sup>53</sup>—though it is a critique of robust virtue epistemologies (of which Greco, Sosa and Millar's accounts are all examples<sup>54</sup>)—is probative with regard to the appeal to cognitive abilities in epistemology. Kallestrup & Pritchard's criteria for some subject to possess a particular ability are as follows. "S possesses [an] ability in virtue of relevant bodily/psychological factors *and* mostly occupying an environment that is conducive for [S] to frequently [succeed at the goal of the relevant ability]" [emphasis added] (7). While the criteria for ability possession do include an environmental factor, Kallestrup & Pritchard should not be seen to advance a broad conception of ability.

Talk of environments does a lot of work for Kallestrup & Pritchard. They delineate three sorts of environment with respect to a subject. A subject's local environment is where she is at a given time, "[containing] the objects and properties that are the proximate causes of her current perceptual experiences" (6). With respect to a subject, a regional environment "is neither where [she] is currently located, nor where she typically forms any beliefs", but nonetheless, "contains the objects and properties with which she might easily have been causally connected" (*ibid*). The global environment is "where [a subject] is normally located although not at present", and which contains, "the objects and properties with which she

---

<sup>53</sup> All references to Kallestrup & Pritchard are from this paper.

<sup>54</sup> I take this to be a non-standard interpretation of Millar. However, this interpretation seems to me to fall out of considering a robust virtue epistemology to be one where the relevant abilities do all of the heavy lifting. Robust virtue epistemologies are those in which there is not a separate, modal stability condition which must be met *in addition* to an ability condition for the target epistemic status to be achieved.



normally interacts” (*ibid*). Given the criteria for possession of ability Kallestrup & Pritchard have put forth, the environment upon which a subject’s possession of an ability depends is her global environment—the environment which she mostly occupies. To see why this dependence does not entail a broad conception of abilities, we need to look at the centrepiece of Kallestrup & Pritchard’s paper.

The primary example in Kallestrup & Pritchard is an epistemic twin earth case (6 – 8). We are asked to imagine a pair of subjects—one on earth and one on twin earth—each of which is a competent speaker of her respective language, but neither of which is a trained chemist. Each of the subjects also has highly reliable perceptual abilities. On earth, in each of the subject’s local, regional and global environments, all watery-stuff is  $H_2O$ ; whereas on twin earth in the twin subject’s *regional* environment much of the watery-stuff is XYZ. Because on each of earth and twin earth the global environment contains exclusively  $H_2O$ , both the subject and twin subject mostly occupy environments which are conducive to forming true water-beliefs. By hypothesis, the subjects on both earth and twin earth are bodily/psychological duplicates, each with highly reliable perceptual faculties. Thus, both the subject and twin subject possess an ability to form true beliefs about water across close possible worlds. Further, as each of the subject and twin subject’s local environments contain exclusively  $H_2O$ , their respective water-beliefs are actually true as well. However, because of the difference in the regional environment between the two subjects, Kallestrup & Pritchard claim that there is an epistemic difference between the two subjects. Particularly, the twin subject’s reliable and true water beliefs are lucky due to the fact that she could have easily been causally interacting with XYZ, as it lurks in the regional environment on twin earth. The reliable, true water-beliefs of the subject on earth are not lucky. Nowhere does XYZ lay in wait for her. The reliable, true water-beliefs formed in the stipulated local environment on twin earth are lucky and hence not knowledge; whereas reliable, true water-beliefs on earth are non-lucky and hence knowledge. Therefore, Kallestrup & Pritchard take themselves to have demonstrated that there is a case which does not differ in terms of ability, but does differ in terms of knowledge.

This example does not suggest a broad conception of ability according to Millar’s distinction because while abilities are individuated with respect to environments in

Kallestrup & Pritchard's account, they are "individuated independently of the environment in which the subject is [currently] placed" (Pritchard et al. 2010: 129). In Kallestrup & Pritchard, the environment in which the subject is placed is strictly the local environment. But this is not the environment upon which ability possession is dependent in Kallestrup & Pritchard's account. Is the possession of ability in Millar's account dependent on what Kallestrup & Pritchard call the local environment? Recall that for Millar, if it is actually the case that F-appearance is not distinctive of F's, then a subject does not have the ability to recognize F's even if she is in fact looking at an F. How should 'distinctiveness' be cashed-out in terms of environments in Kallestrup & Pritchard's sense? When Millar considers fake-barn cases, he draws a distinction between a subject's (Barney) "home territory" and "fake barn territory" (Pritchard et al. 2010: 126). With respect to the former, Barney has the ability to recognize barns. With respect to the latter, he has no such ability according to Millar.

Translating this into Kallestrup & Pritchard's terms will tell us what sort of environmental dependence Millar posits. Barney's home territory in the fake-barn cases cannot be his local environment because he is not currently located in his home territory. Nor can his home territory be his regional environment because Barney's home territory is where he usually forms beliefs. Thus, his home territory is roughly his global environment. How is fake-barn territory classified, then? The natural reading would be to assimilate fake-barn territory to Barney's local environment, because it is what is nearby. However, Kallestrup & Pritchard employ technical senses of 'local', 'regional' and 'global'. While the eponymous structures of fake-barn territory are spatially near Barney, these structures are *not* the "proximate causes of [his] current perceptual experiences", and hence not in Barney's local environment. His local environment is the genuine barn he is looking at. The only remaining choice is that fake-barn territory is roughly the Barney's regional environment. After all, on both the account of Kallestrup & Pritchard and the account of Millar, fake-barns are not what Barney is currently looking at, nor the sorts of structures which populate the places Barney usually finds himself forming beliefs. Further, on both accounts, fake-barns are structures which Barney might have easily been looking at and thus forming beliefs

about. In Kallestrup & Pritchard's terms, Millar contends that ability possession is dependent on regional environments.

According to translated Millar, if a regional environment is such that it renders performances in the local environment luckily successful, then Millar's account claims that subjects in such circumstances lack the relevant ability. Barney lacks the ability to recognize barns in fake-barn country. The twin subject on epistemic twin earth lacks the ability to discriminate between water and twin water and then form water-beliefs on that basis. While Millar would agree that, if the environment in which a subject typically forms beliefs (i.e., her global environment) is one in which F-appearance is not distinctive of F's then the subject does not have the ability to recognize F's, this is in addition to the regional environment dependency he posits.

The epistemic twin earth example is supposed to demonstrate the inadequacy of robust virtue epistemology. Whether or not the case demonstrates the inadequacy of robust virtue epistemology, it can help us adjudicate the disagreement between Sosa and Millar with respect to the kaleidoscope case. Recall that it is reasonable to interpret their disagreement as persisting in the face of both the difference in concern over the type of ability in question (for Sosa, the ability to form true beliefs; for Millar the ability to acquire knowledge) and the difference in the setup of the case (single room versus multiple room).

How would the regions in the kaleidoscope cases be delineated? In both the single and multiple room scenarios, the local environment is one in which the good colour combination (red wall + white light) is present. Thus, *if* subjects in these cases possess abilities to form the relevant true beliefs, then their beliefs will *actually* be true. Do they possess the relevant abilities? To answer this with regard to Kallestrup & Pritchard's account of ability possession, we need to know what the global environment is in each case. This is tricky. In the single room case, the fact that the trickster is poised to implement the bad colour combination (white wall + red light) is described as rendering the kaleidoscope subject's belief *unsafe*. This means that, in close possible worlds, the subject's belief is false. However, recall that for the global environment to be such that a subject possesses some ability, it must be the case that the subject's performances will be successful across relevantly close

possible worlds. Is it the case that, *pace* Sosa, the kaleidoscope subject does *not* manifest the relevant ability, as she does not possess such an ability to manifest?

This verdict issues more from the underdescription of the case rather than a flaw in Sosa's theory. After all, the assumption must be that the kaleidoscope room is an aberration in the subject's life. Before entering the room, she was on an otherwise earth-like world where (in the global environment) red-appearance indicates redness. Further, since the local environment is stipulated to be a favourable one, the only candidate left to explain the instability of the kaleidoscope subject's belief is the regional environment. On this more charitable interpretation, Sosa correctly contends that the kaleidoscope subject has the relevant ability. At least, provided we buy into Kallestrup & Pritchard's account.

What about the multiple-room case? For simplicity, we can assume that the world of the subject in the multiple-room scenario is an endless maze of bare rooms with either the good colour combination or the bad combination. If the rooms which the subject usually inhabits are good combination rooms, then all parties will agree that she has an ability—at least when in that region of rooms—to form true beliefs (or know, in Millar's case) about the colour of the walls<sup>55</sup>. Further, everyone will agree that the subject's local environment is one in which conditions are favourable for forming true beliefs about the colour of the wall. She is, in fact, in a good combination room. The difference will again come in the regional environment. If many of the adjoining (and perhaps more generally nearby) rooms have the bad combination in them, then the subject could have easily been in one of those rooms, and thus in conditions unfavourable for forming true beliefs (or gaining knowledge) about the colour of the walls. If this is the way to translate the multiple-room scenario into Kallestrup & Pritchard's region terminology, then there is no relevant difference between the single and multiple room scenarios. Thus, it seems in both cases Sosa correctly attributes an ability to the subject in question.

Millar will still disagree in both scenarios regarding the attribution of ability because his account ties ability possession to the regional environment rather than the global environment (more on this in a moment). However, bringing to bear the machinery from

---

<sup>55</sup> If this point is not stipulated, it is plausible that all parties will agree that the subject lacks the relevant ability. As such, we need not entertain such a setup.

Kallestrup & Pritchard, we can see that Sosa does not run afoul (when plausibly interpreted) of any independent notion of ability. Tying the possession of abilities to the global environments which are friendly to the general success of the relevant abilities accounts for the notion from Honoré that for someone to have a particular ability they must normally succeed in the outcome towards which that ability is aimed. This is because a subject's global environment is where she usually forms her beliefs, and because it is the case both that this environment is friendly and that she is reliable. Further, the local environment (with which the subject directly interacts) is stipulated to be friendly in both the fake-barn and kaleidoscope room scenarios. (In the former, the subject is looking at a real barn; in the latter, the subject is looking at the good colour combination.) That is, there are no special conditions or circumstances at play. No poor lighting, physical ailment, mental distraction or deception is directly at work. Thus, we have a subject who does succeed in accomplishing something which she usually succeeds at in conditions which do not nullify her success. It seems all of the criteria from Honoré for the attribution of ability are met.

What about the hostility of the regional environment? This fact is supposed, according to Kallestrup & Pritchard, to render the agent's success lucky—and lucky in an epistemically undermining way, no less. But, if an agent's success were lucky, then would not the independent account of ability from Honoré rule that the agent does not possess the relevant ability? In which case, Millar's account would seem to be vindicated.

However, ultimately it is Millar who runs afoul of our independent account of ability. Remember that, in isolation, neither success nor failure (actual or possible) indicates anything about whether an agent possesses some ability. Moreover, when possible failure looms for a performance by an agent to whom the relevant ability is already correctly attributed, the threat of possible failure provides no grounds to retract the attribution of ability to the agent. Millar's account, however, delivers the verdict that ability attributions *should be* retracted in just such cases.

If our runner, Colin, usually completes 42 kilometres in less than three hours in the environments and conditions in which he usually runs 42 kilometres, then he has the ability to run 42 kilometres in under three hours. On a particular occasion, suppose Colin runs a route through a park through which he has never run previously. The whole length of the

path Colin actually runs is paralleled on one side by a path which is all sand and a path on the other side which is all ankle-deep mud (as opposed the packed dirt of the actual path). Colin could have easily run on the sandy or muddy paths, and if he had, he would have failed to run 42 kilometres in under three hours. These alternate paths are in Colin's regional environment: they are things 1) with which he did not on the occasion in question causally interact; 2) with which he does not usually causally interact and; 3) with which he could have easily causally interacted. The packed dirt path is Colin's local environment: i.e., that with which he interacted on the occasion in question. Finally, we will assume that Colin's global environment (where he usually runs) is nothing but packed dirt roads. It is plausible—and in line with Honoré's account—to attribute to Colin an ability to run 42 kilometres in under three hours.

Does Colin lose this ability when he runs in the new park? It seems Millar's account, which ties the possession of abilities to the favourability of the regional environment, indicates that Colin does lose his ability. The difficult paths are nearby possibilities in the same respect as are fake-barns in the Barney case. That is, just as Barney might have easily stopped to look at a fake-barn, Colin might have easily turned onto a different path. This is odd. Not only is it odd, but also it is in tension with what we have taken from Honoré. An agent to whom an ability has already been correctly attributed does not lose that ability in the face of possible failure. Moreover, Millar's account says the same thing about the twin subject on twin earth. When the regional environment is friendly (i.e., there is no XYZ lurking to confound the subject's beliefs), she can discriminate water from twin water and form true water-thoughts on that basis. However, where XYZ hides nearby, the poor twin subject cannot form true water-beliefs on the basis of an ability to discriminate water from twin water. Note that no violence has been done to Millar's theory. The translation of his account into Kallestrup & Pritchard's environment terminology was plausible. Further, the epistemic twin earth and long distance runner cases are structurally identical to the fake-barn and kaleidoscope room cases.

Millar proposed a distinction between broad and narrow abilities. He further claimed that the broad conception of abilities was better equipped to deal with various cases than the narrow conception. However, our examination of the issue has suggested that the narrow

conception is more in line with an independent account of ability. That a broad conception bears better epistemological fruit (if it does) is not sufficient reason to adopt it in the face of independent considerations in favour of a rival conception.

#### § 4: A Way from Epistemic Externalism to Cognitive Externalism?

Given that the best account of ability in general for the epistemic externalist states that an agent's possession of some ability has an element of dependence on that agent's environment, we might wonder whether this commits the epistemic externalist to one theory or another about the nature of cognition. That is, does the view of abilities taken by externalist epistemologists connect with any view of ability in the philosophy of cognitive science? Perhaps an argument such as the following is good.

- (1) An epistemic externalist account of justification must appeal to cognitive abilities  
(*Premise*)
- (2) On epistemic externalist accounts, an agent's abilities depend in part on that agent's environment (*Premise*)
- (3) Cognitive externalism states that an agent's cognitive abilities depend on that agent's environment (*Definition*)
- (4) Therefore, an epistemic externalist account of ability implies cognitive externalism  
(*From 2 & 3*)
- (5: C) Therefore, epistemic externalist accounts of justification must appeal to cognitive abilities as envisioned by the cognitive externalist (*From 1 & 4*)

I will not waste much space explaining this argument, as it should be clear that it is wrong. However, I will briefly explain how it might be taken to work.

(1) is a plausible assumption about what an epistemic externalist account of justification should involve, given the considerations we reviewed in Chapter 2—*viz.*, that for a belief to be justified it must be formed in an appropriate manner. The foregoing discussion led us to conclude that all of the major externalist accounts of ability featured some form of environmental dependence. This was true of both so-called narrow and broad

accounts of ability. Hence, (2) is a plausible premise. (3) offers a rather loose definition of cognitive externalism. On some readings of (some versions of) cognitive externalism, the claim that cognitive processes are in part constituted by factors external to the bodily individual is read-off of relations of reciprocal dependency obtaining between a bodily agent and some external factor. In general, some sort of dependency on external factors is argued for by the cognitive externalist, as we saw in Chapter 1. Given (2) and (3), it might be supposed that (4) follows. If epistemic externalism posits a dependency of agent's abilities in general on that agent's environment, then the same should hold of their specifically cognitive abilities. But since cognitive externalism is (according to (3)) just the thesis that an agent's cognitive abilities depend on her environment, the epistemic externalist should endorse cognitive externalism as a simple consequence of her general view on abilities. Finally, given that the epistemic externalist must appeal to cognitive abilities in her externalist account of justification—and given that her position implies cognitive externalism—the epistemic externalist's account of justification has to appeal to cognitive abilities as construed by the cognitive externalist. That is, given (1) and (4), (5) follows.

There are obvious and fatal problems with this attempt to demonstrate a commitment to cognitive externalism on the part of the epistemic externalist. In philosophy of cognitive science, a cognitive externalist view of abilities is *broad in a constitutive sense*: i.e., abilities are broad if they are partly constituted by factors external to an agent. Likewise, a cognitive internalist view of abilities is *narrow in a constitutive sense*: i.e., abilities are narrow if they are exhaustively constituted by factors within an agent. However, in the epistemological debate, views of abilities are broad or narrow in an *individuating sense*. An ability is broad if it is possessed or individuated with respect to the subject's environment; an ability narrow if it is possessed or individuated independently of the subject's environment. What do either possession or individuation have to do with constitution? Very little, it would seem. That is, the sense of 'depend' in (2) is different than the sense of 'depend' in (3). There is an equivocation in the argument.



Consider content externalism<sup>56</sup>. While various forms of content externalism claim that which contents a subject's thoughts have depends on various factors external to that subject, the further claim that those factors are *part of* the relevant thought does not immediately follow (*contra* Wilson 2004). The fact that my environment contains trees (or the fact that I am the member of a reference-preserving causal chain originating with someone whose environment contains trees; or the fact that I am in a community where 'tree' is used to refer to trees, etc.) is what makes my thoughts containing 'tree' tree-thoughts. It does not follow that the trees (or whatever) in my environment are partly constitutive of my thoughts.

This intuition gives rise to the distinction between passive and active externalism (Clark & Chalmers 1998). The factors cited in content externalism do their thought-individuating work historically. Clark & Chalmers, on the other hand, in their arguments for the existence of extended cognitive abilities, appeal to external factors which do their work ahistorically—that is, in the here-and-now. It is the engaged, reciprocal interaction between an individual and some factor outside the individual which engenders broad cognitive processes. It is not that such processes are labelled as an  $x$ -process or a  $y$ -process because of these external factors—instead, it is that the external factors are part of the process itself. These factors are actively contributing to the accomplishment of some cognitive task.

The environmental factors cited as relevant to the possession or individuation of abilities in the epistemic literature, however, are each passive. Even in Millar's account, the environmental factors with which a subject is actually interacting are not the ones which determine whether she has an ability, or what ability it is. In the epistemological case, it is the local environment with which a subject is actively engaged; whereas it is the global (or, in Millar's case, the global and regional) environment which is counted towards the individuation or possession of abilities. Thus, we should conclude that 'broad' and 'narrow' have the same sense in the epistemological case of individuating abilities as those terms have in the philosophy of mind case dealing with the individuation of mental contents.

---

<sup>56</sup> I limit myself to Putnam-Burge style content externalism here. Other theories of content which give individuating power to external factors may not count as passive. Here, a factor is passive when its contribution is historical. Consider accounts of demonstrative thoughts like Kaplan's (1989). Here the object referred to here-and-now individuates the content of demonstrative thoughts. Thanks to Jesper Kallestrup for pointing this out.

Given that this is the case, there is no direct argument from the notion of ability deployed in the externalist epistemological literature to a commitment on such theorists' part to some form of cognitive externalism. If I am to make a case for such a commitment, I will need to bring more considerations to bear.

## Chapter 4

### From Epistemic Externalism to Cognitive Externalism

#### *Introduction*

In the previous chapter, we noted that there was no way to argue directly from the account of ability used by externalist epistemologists to a commitment on their part to any version of cognitive externalism. This was because the environmental dependencies invoked in both the narrow and broad epistemological accounts of ability were passive dependencies, which were *individuating* of abilities rather than *constitutive* of them. However, the difference between broad and narrow abilities in the debate in philosophy of cognitive science was regarding whether abilities are broadly or narrowly constituted. We harkened back to the well-trodden debates concerning content externalism, recalling that it does not follow from some factor's individuating role that it has a constitutive role. Hence, any attempt at direct argument for a commitment to broadly constituted abilities on the part of the epistemic externalist was blocked.

There is more to be said. A case can still be made for the epistemic externalist's commitment to some form of cognitive externalism. In this chapter, the question will adduce additional considerations. Before, we only considered whether the account of ability given by

the epistemic externalist committed her to cognitive externalism. Now we will ask whether stipulating the truth of epistemic externalist positions (which rely on an ability condition) entails cognitive externalism. This differs from the method of the last chapter because by stipulating epistemic externalism, we must rule out all forms of epistemic internalism, and any positions regarding the nature of abilities wedded to epistemic internalism. First, we will need to recall the confirmation conditions set out in Chapters 1 and 2.

*(Cognitive Externalism Confirmation Condition)* There is at least one actual instance of a cognitive phenomenon the constitutive machinery of which:

- a) Extends beyond the brain (and nervous system) of an individual
- And
- b) Extends at least beyond the non-neural portions of an individual which directly subserve, or interface with, an individual's brain (and nervous system)

*(Negative Epistemic Externalism Confirmation Condition)* There is at least one possible instance of justified belief for some subject the basis for which is:

- a) Not accessible to the relevant subject, in principle, by reflection alone
- And
- b) Not part of the relevant subject's mental life (construed as internal)

*(Positive Epistemic Externalism Confirmation Criteria)* All possible instances of epistemic justification are instances of beliefs arrived at via the exercise of reliable, cognitive abilities

Clearly, a confirming instance of negative epistemic externalism will not necessarily count as a confirming instance of cognitive externalism. Negative epistemic externalism is true just in case epistemic internalism is false. Because internalism is a thesis which attempts to articulate a necessary truth about the nature of justification, it is false if possibly false. Hence, negative epistemic externalism is true if possibly true: i.e., if there is at least one possible instance of justification which violates epistemic internalist criteria. However, the

forms of epistemic externalism which I will argue accrue a commitment to cognitive externalism are positive theses, which (I take it) aim to capture the way things are *at least* in the actual world. Hence, while a confirming instance of negative epistemic externalism is not clearly in the running for also being a confirming instance of cognitive externalism, a confirming instance of positive epistemic externalism clearly is in the running to be a confirming instance of cognitive externalism. But does the confirmation of positive epistemic externalism also count as the confirmation of cognitive externalism? I will argue that it does.

The motivation for asking this question stems from the deployment of the term ‘cognitive’ in epistemology. Recall that basic reliabilism faced two primary problems—*viz.*, the direction-of-fit problem and the generality problem. The first problem—which concerned whether epistemic success was properly attributable to relevant subjects—was solved by appealing to subjects’ abilities. However, the solution for the direction-of-fit problem left the generality problem untouched. Abilities of any grain can be specified for a given situation, and which grain of ability one chooses will determine whether the cited ability is sufficiently reliable for the attribution of some positive epistemic status. The solution to this further problem, I suggested, is to appeal to a specific class of ability—*viz.*, cognitive ability.

If deploying the term ‘cognitive’ is to do any real work for the externalist epistemologist, she must deploy it in a way which respects the development of the concept in the empirical and philosophical discourse devoted to it; else, she must develop an *independent* account of cognition that goes along with her account of justification<sup>57</sup>. The reason for this is twofold. First, the epistemologist’s use of ‘cognitive’ cannot have the content, “whatever gets me out of this theoretical jam”. This is just to say that the manoeuvre cannot be *ad hoc*. Second, if the account she adopts is not properly sensitive to

---

<sup>57</sup> More likely, an epistemologist should deploy the concept as it is delineated in the sciences devoted to it. This has been the position of philosophers with epistemological views as diverse as Goldman (1986) and Conee (2007). According to Goldman, “The strict role of primary epistemology is to borrow results from cognitive science and assess repercussions of these results” (182). We have already seen, in chapter 2, Conee’s contention that philosophical theories should ideally not be overburdened with extraneous empirical content.

the empirical and philosophical issues concerning cognition which have been identified by those theorists directly working on such topics, the externalist epistemologist's appeal to specifically cognitive abilities is not likely to do the work necessary to solve the generality problem. That is, the abilities taken as paradigmatically cognitive by (philosophers of) cognitive science seem to be of precisely the right grain *and* reliability level the epistemic externalist needs them to be. Human beings get around just fine.

No externalist epistemologist (of whom I am aware) gives a full account of cognition along with her account of justification (or whatever positive epistemic status she happens to be concerned with). Instead, epistemologists tend to give a general characterisation of the abilities as those for discriminating, gathering, processing, remembering and deploying various pieces of information. Thankfully, this is more or less how such abilities are characterised in (philosophy of) cognitive science. What remains to be seen is with which sort of (philosophy of) cognitive science the externalist epistemologist has most in common.

Why might we think that an externalist epistemologist is committed to cognitive externalism? Why cannot she think that cognition is a brain-bound or organism-bound phenomenon? After all, prominent externalists have claimed just that. I claim that if the externalist epistemologist holds cognition to be what Adams & Aizawa and Rupert maintain, then her position will collapse into epistemic internalism. Specifically, an externalist epistemology which operates with a brain-bound or organism-bound notion of cognition will collapse into a form of internalism consistent with mentalism. However, considering that any epistemic theory consistent with mentalism must be an internalist epistemic theory, this will be a bad result for the epistemic externalist.

Now, I must note that there is no real barrier to the accessibilist and the epistemic externalist having the same notion of cognition. Philosophers on each side could agree what the nature of cognition is. The real disagreement between them regards *which cognitive processes* (however these may be delineated) *are necessary for justification*. Accessibilists claim that (possibly) reflection is necessary for justification while epistemic externalists deny this. The accessibilist says that what is relevantly internal for justification are those facts which a subject has reflective access to, where reflective access is "*a priori* reasoning, introspective awareness of one's own mental states, and one's memory of knowledge acquired in those

ways” (Pryor 2001: 104). Cognitive externalism could be true in virtue of some cognitive process other than reflective ones, and the epistemic internalist and externalist could go on disagreeing about justification. Note that no problems arise from arguments like Gibbons’ (2006) which claim that *some sort* of accessibility intuition can be satisfied by an externalist theory. Such arguments constitute a denial that it is specifically *reflective* access which is relevant to justification. If we do insist upon reflective access, as Pryor has defined it, Gibbons (and also Vahid 2002 and 2011) would concede that *that* notion of access is internalist, but proceed to argue that the internal notion of access is not obligatory, and does not have a monopoly on satisfying certain intuitions about epistemic perspectives.

But what if the confirming case of cognitive externalism is *a priori* reasoning, introspection or memory of beliefs formed by either of these mechanisms? That is, what if one of these cognitive processes is in part constituted by factors outside bodily individuals? A variant of Clark & Chalmers’ (1998) Otto example can serve to illustrate the worry.

Suppose Otto’s reasoning capabilities and introspective capacity are normal. (This may not be typical of Alzheimer’s disease, which Otto is stipulated to suffer from. Instead, imagine the case of a subject who cannot not form new long-term memories, as in the case of Guy Pearce’s character Lenny from *Memento* [Clark 2006].) Now, Otto reasons to some conclusion via *a priori* methods (say, that all female foxes are vixens). Because he knows that he will not remember this if he does not write it down, Otto records in his notebook, “All female foxes are vixens”. Later, someone points out a fox to Otto and says, “That’s a female fox”, whereupon Otto consults the fox-entry in his notebook and forms the belief that the creature pointed out to him is a vixen. Essential to the formation of his belief that the creature pointed out to him is a vixen is the entry in the notebook. Now suppose a form of accessibilism which demands that for a belief to be justified a subject must actually reflectively access the facts which are essential to her belief. Further, suppose that, in accordance with Clark & Chalmers, Otto’s looking-up of the fox-entry in his notebook counts as remembering. The relevant piece of information in the fox-entry (*viz.*, that all female foxes are vixens) was arrived at by *a priori* reasoning, which was then accessed by

remembrance. Hence, we have a belief which is justified by accessibilist criteria, but which involves a factor plainly external to the bodily individual under consideration<sup>58</sup>.

Should the accessibilist be worried about this result? It might seem that something has gone wrong if things like notebooks get to count as internal. However, this is only if one takes ‘internal’ in its spatial sense. One might have thought that, as a matter of contingent fact, all reflective cognitive processes were internal in the spatial sense. However, the following is not a plausible reconstruction of the accessibilist’s thoughts.

- (i) Every factor relevant to a subject S’s justification for her beliefs at *t* is spatially internal to S at *t*.
- (ii) Every factor justificationally relevant to S’s beliefs at *t* spatially internal to S at *t* is (in principle) reflectively accessible to S at *t*.
- (iii) Therefore, every factor justificationally relevant to S at *t* is (in principle) reflectively accessible to S at *t*.

Instead, the accessibilist thinks in the following way.

- (i\*) Every factor relevant to a subject S’s justification for her beliefs at *t* is (in principle) reflectively accessible to S at *t*.
- (ii\*) Every factor (in principle) reflectively accessible to S at *t* is a mental state of S at *t*.
- (iii\*) Therefore, every factor relevant to S’s justification for her beliefs at *t* is a mental state of S at *t*.

---

<sup>58</sup> Depending upon what sort of cognitive externalism is brought to bear, it might be more plausible to construe *a priori* reasoning or introspection in a cognitive externalist fashion. Briefly, a cognitive externalist might give *a priori* reasoning an externalist treatment by claiming that it depends upon certain symbolic systems of social practices. Introspection might be made cognitive externalist-friendly by claiming that when we introspect our mental states (e.g., the visual experience of a stack of books in front of me), we actually attend to the objects which those states represent (e.g., the stack of books in front of me). We might call such a view “introspection as outrospection”. See Michael Tye (2009) for an example of such a view.



(i\*) is just a restatement of Pryor's simple internalism. (ii\*) is a plausible further assumption. Clearly, (iii\*) follows from (i\*) and (ii\*). Notice that it would be a further step which links epistemic internality to spatial internality via the contention that mental states are spatially internal to subjects. But clearly, this is an unnecessary step for the accessibilist's purposes. That is, the spatial internality of justificationaly relevant factors is not explanatorily relevant to the accessibilist. Should spatial and epistemic internality coincide, this would be an interesting coincidence to the accessibilist—but that is all. Neither of the starred premises nor the starred conclusion have anything in them which suggests even *prima facie* tension with cognitive externalism. What does all the explanatory work for the accessibilist is that the subject can be in some specific kind of cognitive relation with the facts which justify her beliefs. We have already discussed, in Chapter 2, the accessibilist's reasons for placing these access restraints.

Because the sense of 'internal' in the debate between accessibilists and epistemic externalists does not impinge upon the nature of cognitive processes, the externalist does not commit herself to any claims about the location of factors constitutive of cognitive abilities by denying accessibilism. Of course, if no such commitment is engendered by rejecting accessibilism, then no commitment to cognitive externalism can be derived here<sup>59</sup>.

It is the conflict with mentalism which forces the externalist to reject cognitive internalism. Consider what the epistemic externalist would be appealing to were she to invoke the senses as cognitive abilities, while treating cognition within the cognitive internalist framework. On Adams & Aizawa's intracranialist cognitive internalism, vision, audition, olfaction, etc. could not be constituted by either any non-neural parts of the body (e.g., the vitreous and aqueous humours of the eye, the tympanic membrane), or the active use of—or engagement with the world by—the senses (e.g., looking around, manipulating, etc.). Rupert's organism-centred cognitive internalism could allow that some genuine cognition goes on in (a limited amount of) non-neural portions of the body, but would

---

<sup>59</sup> Given the way I present accessibilism (in chapter 2) as being entailed by—but not entailing—mentalism, the epistemic externalist will reject accessibilism by rejecting mentalism. Hence, any commitments which stem from the rejection of mentalism will be accompanied by a rejection of accessibilism. However, these latter commitments will not arise *because* of the rejection of accessibilism.

disallow that the active use of—or engagement with the world by—the senses constitutes cognition. Hence, for the intracranialist, olfaction would be the brain areas and nerve tissue which carry and process olfactory data. (Likewise for the other senses.) For the organism-centred theorist, olfaction would be something like the suite of inner resources tied to a legitimate cognitive explanandum by a relation of sufficiently high conditional probability (which is going to end up delivering something more or less like the picture Adams and Aizawa paint, in this case). Thus, the epistemic externalist's appeal to cognitive/intellectual abilities would advert only to facts internal to the subject in question which are relevant to cognitive/mental phenomena. But this is exactly the mentalist's notion of what is relevantly internal for justification: she appeals to sensory states, beliefs, experiences or what have-you, where these are taken to be internal to the subject in question. As such, if the epistemic externalist wishes to remain distinct from the mentalist, the former cannot appeal to the cognitive internalist notion of cognitive abilities.

*§1: A Way from Epistemic Externalism to Cognitive Externalism—Redux*

In the introduction to this chapter, I gave a brief sketch of a second argument to the effect that epistemic externalists are committed to cognitive externalism. The basic idea was that, as a consequence of her obligation to deny epistemic internalism, the epistemic externalist's appeal to cognitive abilities must take a specific form—*viz.*, an appeal to cognitive abilities, as they would be described by the cognitive externalist. As this argument is long and complex, let us go through it slowly. I will present the premises of the argument in groups.

*§1.1: Premise Group 1*

The first group is as follows.

- (1) Epistemic externalism entails the rejection of epistemic internalism.
- (2) Mentalism is a form of epistemic internalism.
- (3) Therefore, epistemic externalism entails the rejection of mentalism.

(1) is a definition. Frequently, epistemic externalism is negatively defined as the denial of epistemic internalism (e.g., Gibbons 2006; Brown 2007; Goldberg 2007). Conee & Feldman (2004) and Conee (2007) argue that mentalism is the simplest formulation of epistemic internalism. We have briefly examined their arguments to this effect in Chapter 2, and will do so again in Chapter 5. For the time being, I will simply concede the argument to Conee & Feldman, and treat (2) as an assumption. From (1) and (2), the straightforward conclusion is (3). If the epistemic externalist must deny epistemic internalism, and if mentalism is a form of epistemic internalism, then the epistemic externalist must deny mentalism.

### *§1.2: Premise Group 2*

Now we move on to the second group of premises.

- (4) An account of justification which appeals to cognitive abilities, where those abilities are exhaustively constituted by factors internal to the individual of whom said abilities are predicated, is consistent with mentalism.
- (5) Cognitive internalism is the claim that cognitive abilities are exhaustively constituted by factors internal to the individual of whom said abilities are predicated.
- (6) Therefore, an account of justification which appeals to cognitive abilities and is consistent with consistent with mentalism implies cognitive internalism.

The claim in (4) is that there is a possible internalist position which could appeal to cognitive abilities to do the same sort of work that the epistemic externalist appeals to cognitive abilities to do. That is, there is a theory of justification (perhaps merely in logical space) which both appeals to cognitive abilities and is consistent with mentalism. What would such a theory look like? And why would an epistemic internalist appeal to cognitive abilities? As we saw in Chapter 2, epistemic internalists and externalists alike are under the jurisdiction of the anti-luck intuition. In the case of reliabilism, the appeal to cognitive abilities is the way

we saw epistemic externalists respond to the direction-of-fit problem with the unadorned version of their theory. Further, this was seen as a way to satisfy the anti-luck intuition. In the case of knowledge, if a subject's true belief is formed due to the exercise of an (reliable, integrated) ability of hers, then that belief is creditable to her rather than to luck, and hence is knowledgeable. With justification, if a subject's belief (true or not) is formed due to the exercise of an (reliable, integrated) ability of hers, then that belief is appropriately formed and hence justified.

An internalist can say essentially the same thing. Zagzebski (1996), for example, defines justified beliefs as, “[I]nstances in which believers have and exhibit intellectual virtue in the process of belief formation . . .” (293). While Zagzebski indicates that her theory of knowledge is an internalist-externalist hybrid, the account of justification it contains is internalist. For Zagzebski, possessing an intellectual virtue is being motivated by the goal for knowledge, understanding or (more generally) “cognitive contact with reality” (1996: §4.1), where ‘motivation’ is understood as something which is “at least weakly cognitively accessible” (*op cit* 299)<sup>60</sup>.

Talk of virtues may be translatable into talk about cognitive abilities. Indeed, our externalist exemplar, Greco, often talks about cognitive abilities as virtues or excellences. While Zagzebski, with her strong Aristotelian sensibilities, explicitly rejects such an elision, an internalist need not follow her in this. I assume, then, that we can adjust Zagzebski's account of justification into a more general accessibilist principle.

*(Ability Accessibilism)* A subject *S* is justified in her belief that *p* at *t* if and only if *S*'s belief that *p* at *t* is formed on the basis of cognitive abilities which *S* can (in principle) reflectively determine that *S* utilized to form *S*'s belief that *p* at *t*.

---

<sup>60</sup> I am not foisting the identification of “weakly cognitively accessible” and “internalist” on Zagzebski. The context of the short excerpt above is as follows.

The accounts I have given of knowledge and justified belief are derived from an account of intellectual virtue, and since intellectual virtue has an element which is at least weakly cognitively accessible (the component of motivation) as well as an element that may be inaccessible (the component of success in reaching the end of the virtuous motivation), my theory combines internalist and externalist elements (*idem*).

This principle is compatible with the simple internalism of Pryor because it does not require that a subject have any view about the justificatory potency of the abilities she deploys.

However, if such a definition seems too weak, it can easily be amended to resemble access internalism more than simple internalism. For our purposes, what is important is that an epistemic internalist can appeal to cognitive abilities to do the same theoretical work as the epistemic externalist—*viz.*, to give a theory of epistemic justification.

What about mentalism? There are no extant mentalism-like positions which make an appeal to abilities (Conee & Feldman's mentalism is the only variety on offer). However, we can imagine that if an epistemic internalist were to make an appeal to cognitive abilities consistent with mentalism, she would have to do so within the strictures of the supervenience thesis characteristic of mentalism (i.e., (S), from Chapter 2 of this work, p. 32). Mentalism cannot countenance the justificatory potency of factors external to a subject's mental states, broadly construed. That is, the cognitive abilities to which an account of justification consistent with mentalism would appeal would have to be ones constituted by factors circumscribed by the boundaries of the individual at issue. In the case of Conee & Feldman, this clearly means the boundary of the skin at the outmost.

(5) is a definition. Cognitive internalism is nothing more than the claim that all of the factors which constitute cognitive phenomenon are circumscribed by the boundary of an organism (i.e., the skin), or indeed "further in" (e.g., the central nervous system). (6) follows from (4) and (5), on the further assumption that the mentalist is stuck with Conee & Feldman's adoption of the skin as the boundary of the individual. (We will question this further assumption at a later point.) Hence, a mentalist appeal to cognitive abilities is an appeal to cognitive abilities as construed by the cognitive internalist.

### *§1.3: Premise Group 3*

Now we move on to the third group.

- (7) An epistemic externalist account of justification must appeal to cognitive abilities.

- (8) If an account of justification which appeals to cognitive abilities and is consistent with mentalism implies cognitive internalism, then epistemic externalism must reject cognitive internalism.
- (9) Epistemic externalism must reject cognitive internalism.
- (10) The rejection of cognitive internalism entails the acceptance of cognitive externalism.
- (11) Therefore, epistemic externalism implies cognitive externalism.

(7) requires some argument. In response to the problems which arose for the basic externalist position, reliabilism, I noted only the line of response which culminates in an appeal to cognitive abilities. However, there have also been attempts to formulate externalist epistemology with only modal or anti-luck conditions. However, Pritchard (forthcoming) and Pritchard (in Pritchard et al. [2010]) argue that once anti-luck and modal conditions are adjusted to deal with counterexamples, they become tantamount to ability conditions. I will develop no novel arguments for (7) here, instead assuming that the arguments of the aforementioned authors are good. If so, then the only viable way to improve generic reliabilism involves adopting an ability condition.

Obviously, the appeal to cognitive abilities must be epistemically relevant. That is the appeal must be made to satisfy *epistemic intuitions about positive epistemic statuses*. In our case, the intuition (identified in Chapter 2) is that a (true or false) belief is not a justified belief unless it is formed the right way. Because this intuition concerns when beliefs are (epistemically) justified, it is an intuition about a positive epistemic status. However, even if it were formed incorrectly, a belief could be evaluated positively along (for instance) a moral dimension. Suppose Fred believes, as the result of a dart toss, that genocide is wrong. From a moral perspective, this belief will be appropriate—perhaps because the belief will lead to greater happiness for a greater number of people, or because it is what the moral saint would believe, etc. However, from an epistemic perspective, Fred's belief about genocide will be

inappropriate even if true<sup>61</sup>. Whether this is because truth is the goal of beliefs and dart tossing is not a good way to achieve truth, or simply because it is a brute fact that a justified belief must be formed in accordance with certain rules (etc.), Fred's belief will be faulty. Thus, not only is this an intuition about a positive epistemic status (*viz.*, justification), it is an intuition driven by epistemic considerations (rather than moral or practical ones, for instance). In summary, (7) is the claim that epistemic externalism has to invoke cognitive abilities in order to do the theoretical work it sets out for itself.

The motivation for (8) is not that the epistemic externalist must be against whatever the epistemic internalist advocates. Nor is it that, generally speaking, someone inhabiting an externalist position with regard to some epistemic status must be against whatever an internalist about some (possibly different) epistemic status advocates. Obviously, epistemic internalists should be able to agree with their epistemic counterparts about (for instance) the address of the local Mexican restaurant, or which of nihilism or universalism regarding composition is true (if indeed either position is true). Further, as we noted in the introduction to Chapter 2, being an externalist about one epistemic status need not commit one to externalism about a different epistemic status. (Likewise for being an internalist about some epistemic status.) The context of my argument is a debate between internalists and externalists about a particular epistemic status, construed in a particular way. Recall that in Chapter 2 we saw an argument from Greco to the effect that it was possible to frame the debate about epistemic justification in such a way that an internalist and an externalist could have a substantive and interesting disagreement. Given that the internalist and externalist at issue here are at loggerheads over the very same epistemic status, one party cannot accept anything consistent with what the opposed party accepts *with respect to justification*. Thus, if an account of justification which appealed to abilities delineated in the cognitive internalist mode is consistent with mentalism, the epistemic externalist must reject such an appeal because it pertains to the epistemic status at issue between her and her mentalist counterpart—*viz.*, justification. (8), then, is just an instance of (3) on the assumption of (7).

---

<sup>61</sup> It might be the case that Fred's belief is epistemically appropriate for other reasons as well. For instance, believing that genocide is wrong may not cohere with other beliefs that he has. What is important here is that Fred's belief is inappropriate specifically with respect to epistemic justification.

Given (6) and (8), (9) follows. If the mentalist can appeal to cognitive abilities to do the same theoretical work as the epistemic externalist appeals to cognitive abilities to do, then the two theorists' respective notions of cognitive abilities must be different. This is an instance of the more general fact that the epistemic externalist cannot appeal to the same factors to do the same theoretical work as her internalist counterpart. Else, the epistemic externalist's theory will not differ from the internalist's theory. But the epistemic externalist must differ—and significantly so—from the epistemic internalist by definition. In the situation at hand, this means repudiating cognitive internalism.

(10) is a definition in the same way as (1) is. Cognitive externalism is partly defined as the denial of cognitive internalism. That is, the cognitive externalist simply denies that the factors constitutive of cognition are circumscribed by the boundary of the skin. Taking this definition together with (9), we arrive at (11)—*viz.*, that the epistemic externalist must accept cognitive externalism. This is because the epistemic externalist is forced to repudiate cognitive internalism as an instance of her obligation to deny mentalism (which is, in turn, an instance of her obligation to reject epistemic internalism in general). However, cognitive externalism just *is* the denial of cognitive internalism. Thus, by definition, the epistemic externalist's commitment to cognitive externalism ensues.

#### *§1.4: Conclusion and Summary*

Finally, we arrive at the conclusion.

(12) Therefore, an epistemic externalist account of justification must appeal to cognitive abilities as envisioned by cognitive externalism.

This is the conclusion, rather than (11), because I aim to establish that the epistemic externalist must adopt cognitive externalism to do the theoretical work she wants with respect to justification. (12) follows from (11) and (7). The epistemic externalist needs to appeal to cognitive abilities to get her account of justification working. That is the point of (7). Further, because the epistemic externalist is obliged to deny accounts of justification compatible with the mentalist's account, she cannot endorse an account of justification



which appeals to cognitive abilities as envisioned by cognitive internalism. That is what (11) says. Putting these two claims together, we see that the epistemic externalist must take up cognitive externalism in order to give an account of justification which relies on an ability condition. Hence the conclusion, (12). Here is the argument, then, fully put together.

(1) Epistemic externalism entails the rejection of epistemic internalism.

*[Definition]*

(2) Mentalism is a form of epistemic internalism. *[Premise]*

(3) Therefore, epistemic externalism entails the rejection of mentalism. *[From 1 & 2]*

(4) An account of justification which appeals to cognitive abilities, where those abilities are exhaustively constituted by factors internal to the individual of whom said abilities are predicated, is consistent with mentalism. *[Premise]*

(5) Cognitive internalism is the claim that cognitive abilities are exhausted by factors internal to the individual of whom said abilities are predicated.

*[Definition]*

(6) Therefore, an account of justification which appeals to cognitive abilities and is consistent with mentalism implies cognitive internalism. *[From 4 & 5]*

(7) An epistemic externalist account of justification must appeal to cognitive abilities. *[Premise]*

(8) If an account of justification which appeals to cognitive abilities and is consistent with mentalism implies cognitive internalism, then epistemic externalism must reject cognitive internalism. *[From 3 and 7]*

(9) Therefore, epistemic externalism must reject cognitive internalism. *[From 6 & 8]*

(10) The rejection of cognitive internalism entails the acceptance of cognitive externalism. *[Definition]*

(11) Therefore, epistemic externalism implies cognitive externalism. *[From 9 & 10]*

(12: C) Therefore, an epistemic externalist account of justification must appeal to cognitive abilities as envisioned by cognitive externalism. [From 7 & 11]

Where does this leave us with respect to the rephrased question I posed at the beginning of this section? We wanted to know whether the confirmation of positive epistemic externalism also counted as the confirmation of cognitive externalism. Recall that positive epistemic externalism is confirmed when all possible instances of epistemic justification are beliefs arrived at by the exercise of a reliable, cognitive ability. For this to be a confirmation of cognitive externalism, at least one of the reliable, cognitive abilities issuing in a belief would need to be in part constituted by factors outside of the bodily individual of the relevant believer. Above I have argued that the epistemic externalist (who invokes an ability condition) cannot treat of cognitive abilities in a particular way—*viz.*, the way a mentalist would. Because the mentalist would treat cognitive abilities as being wholly constituted by factors within the bodily individual, the epistemic externalist cannot treat them that way. Thus, the epistemic externalist must, at least some of the time, invoke abilities which are in part constituted by factors outside of the bodily individual. But if she does so even once, the epistemic externalist is committed to the truth of cognitive externalism. The answer, then, is that the confirmation of positive epistemic externalism does also count as the confirmation of cognitive externalism.

### §2: *Objections and Replies*

There are a number of possible objections to the argument I have presented. I will attempt to delineate and respond to some of these objections here. Chapter 5 will go into greater depth considering the more serious (and independently interesting) objection which I will identify. There will be one objection concerning the soundness of the argument, and two concerning its validity. The first objection concerning the argument's validity centres around premise (8). This objection claims that treating (8) as an instance of (3) is illicit. The second objection concerning the argument's validity is made with regard to (4). This objection claims that, in order for (4) to do the work I need, a further assumption is actually needed

and as such the argument as stated is invalid. I will defend the validity of the argument first. Then I will proceed to outline the objection to the argument's soundness via rejecting (2). The objection to (2) will be briefly delineated here, and then treated of at some length in Chapter 5.

Before moving on, I should note that a further objection might come in the form of a denial of (7). That is, it might be denied that the epistemic externalist must make an appeal to cognitive abilities in her account of justification. As stated in Chapter 2, I will assume that other authors have successfully defended the claim that purely modal or anti-luck externalist strategies fail<sup>62</sup>. Hence, I will not mount a further defence of (7)

### §2.1: *Objection to (8)*

The objection concerning (8) is likely to go as follows. Epistemic externalism does not have to reject cognitive internalism even if mentalism implies cognitive internalism because any such implication is irrelevant. This is so even if the epistemic externalist must make an appeal to cognitive abilities in her account of justification (which we assume by invoking [7]). After all, what is at issue between the epistemic externalist and the mentalist is not whether cognitive abilities are constituted broadly or narrowly, but whether things other than the abilities themselves contribute to justification. Recall the discussion of ability in Chapter 3. We concluded that the possession of abilities was dependent on a subject's global environment—*viz.*, the environment in which she forms most of her beliefs. That there was no direct argument from this sort of environmental dependence to a commitment to cognitive externalism meant that the account of ability on offer by externalist epistemologists was neutral between broadly and narrowly constituted abilities. Moreover, the epistemic externalist's rejection of mentalism involves denying that only factors within a subject's mental life are relevant to justification. Concerning Fred's belief that the toast is burning, the mentalist claims that it is justified only in virtue of what is actually going on with Fred, cognitively speaking. The epistemic externalist, on the other hand, says that Fred's belief that the toast is burning is justified in virtue of what's actually going on with Fred, cognitively speaking, *and* how Fred's cognitive goings-on relate to the world. The mentalist claims, and

the epistemic externalist denies, that Vat-Fred is justified (to the same extent) in believing that the toast is burning as Fred is. To screen off worries from content externalism, it is stipulated that Vat-Fred was until recently not envatted. That is, Vat-Fred and Fred share the same global environment. The externalist claims that Vat-Fred is failing epistemically due to his peculiar local environment. However, as we have noted, actual failure of an occasion provides no grounds to revoke ascriptions of a particular ability to an agent of whom that ability has been appropriately predicated previously. So, in fact, the mentalist and the epistemic externalist exactly agree concerning the possession of abilities. They could even agree that a belief is justified if (and only if) it is the product of the exercise of a cognitive ability. In this case, the mentalist and the epistemic externalist would disagree on when an ability is exercised. The mentalist would claim, and the epistemic externalist would deny, that abilities could be exercised in unfavourable local environments.

This agreement is an illusion. Recall the aside from Chapter 3 concerning the notion of a reliable ability. I noted that, given the accounts of ability on offer from epistemic externalists, reliability was built into the notion of an ability. Hence, any talk of reliability on top of ability is redundant. But, if this is the case, then it cannot be that the mentalist and the epistemic externalist agree on the nature of abilities as the mentalist cannot countenance any reliability constraints in her theory of justification. Thus, the difference between the mentalist and the epistemic externalist is not merely the appeal to reliability statistics.

Nevertheless, my reply may be met with a shrug. All I have shown is that the mentalist and the epistemic externalist necessarily differ on the nature of abilities (or, more cynically, that to which the label “ability” ought to be applied). However, nothing I have said so far demands that the way in which they must differ with regard to the nature of ability is the same way in which the cognitive internalist and the cognitive externalist differ. The difference between the mentalist’s notion of ability and the epistemic externalist’s notion is just that the former notion does not include reliability and the latter notion does. But why should including reliability in the notion of ability imply that abilities are the way the cognitive externalist would have it?

---

<sup>62</sup> See footnote 35 of this work.

What is really at issue is what relation needs to obtain between a subject and things outside the subject for a positive epistemic status to arise. Recall that in the Kallestrup & Pritchard paper, though so-called robust virtue epistemologists and modest virtue epistemologists are portrayed to agree on the possession conditions for ability, there is supposed to be a difference between them. In terms of dividing explanatory work, the robust virtue epistemologist assigns all the important explanatory work to the ability itself, whereas the modest virtue theorist will allow some work to be done by an independent condition (e.g., an anti-luck conditions). In the modest theorist's case, she can say that there is a friendliness relation between the subject and the environment, such that when this relation obtains the subject's abilities will yield beliefs with positive epistemic statuses. In fact, whether the friendliness relation obtains is what explains the epistemic difference between the two subjects in Kallestrup & Pritchard's epistemic twin earth case.

Recall that the case is set up so that there is a subject on earth, whose local, regional and global environments all contain only H<sub>2</sub>O. The subject's twin on twin earth is physically the same, but differs with respect to her regional environment, in which there is some XYZ. Because the twin subject's regional environment (with which she could have easily been causally interacting) is unfriendly to the formation of true water-beliefs based on the ability to discriminate water from twin water, that she forms true water-beliefs about local samples of watery-stuff is lucky, and therefore not knowledgeable. The earth subject, however, is subject to no such luck and so does form knowledgeable water-beliefs. However, the robust theorist cannot explain the epistemic difference between the two subjects because she does not appeal to the environmental-friendliness relation in addition to the ability condition. Moreover, she cannot claim, as Alan Millar does, that the twin subject simple lacks the relevant ability. After all, we concluded that the robust theorist (of the Greco and Sosa variety<sup>63</sup>) agreed with the modest theorist (like Kallestrup & Pritchard) about the possession

---

<sup>63</sup> In replies to criticisms of his theory, Greco appears to take something like Millar's line in denying that the subject has an ability. In Greco (2007) he makes such a claim in he context of a case of sabotage at a newspaper. These saboteurs replace some true claim ( $p$ ) with a false one ( $\text{not-}p$ ), which the editors do not catch until many papers have been printed. Upon discovering and correcting the mistake, the editors print new copies to put into circulation with the sabotaged ones. Now, suppose Fred reads a corrected newspaper. Does he know that  $p$  on the basis of reading the newspaper? Intuitively Fred does not. But how can a robust virtue theorist like Greco account for this,

of abilities. And even were that not the case, she should not pursue Millar's route and hitch the individuation of abilities to regional environments, as it was concluded that doing so ran afoul of our independent account of ability. Millar's motivation in this case was the looming possibilities of error. If these possibilities were sufficient, Millar suggested that a subject actually did not have the ability at all. However, possible failure was no reason to revoke a previous, proper attribution of ability.

At this point, the robust theorist's options are unclear. She needs the ability condition to do all the heavy epistemic lifting in her theory and she needs her appeal to abilities to be distinct from a mentalist appeal to cognitive ability. Reliability is already built into her notion of ability, and it is difficult to explain how this works. She will not appeal to a separate condition regarding the friendliness of her environments because this would be to off-load work into a non-ability condition. My suggestion, then, is that she can copy the work that the modest theorist does by somehow including the environment in the ability condition. A straightforward way to do this is to include factors external to the bodily individual into what makes up the ability. In short, this is to embrace cognitive externalism, which is what the argument of this chapter has suggested the epistemic externalist must do.

This might appear to be something of a concession, as it seems that my argument applies only to the robust virtue theorist. However, this is not straightforwardly the case. The considerations thus far adduced in this section have all been with respect to *knowledge*, rather than justified belief. The robust virtue theorist takes the invocation of cognitive abilities to be able to do all of the work required of an account of knowledge, whereas the modest virtue theorist claims that an additional condition is needed. The primary problem with robust virtue epistemology identified by Pritchard (forthcoming) and Kallestrup & Pritchard (forthcoming) is that it has trouble dealing with fake-barn style Gettier cases. In such cases, it is argued, the agent is clearly in possession of a (reliable, integrated) cognitive

---

considering that Fred's ability to read is in working order, and he has discriminately chosen the newspaper he reads for its reliability? Here is Greco. "[I]n the newspaper case, we can say that [Fred] does not believe the truth from an ability. Relative to the environment [he] is in [i.e., one with many mistake containing newspapers], [Fred] does not even *have* the ability to form true beliefs of the relevant sort" (67) [original emphasis]. In Kallestrup & Pritchard's terms, this is clearly tying the possession of ability to regional environments as we saw Millar does.

ability and clearly manifesting that ability. That is, Barney is attentively looking at a real barn on a clear, sunny day with his excellent eyesight, etc. The problem is that the nearby fake-barns foil Barney's knowledge. He is, however, perfectly well-justified in believing that the building he is looking at is a barn. This is the epistemic status we are interested in here. Hence, the difference between the robust and moderate virtue epistemologist *with respect to knowledge* does not arise.

To whom, then, does my argument apply? Throughout, the epistemic status with which we have been concerned is justification. The specifically targeted externalist accounts of justification have been those which appeal to an ability condition. What the discussion above demonstrates is that, more specifically, the argument applies to those theorists who *only* appeal to an ability condition. In the case of knowledge, there is an alternative externalist theory which appeals to separate ability and anti-luck conditions. If there were such a theory of justification, then my argument would not apply to it. However, there is no such theory of justification. In the discussion of knowledge, the reason the modest virtue epistemologist invokes a second, distinct condition is because the strong virtue epistemologist's position is seen to remain vulnerable to Gettier-style counterexamples. Thus, the second condition is essentially an anti-Gettier one. However, Gettier issues do not come up in the discussion of justification *per se*. Hence, there is no reason to appeal to a second condition when giving an account of justification.

My argument applies to the following sorts of positions, then. First, it applies to externalist accounts of justification, insofar as they appeal to an ability condition alone. For the various reasons discussed in Chapter 2, an ability condition seems to be an ideal condition on which to build an externalist account of justification. Indeed, this is how Greco gives his account of justification. Second, my argument will also apply to externalist robust virtue accounts of knowledge (e.g., Greco and Sosa) as they invoke only a cognitive ability condition. More generally, the argument will apply to any externalist theory of any epistemic status in which an ability condition is deployed to do all of the theoretical work, on the assumption that a mentalism-style internalist account of the very same epistemic status can be given in terms of an appeal to cognitive abilities (this assumption applies in the two specific cases above as well).

§2.2: *Objection to (4)*

The objection to (4) is actually a claim that there is a suppressed premise in my argument. I claim that an internalist account of justification consistent with mentalism is also consistent with cognitive internalism. However, the objection goes, the work I put (4) to actually requires the further assumption that a mentalist account of justification which appeals to ability is *not* consistent with cognitive externalism. If the mentalist can happily accept cognitive externalism, then I cannot derive any commitments to it on the epistemic externalist's part on the basis that she must deny mentalism. Worse, the form of my argument would suggest that, because mentalism is consistent with cognitive internalism *and* cognitive externalism, the epistemic externalist must reject both of them. However, on the assumption the cognitive internalism-externalism distinction is exhaustive, this would be an absurd conclusion.

However, remember what Conee (2007) had to say about the mental in his defence of mentalism. Conee leaves it in the hands of the relevant scientists to determine which mental states and processes there are and what they are like (56). It is inessential, claims Conee, that we have any particular sort of access to our mental life for it to do its justificatory work. All that matters is that our mental life is inside of us. I can see no plausible alternative interpretation of what Conee means here other than what is mental is exhausted either by what is neurally-bounded or by what is organism-bounded. Remember that Conee & Feldman say as much. "The relevant notion of what is internal is primarily spatial—the interiors of two *people* are the same in their *intrinsic qualities* [emphasis added] (82)."

It might be objected, however, that Conee & Feldman's cognitive internalist assumption is not obligatory for the mentalist. Indeed, if some version of cognitive externalism is true then such an assumption is indefensible. Further, if the mentalist should shed the cognitive internalist assumption, would not she be able to maintain that justification is internal to the mind, though the mental is now acknowledged to incorporate factors beyond the brain and body? Despite the initial plausibility of this move, it creates tension with other key epistemic internalist claims. Here, I will only briefly sketch an argument to the effect that mentalism cannot divest itself of cognitive internalism. The next chapter will explore this issue in greater depth.



Part of what typical epistemic internalists see as an advantage of their type of position is that, unlike epistemic externalist positions (such as reliabilism), epistemic internalism does not allow that justifiers can do their justificational work contingently. This stems from the fact that such internalists tend to accept the New Evil Genius intuition.

*New Evil Genius (NEG)* The extent to which  $S$  is justified at  $t$  in believing  $p$  is just the same as the extent to which  $S$ 's envatted duplicate is justified at  $t$  in believing  $p$  (Pritchard 2011: 365).

Giving a mentalist reading of NEG is to hold that only facts about which mental states a subject has—and not those states' causal histories—matter with respect to whether those states do any justificational work (Greco 2005). Hence, the justificational power of mental states is independent of their causal histories and embedding environment<sup>64</sup> and any subject who possesses a mental state with justificational power will possess a state which non-contingently has that power (i.e., it will possess its justificational power irrespective of its causal origin)<sup>65</sup>. That is, the epistemic internalist wants to rule out causal, environmental and social factors from possessing justificational potency (Conee 2007: 59). However, this does not sit well with the inclusion of brain-and-body external factors within the category of the mental. A problem immediately arises concerning how to reproduce the cases which illustrate NEG. The usual envatment scenarios presuppose that neurological duplication entails mental duplication, which in turn entails epistemic duplication. It is easy enough to replicate something like envatment for bodily individuals with a *Matrix*-style scenario. However, if factors external to a bodily individual are thrown into the mix, it starts to look like epistemic duplication requires the duplication of large portions of the environment

---

<sup>64</sup> The epistemic internalist may allow some environmental dependency in order to accommodate content externalism. An argument to this effect by Conee (2007) will be reviewed in chapter 5.

<sup>65</sup> Epistemic internalists do not say—and I do not interpret them as saying—that subjects have such states non-contingently.

subjects inhabit, including their interactions with other agents<sup>66</sup>. However, this appears to be at odds with epistemic internalism.

One might try to reproduce the NEG intuitions with a case adapted to the inclusion of external factors within mental states. Instead of an envatted person and her non-envatted counterpart, one might compare the victim of a *Truman Show* sort of scenario (where the victim lives what appears to her to be a normal life, but which in fact occurs entirely on an elaborate movie set wherein all the people she interacts with are actors playing roles and all the events of her life are scripted) with an individual whose life is genuine and unscripted. It is not clear, however, that the mentalist can sustain her supervenience claim in the *Truman Show* scenario. If mental states are in part constituted by the external features of the environment with which the subject interacts, then the *Truman Show* victim and her unscripted counterpart will not have the same mental states. (The “sky” in the scripted life is actually a painted ceiling; the “wind” is produced by large fans and not temperature gradients; and so on.) But, if the two subjects are not mental duplicates, then on what basis can it be claimed that they are justificatory duplicates? The strong supervenience of justification on the mental claimed by Conee & Feldman does not require that two subjects have identical mental states in order that they are justified in believing the same propositions (and to the same extent). Thus, the mere fact that two subjects are not maximal mental duplicates does not imply that they are not justified with respect to believing the same propositions. However, in the case the two subjects are not mental duplicates, any conclusion that they are nonetheless justificational duplicates cries out for explanation. However, the mentalist attempting to cope with cognitive externalism cannot satisfy this need. Further, the mentalist is faced with this problem *because* of the manner in which mental

---

<sup>66</sup> Hurley (2010) makes a similar point about internalist explanations of content. Hurley distinguishes the truth of an internal supervenience claim for the possibility of a “controlled STE” (i.e., a supervenience thought experiment).

Internal supervenience merely requires that thought contents do not vary *if* the relevant internal factors are not duplicated across different environments: unplugged and replugged. But the truth of this conditional claim does not require that the relevantly controlled STEs are possible: it may not be possible for the internal factors to remain constant while external factors vary, to be unplugged from one environment and replugged into another. They may vary together in the relevant possible worlds so that they are not explanatorily separable (8).

states are said to be constituted by the cognitive externalist. The mentalist's inability to explain the NEG intuition given cognitive externalism suggests that she cannot (easily) accept cognitive externalism.

Note that this is not like the argument (found, for instance in Pritchard 2011) that envatted counterparts may differ in their mental states due to external factors playing an individuating role with respect to the content of mental states. Instead, the cognitive externalist alluded to here claims that external factors play an individuating role with respect to what are sometimes called the vehicles of mental content: i.e., the mental states themselves (Hurley 1998b). To each object with which a subject interacts, there corresponds a distinct mental state<sup>67</sup>.

I have not claimed that it is impossible for the mentalist to reproduce the NEG intuitions for cognitive externalist cases; but I do think it is up to the mentalist to provide such a reconstruction<sup>68</sup>. Until she does, I maintain that the mentalist cannot simply accept cognitive externalism: it is not so easy to shed the cognitive internalist commitment evinced in Conee & Feldman. (I have been very brief here. But I will return to these issues later, as they are the focus of Chapter 5.)

---

<sup>67</sup> This may seem to perversely increase the number of types of mental states. For an argument to the effect that positing so many types of mental state may be necessary, see Burge 2010. Indeed, Burge's criteria of individuating mental state types results in even more types than the cognitive externalist one I have mentioned. Not only does Burge posit a distinct mental state type for each object of perceptual, but also for each perspective on each such object (Chp. 2, 34 – 42). While it is clear that Burge does not intend to argue for any sort of cognitive externalism (he's elaborating his anti-individualist take on content externalism), this style of argument can nonetheless be deployed by someone who is so arguing.

<sup>68</sup> The following way for the mentalist to preserve NEG has been suggested to me. Suppose that, in accordance with cognitive externalism, Otto's notebook (Clark & Chalmers 1998) is included as a constituent of Otto's mental states. For Vat-Otto, his mental states could include the portions of the computer simulation dedicated to producing in Vat-Otto the experience-as-of interacting with a notebook. The way Otto and Vat-Otto take the world to be is now the same, and so NEG is preserved.

This reply fails to address the cognitive externalist thesis. According to the sort of cognitive externalism under consideration, mental states are partly constituted by the factors with which a subject is actively engaged. Hence, for the cognitive externalist, Otto and Vat-Otto are in different mental states even if they are constituted by duplicated brains (or duplicated bodies). The envatted brain or body will be actively engaged with elements of a computer simulation (if it is actively engaged with anything at all), while the non-envatted body or brain will be interacting with things in

Nor is the fact that Conee speaks in terms of the ‘mental’ and Adams, Aizawa and Rupert in terms of the ‘cognitive’ a problem. The same passage which evinces Conee’s deference to the sciences of psychology and neuroscience seems to suggest that he would be happy to treat the difference in word choice here as insubstantial<sup>69</sup> (56). Adams and Aizawa (2007: 31) treat the cognitive as more than “[merely] biological, chemical or physical . . .” indicating that they are interested in the more abstract state types (e.g., psychological states) which Conee suggests are relevant to the mentalist. Hence, while Conee does insist that the epistemic internalist must countenance the mental, there is no reason to suppose that he would require that the mental be anymore than what Adams, Aizawa and Rupert claim the cognitive is. Conee’s deference will allow him to balk only at hardcore eliminativist projects (2007: 56 – 57). Of course, Adams & Aizawa and Rupert would balk at such projects as well.

### §2.3: *Objection to (2)*

The objection to (2) is more complicated. Because, in Chapter 3, I concluded there was no direct argument from the epistemic externalist’s account of ability to cognitive externalism, I had to move to the argument I have used in this chapter. The argument under consideration attempts to derive a commitment to cognitive externalism on the epistemic externalist’s part by appealing to her appeal to ability considered in the light of her obligation to deny epistemic internalism. In particular, I have claimed that it is the epistemic externalist’s clash with mentalist internalism which burdens the former with a commitment to cognitive externalism. However, it may be argued that, insofar as it presents something distinct from accessibilism, mentalism does not present epistemic internalism. But if this is the case, then it is not clear that the epistemic externalist must reject mentalism. Hence, my second argument to the epistemic externalist’s commitment to cognitive externalism fails.

Whether or not the mentalist presents a distinct vision of internalism is a significant issue in its own right, apart from how it effects my argument. As such, I will devote the next

---

the world. Indeed, if NEG could be preserved in this manner, the detour through *Truman Show* scenarios would not have been necessary at all.

<sup>69</sup> I draw this conclusion from the fact that Conee moves from talk about *mental* states to talk about *psychological* states without remark, and indicates that psychological science could tell us things about the mental.

chapter to exploring the issue to ensure justice is done to it. Of course, I will also discuss the implications of the conclusion I reach with respect to mentalism for the main argument of this work.

## Chapter 5

### Mentalism is Not Epistemic Ur-Internalism\*

#### *Introduction*

In Chapter 2, we saw that Earl Conee and Richard Feldman have developed a view of epistemic internalism which they call “mentalism”. The position is delineated explicitly in contrast to another view of epistemic internalism, accessibilism, which is the orthodox and dominant view of internalism. Accessibilism argues that there is some sort of access constraint on justifiers, and that endorsing some such constraint is definitive of epistemic internalism. Mentalism, on the other hand, argues that the epistemic internalist is not obliged to endorse any access constraint. In short, both accessibilism and mentalism are supposed to be epistemic ur-internalism: i.e., accounts which isolate the basic core of internalism in epistemology.

It was assumed in Chapter 4 that mentalism promulgates a thesis which any epistemic externalist is bound to repudiate. The epistemic externalist’s commitment to deny the mentalist thesis—combined with her appeal to cognitive abilities—was argued to commit her to some form of cognitive externalism.

---

\* A version of this chapter was presented as a stand-alone paper at the 2011 European Epistemology Network meeting at the University of Lund in Lund, Sweden.

In this chapter, I examine whether mentalism actually does put forth a thesis which the epistemic externalist must deny by questioning whether mentalism successfully stakes out an ur-internalism distinct from the accessibilist ur-internalism. If mentalism presents no epistemic internalist thesis distinct from accessibilism, then rejecting it requires nothing more of the epistemic externalist than the rejection of accessibilism. However, recall that in Chapter 4 we determined that the disagreement between accessibilists and epistemic externalists provided no reason to conclude that the latter incur any commitment to cognitive externalism. Thus, if mentalism fails to present anything new for the epistemic externalist to reject, the basis for my argument to the effect that epistemic externalists incur a commitment to cognitive externalism collapses. This chapter, then, provides a possible escape from the threat of such commitment (and—albeit incidentally—perhaps also provides a critique of mentalism).

The motivation for the arguments against mentalism in this chapter comes, oddly enough, from considering cognitive externalism. However, as will be seen, I am not concerned here with whether or not cognitive externalism is true—but instead will simply assume it presents a possible scenario. I argue that even this may be enough to threaten mentalism's bid for ur-internalist status.

In brief, my argument is that when the mentalist is faced with the possibility of cognitive externalism she cannot accomplish the tasks incumbent upon her to achieve ur-internalist status for her position. The tasks are as follows: a) distinguish her position from accessibilism, b) demonstrate that her position captures what is relevantly internal in epistemic internalism, c) demonstrate that her position preserves important internalist judgments concerning cases, and d) demonstrate that her position has theoretical advantages over accessibilism. I will explain later why the mentalist must do all these things.

From the outset, I should distinguish the arguments against mentalism presented here from several others extant in the literature. My argument will share some similarities with those of other authors, but its fundamental motivation is different from other critiques. What I say will be most similar to the arguments noted by Matthias Steup (2006) and Richard Fumerton (2007). Fumerton says, concerning the internalism-externalism debate in philosophy of mind, “[the] epistemic ‘internalist’ identifying evidence with mental states is

really an externalist ‘in disguise’” (2007: 35). Here, Fumerton means the debate about mental content. However, I think Conee (2007) has argued convincingly that content externalism poses no significant threat to epistemic internalism construed as mentalism. (We will survey Conee’s argument later.) I will be attempting to establish a point similar to Fumerton’s, except the philosophy of mind debate I will be concerned with is *cognitive* externalism-internalism, rather than content internalism-externalism. Steup points out (in a footnote to his Stanford Encyclopaedia entry ‘The Analysis of Knowledge’) that the mentalist may face some difficulties distinguishing her position for accessibilism.

The problem here is that internalists would not want to count, for example, the state of being reliably clairvoyant as a mental state . . . Thus the need arises to differentiate between neurophysiological states that are, and those that are not, mental states . . . But if the direct recognizability criterion is needed to define mental states, then mental state internalism does not appreciably differ from accessibility internalism (2006, fn 14).

Here, Steup is claiming that the mentalist may need to retreat to accessibilist claims in order to preserve a key internalist intuition about cases. Again, I will also claim that the mentalist may need to retreat to accessibilist claims. However, I will not motivate this by claiming that the mentalist needs criteria to distinguish mental states from merely neurophysical states. Hence, with respect to arguments against mentalism like Fumerton’s and Steup’s, I consider similar possible flaws in mentalism. The difference in my arguments will be the motivations and routes to these critical conclusions. That is, I am not merely restating or evaluating extant objections, but exploring new ones.

There are two further arguments against mentalism I wish to mention: one from Sanford Goldberg (2007) and another from Jessica Brown (2007). Sanford Goldberg considers two arguments which I see as attacks on mentalism’s claim to ur-internalist status. First, he takes Conee & Feldman to advance mentalism as a conjunction of the claims “(a\*) only mental states can be epistemic justifiers and (b) [epistemic justification . . . is a necessary condition on knowledge]” (2007: 135, FN 5). As a thesis which attempts to articulate the unelaborated core of epistemic internalism, if mentalism does not count paradigmatically epistemic internalist theses as internalist, it fails as ur-internalism. However, Goldberg



claims, many theorists who plausibly self-identify as epistemic internalists do not endorse (b) (2007: 136, FN 5). Hence, if (b) is part of mentalism, then mentalism will deliver an implausible picture of the internalism-externalism landscape. Second, Goldberg argues that mentalism abandons what he takes to be basic internalist intuitions about what makes something internal—*viz.*, the intuition that “only states that are first-personally accessible can be justifiers” (2007: 135, FN 5).

For my purposes, I will construe mentalism as only putting (a\*) forward. That is, I will take the mentalist to be concerned only with epistemic justification. Even if Conee & Feldman advance a mentalist account of internalism about knowledge, they can advance an interesting mentalist account of internalism about justification separately<sup>70</sup>. Further, I will give the mentalist a chance to demonstrate that her position can account for intuitions about first-personal access. After all, as we saw in Chapter 2, mentalism construed as *ur*-internalism is supposed to include all theories included by accessibilist *ur*-internalism. She does not abandon intuitions about first-personal access; she just demotes them in importance.

Jessica Brown’s objection to mentalism is based on the apparent inadequacy of the cases Conee & Feldman use to motivate their position.

But Conee and Feldman provide no reason for thinking that the six cases [in favour of mentalism] exhaust all possible cases, admitting they ‘have no proof that there is no exception to the pattern exhibited by the examples’ [quoting Conee & Feldman 2004: 238] (2007: 26).

Brown makes this claim in the service of arguing that no viable form of epistemic internalism escapes the threat of content externalism. She does *not* claim that the mentalist can give no argument securing the compatibility of her position with content externalism. Instead, she claims that the mentalist never gets a chance to make such an argument, because

---

<sup>70</sup> Goldberg’s first objection would apply to a mentalist account of internalism about knowledge—call it “k-mentalism”. However, in both Conee & Feldman (2004) and Conee (2007), mentalism is presented as an account of internalist theories of justification—call it “j-mentalism”. Feldman (2004) is the source Goldberg cites as evidence that mentalism is k-mentalism. However, nowhere in Conee & Feldman (2004) or Conee (2007) is it claimed that mentalism itself—characterized by their claim of strong supervenience of justification on the mental—is a thesis about knowledge. They may claim that justification is necessary for knowledge, but that claim is separate from the thesis of mentalism.

her position is not sufficiently motivated from the beginning. For my purposes, however, we will give Conee & Feldman the benefit of the doubt and assume that their cases for mentalism are exhaustive of all possible cases, or that there is reason to suspect that there is no exception to the pattern exhibited by the examples they furnish. My argument is thus clearly distinct from Brown's.

In the rest of the chapter, I will proceed as follows. In Section 1, I recapitulate mentalism and its rival, accessibilism, and then provide evidence that the proponents of these accounts are not talking past each other. In addition, I will demonstrate that mentalism and accessibilism aim at the same goal—*viz.*, to develop an account of ur-internalism. Further, given that the dispute between the mentalist and the accessibilist is substantive, I will outline what I take to be incumbent upon the mentalist if she wishes to carry the day for her position. That is, I will explain in more detail tasks (a) – (d) mentioned earlier in this section.

Section 2 will begin with a recapitulation of cognitive externalism in the philosophy of mind. I will argue that, given merely the possibility of cognitive externalism, the mentalist's ability to accomplish the tasks set out for her in Section 1 appears to be under threat. Finally, Section 3 will explore what options are available to the mentalist in the face of the threat which I argue cognitive externalism poses to her position. After surveying the options, I conclude that the mentalist cannot succeed in carrying the burden set out in Section 1.

### *§1: Accessibilism versus Mentalist—Redux*

Before making an argument, we must review mentalism and accessibilism. Recall that mentalism in epistemology is explicitly characterized by two theses:

S: The justificatory status of a person's doxastic attitudes strongly supervenes on the person's occurrent and dispositional mental states, events and conditions.

M: If any two possible individuals are exactly alike mentally, then they are alike justificationaly, e.g., the same beliefs are justified for them to the same extent<sup>71</sup> (Conee & Feldman 2004: 56)

What is supposed to make this position distinctively internalist is that it captures what is essentially internal about justification—*viz.*, that it supervenes on the mental, generally. Like other versions of internalism, it discounts reliabilism and other epistemic externalist positions by denying that “contingent or purely environmental variation [can affect] justification” (Conee 2007: 59; see also *op cit* 55, fn 5). This benefit is gained by treating the mental itself as strongly supervening on factors internal to the subjects under consideration. Mentalism is distinguished from the traditional view of internalism by omitting any requirements regarding the reflective accessibility of an agent’s justification by that agent: i.e., positions which rule out as justifiers any conditions which an agent cannot, in principle, access via reflection. Positions characterized by such requirements are generally known as accessibilism. James Pryor flatly states that epistemic internalism is always concerned with “. . . facts to which one has a special kind of access” (2001: 103). Hence, Fred is justified in his belief that the toast is burning when, by reflecting upon his beliefs and experience, he discovers what gives support to his belief (e.g., olfactory impressions of carbonizing bread, visual impressions of smoke issuing from the toaster, etc.). Recall Pryor’s “simple internalism”.

*Simple Internalism* (SI) Whether one is justified in believing P supervenes on facts which one is in a position to know by reflection alone (2001: 104).

Here, ‘reflection’ is stipulated to include, “*a priori* reasoning, introspective awareness of one’s own mental states, and one’s memory of knowledge acquired in those ways” (*ibid*). That is, recourse to environmental, behavioural or statistical data is not relevant to the justificatory status of beliefs. This notion of internalism is simple because it leaves unspecified both the strictness of the conditions upon the access which a subject must have to the facts which

---

<sup>71</sup>In a footnote, Conee & Feldman note that, “[whether] (S) implies (M) depends on depends on details of the supervenience relation which we will not discuss here” (2004: 56, FN 12). I will assume that (S) entails (M).

justify her beliefs, and whether that access must include a perspective on the justificatory relation of facts to beliefs. The strictness of conditions upon a subject's access to justifiers varies along two axes. First, the access can be required to be either only *in principle* access, or *actual* access. The latter is stricter than the former because it requires that a subject does reflect upon and discover the justifiers for the beliefs she forms, whereas the former requires that a subject merely *be able* to reflect upon and discover the justifiers for the beliefs she forms. Second, the breadth of justifiers which the subject must (actually or possibly) reflect upon may be varied. In the most permissive instance, the internalist might require only that a subject access *some* of the justifiers for her beliefs. A maximally strict interpretation, on the other hand, would require that a subject access *all* of the justifiers for her beliefs. Further, the internalist might opt for a qualitative restriction—rather than the quantitative sort already mentioned—requiring that a subject access at least the *crucial* justifiers for her belief.

Further, SI might be escalated to what Pryor refers to as “access internalism” (not to be confused with the label I use for Pryor's theory of internalist theories, “accessibilism”). Here is the principle which characterizes the further escalation.

(*Access Internalism*) One always has ‘special access’ to one's justificatory status  
(2001: 105)

With this principle, the internalist claims that a subject, when justified, will be able to tell (actually or possibly) that she is justified. This is a quite strong position. Though it has its defenders (e.g., Chisholm 1989; Bonjour 1980), we will not dwell on it here. I mention it only to emphasize the point that simple internalism is a baseline accessibilist principle which can be augmented in various ways to yield a variety of epistemic internalist positions.

Earl Conee states that such commitments are extraneous. A position can count as a version of epistemic internalism without imposing these commitments. “It seems extra to imply that our innards have any particular psychology. It also seems extra to imply that knowledge of our innards is definitely available to us” (Conee 2007: 56). Importantly, taking accessibility as the core notion of internalism rules out as internalist certain positions which intuitively warrant the classification, according to Conee.

For example, a philosopher might regard *a priori* reasoning and introspection as discredited categories of folk psychology. She might nonetheless affirm, for instance, that one's epistemic justification strongly supervenes on one's sensory states. A 'Sensationalist' of this sort ought to qualify as an internalist (*ibid*).

In short, Conee claims that the access criteria Pryor takes to characterize any epistemic internalist position—historical, contemporary or unexplored—both fail to do this and further burden the internalist with troublesome commitments. Conee understands mentalism as isolating the unelaborated core of epistemic internalism: it bars the same reliabilist intuitions that accessibilism does, but carries no extraneous psychological or philosophical commitments. Just as Pryor's simple internalism can be in various ways adjusted to be stricter, mentalism can be made stricter by *re-adding* the access requirements. While the mentalist considers reflective access requirements to be *unnecessary* to an internalist position, she does not consider them to be *insufficient* for a position to be an internalist one. Hence, mentalism is clearly intended to be the simplest version of internalism in the same way Pryor's simple internalism is. To avoid confusion by couching the disagreement between accessibilists and mentalists as giving competing formulations of simple internalism, I have coined "ur-internalism" to describe that which the two parties are in competition to give the best formulation of.

We might be concerned that the mentalist (represented by Conee & Feldman) and the accessibilist (represented by Pryor) merely talk past each other, or are simply interested in different things. That is, they merely identify different aspects of internalism or use the term 'internalism' differently. However, this is not the case. Each is aware of the other's way of characterizing what is relevantly internal to epistemic internalism, and each considers the other's way to be faulty. Pryor bruits that the construal of internal used in debates about content in the philosophy of mind does not capture what is at issue in the epistemological case.

Internalism in epistemology is a different sort of issue than internalism in the philosophy of mind. Philosophers of mind are concerned with whether

certain properties supervene on *one's intrinsic make-up*. Epistemologists are more concerned with the connection between one's justificatory status and *facts to which one has a special sort of access . . .* Clearly, not every state that is 'internal' in the philosophy of mind sense will be among these states (2004: 103) [emphasis added].

Thus, Pryor indicates that to appealing to mental states beyond those to which one has some special sort of access is to lose sight of what internalism in epistemology is about. Indeed, though Pryor does not discuss it, this sort of access could obtain between a subject and facts which are not internal in the content internalist's sense at all<sup>72</sup>. Hence, Pryor's claim might be strengthened to say that, in addition to labelling some states as (epistemologically) internal which ought not to be so labelled, the mentalist will fail to label some states as (epistemologically) internal which ought to be so labelled.

Further, the mentalist directly engages with and repudiates the characterization of internalism which Pryor puts forth. We saw earlier that the processes and methods Pryor takes to characterise reflective access are the same processes and methods which Conee claims to be the burdensome and unnecessary commitments which accessibilism foists on epistemic internalism. In addition, it is clear that Conee & Feldman take up the exact interpretation of 'internal' which Pryor finds wanting. "The relevant notion of what is internal is primarily spatial—the interiors of two people are the same in their intrinsic qualities" (Conee & Feldman 2004: 82).

Clearly, the accessibilist and the mentalist do not talk past each other. Moreover, they are obviously attempting to accomplish the same thing. That is, each is trying to isolate the core notion of epistemic internalism—shorn of the trappings of particular theories—thereby formulating ur-internalism. From this I conclude that there is a substantive disagreement

---

<sup>72</sup>This is quite speculative (and inessential). What I have in mind are considerations like those found in Michael Tye (2009), where it is claimed that when we introspect, what we attend to are features of the world. For instance, when introspecting our experience of a red apple, focusing on the red part of the experience is just to attend closely to the redness of the apple. Such an account would furnish us with examples where an access relation of just the type Pryor has in mind obtains between an agent and a factor which would clearly be ruled as external on a content internalist-style notion of the external-internal distinction. Importantly, these factors are taken to be individuating of a mental state, but *not* part of the state. Hence, the factor upon which a subject reflects in such an instance is not a mental one. Thus, the mentalist supervenience thesis would be violated.

between the two parties: an internecine conflict the resolution of which is of interest to everyone with a stake in the externalism-internalism debate in epistemology<sup>73</sup>.

How can this dispute be resolved, and in which side's favour will the resolution be? To determine this, some attention needs to be paid to the dialectical situation. Accessibilism is the traditional, orthodox way of cashing out what is relevantly internal to internalism. Talking about justification in terms of 'internal' versus 'external' goes back at least to Roderick Chisholm (1977 and 1989), who gives the ability to access justifying facts prominence<sup>74</sup>. This sort of position is also taken to be implicit in earlier epistemological theorizing<sup>75</sup>. Mentalism, then, is a challenger to orthodoxy and bears the burden of proof. The mentalist must, to establish the supremacy of her thesis, accomplish a few tasks: a) distinguish her position from accessibilism, b) demonstrate that her position captures what is relevantly internal in epistemic internalism, c) demonstrate that her position preserves important internalist judgments concerning cases, and d) demonstrate that her position has advantages over accessibilism.

The mentalist must accomplish (a) for obvious reasons: if her position is not distinct from the one with which it is in competition, then she is offering just another among many versions of accessibilism. However, the mentalist aims not to put forth a *specific position* about the nature of justification, but an account of positions about the nature of justification. Moreover, she aims to give an account of positions concerning the nature of justification which is importantly different from accessibilism (itself not construed as a specific position, but a way of categorizing accounts of justification). For similar reasons, task (b) is incumbent upon the mentalist. She argues that the accessibilist account of epistemic internality involves

---

<sup>73</sup>One might wonder if interest in this question is mere idiosyncrasy, not indicative of the issue's fruitfulness with respect to the broader project of epistemology. After all, who cares which positions are *called* internalist or externalist? There is a quick answer. Conee and Feldman seem to care—as does Pryor. Others have bothered to write that mentalism does not deserve to be called internalism. That is enough for me.

<sup>74</sup>I understand, however, that the terms 'internalism' and 'externalism' for epistemological positions concerning knowledge go back to David Armstrong (1973).

<sup>75</sup>See, e.g., Prichard 1950. Prichard plausibly reads an access criterion into Descartes' philosophy. Moreover, the ability to *provide* justification for beliefs has been read into epistemological theory as far back as Plato. Intuitively, provision of reasons involves access to justifications.

extraneous commitments and restrictions. If the mentalist is going to make a positive case as well, she had better put forth an alternative to the account she criticizes.

Insofar as the mentalist doesn't want simply to retool 'internal' for her own epistemological purposes, she will want to accomplish task (c), preserving as many judgments about cases usually associated with internalism as possible. Indeed, as we saw earlier, Conee claims that every theory of justification labelled "internalist" by accessibilism will be given the same label by mentalism. Additionally, mentalism will label some theories of justification "internalist" which accessibilism fails (and culpably so, claims Conee) to so label. To do this, it would seem that the mentalist wants to preserve *all* judgments about cases that extant internalist theories have made.

Of course, it would not be enough merely to accomplish (a), (b) and (c). Theoretically, this would put mentalism and accessibilism in a stalemate, with no apparent considerations to suggest adopting one over the other. However, dialectically, this would leave accessibilism with the upper hand as the default position. If the mentalist wants to unseat accessibilism as ur-internalism, she had better also demonstrate that her position has theoretical advantages over the accessibilist's. I will argue that the mentalist cannot accomplish these tasks.

### §2: *Mentalism Defeated*

Mentalism's claim to ur-internalist status is supposed to arise from the claim that justification strongly supervenes on the mental, where the mental itself supervenes on factors internal to the relevant subjects. The claim about the supervenience base of the mental is not part of the formal apparatus of mentalism (i.e., it is not stipulated in either (S) or (M)), but it does appear to be what motivates Conee & Feldman's contention that supervenience on the mental captures what is relevantly internal for epistemic internalism. For this to be the case, the mental states, events and conditions we attribute to an agent need to occur inside the subject of interest. This is clearly borne out by Conee and Feldman's comparison of their mentalist position to content internalism:

What internalism in epistemology and the philosophy of mind [content internalism] have in common is that being in some condition which is of



philosophical interest . . . is settled by what goes on inside of cognitive beings (Conee & Feldman 2004: 57).

Despite the similar motivations for content internalism and epistemic internalism, Conee (2007) has argued that the two do not stand or fall together. Conee cites the development of the notion of a “counterpart proposition”.

If, according to the content externalist, two people who are internally alike believe a proposition that each could express by the same sentence, then the proposition that one believes is a counterpart of the proposition that the other believes. If they are believing different propositions due to external differences, then the two propositions are counterpart propositions for those two people at that time. If they are believing the same propositions, then that proposition is a counterpart of itself for them at the time (Conee & Feldman 2004: 82).

Thus, even though an earthling and her twin-earth counterpart are entertaining strictly distinct propositions (e.g., the former about H<sub>2</sub>O and the latter about XYZ), the two are justified in believing their respective counterpart propositions to the same degree.

Conee further claims that this epicycle is not necessary to render mentalism consistent with content externalism. He takes it as plausible that, “. . . a proposition has some justificatory status for someone only if she has the concepts enabling her to engage the proposition” (Conee 2007: 52). If the content externalist believes not only that the propositional contents of mental states are externally individuated, but also that the concepts an agent has are in part determined by environmental factors, then the earthling and his twin-earth counterpart aren’t mental duplicates in the first place.

Mentalism takes ‘internal’ duplicates to be mental duplicates. Mental duplicates must be justificatory duplicates. Given the conceptual requirement, the twins are not mentally alike . . . So mentalism does not imply that the same propositions are justified for each of them (*op cit* 53).

In any case, mentalism and content externalism are at least consistent<sup>76</sup>. Note Conee's statement that mentalism need not change to accommodate content externalism suggests that, as stated, the mentalist's position is neutral with respect to the content internalism-externalism debate.

However, mentalism is *not* consistent with externalism about *mental states themselves* (as opposed to the content of those states). Recall that this is one of the positions for which I have adopted the term "cognitive externalism". Such positions claim that mental states are in part constituted by what goes on outside of individuals (e.g., Clark 2009; Hurley 2010; Noë 2004). Cognitive internalism, on the other hand, contends that mental states are exhaustively constituted by factors *within* individuals (Adams & Aizawa 2009; Rupert 2008). In each case, there is an intuitive notion of what an individual is—*viz.*, a roughly person-sized chunk of space-time. I have been referring (and will continue to refer) to this as the "bodily individual".

It might be immediately claimed that there is a confusion here. Cognitive externalism, after all, is concerned with cognitive states/processes, and whether or not they extend beyond bodily individuals. However, this need say nothing about *mental* states/processes; and it is particularly the mental—rather than the cognitive—that the mentalist adverts to. The distinction between the mental and the cognitive is acknowledged in some cognitive externalist arguments. Clark & Chalmers (1998) present two arguments: one for extended cognitive processes and another for extended mental states. Further, they explicitly acknowledge that these are separate arguments, claiming that arguing for extended cognitive processes is not arguing for extended mental ones (12). Does this stop my argument before it begins? I do not think it does.

Let me give the weakest reason first. Conee & Feldman say that the factors relevant to justification are "inside" *cognitive* beings (2004: 57). If, for Conee & Feldman, the cognitive and the mental were importantly distinct, then it is not at all clear why what goes on inside of cognitive beings would capture all the mental phenomena Conee & Feldman wish to

---

<sup>76</sup> However, see Williamson (2000: 58). He argues that epistemic internalism is false on the assumption of a certain kind of content externalism. See Conee (2007) for a direct response to this claim.

appeal to. Of course, this could just be due to looseness of terminology. Nevertheless, further consider the sorts of things which are supposed to count as mental under mentalism: sensory states of the kind that would be endorsed by an empirical sensitive philosopher who takes, “. . . *a priori* reasoning and introspection [to be] discredited categories of folk psychology” are counted as mental states. Plausibly, under the operative description, sensory states are just states of the sensory organs and sensory processing areas of the brain. But such states seem to count as cognitive as much as they count as mental.

Finally, recall one of the independent critiques of mentalism mentioned at the beginning of the chapter. Steup claims that to retain core internalist judgments about cases, “. . . the need arises [for the mentalist] to differentiate between neurophysiological states that are, and those that are not, mental states . . .” (2006, FN 14). The plausible distinction, according to Steup, is that mental states are neurophysiological states to which subjects bear a certain sort of access relation. Of course, the mentalist does not want to invoke this criterion to identify mental states. Nor do Conee & Feldman suggest an alternative way to make the distinction. Thus, I do not take there to be any problem in basing my objection to mentalism on considerations of cognitive externalism.

While neither (S) nor (M) rule out cognitive externalism, the mentalist’s insistence that appeals to mental states capture what is relevantly internal about justification does rule out cognitive externalism. Cognitive externalism is in tension with the spirit of mentalism, rather than with the letter. Some explanation is in order.

The consistency of (S) and (M) with cognitive externalism is straightforward. Nothing in either the supervenience or cross-world comparison claims specifies any restrictions of the base properties other than that they be mental. Recall that the supervenience base in (S) is “the person’s occurrent and dispositional mental states, events and conditions” (Conee & Feldman 2004: 56). For all this says, Clark & Chalmers’ (1998) example of Otto and his notebook will satisfy Conee & Feldman’s supervenience thesis. Let us rehearse the case. The example involves an Alzheimer’s sufferer, Otto, and the notebook he uses to augment his faulty memory. Otto records information he is likely not to remember in his notebook. Hence, when he hears about a new exhibit at a local museum and forms a desire to go see it, he consults his notebook to determine the location of the museum. Otto’s desire, combined

with his consultation of the information in the notebook, is supposed to be the best explanation of Otto's trip to the museum. Moreover, it is supposed to be exactly parallel to the explanation given for an individual with normal functioning biological memory, Inga, when she makes the same trip. The only difference is the location of some of the pertinent factors; which difference cannot be appealed to as relevant on pain of begging the question against cognitive externalism. With respect to mentalism, the situation can be recast dealing with justification rather than explanation for action. The information in Otto's notebook in part constitutes a dispositional belief about the museum's location, just as the information in Inga's biological memory constitutes a dispositional belief. The mentalist cannot insist that the difference in location of the information makes a difference to its justificatory role anymore than the opponent of cognitive externalism can insist that the location of the information makes a difference for the explanation of certain actions. Further, the mentalist does not want to give an argument about the nature of mental states at all: she takes it that she can assume mental states are internal to the relevant individual. I suggest she cannot assume this.

One point of Clark & Chalmers' paper is to establish that the underpinnings of some mental states (in particular, dispositional beliefs) can go beyond the body boundary of the individual to whom we attribute those mental states. The mentalist wants to leave it open that any mental state might do justificatory work. Given this, cognitive externalism is consistent with (S) as stated. Further, because cognitive externalism is consistent with (S), and (M) is entailed by (S) (Conee & Feldman 2004: 56; and see fn 12), the cross-world comparisons come off smoothly as well. If we stipulate that facts about Otto and his notebook on the one hand—and Otto\* and his notebook on the other—are held fixed in their respective worlds, the two should be justified in believing all of the same propositions and to the same degree.

If the mental states that mentalists appeal to are correctly characterized by cognitive externalism, they no longer exemplify what is relevantly internal about justification. Recall that Conee & Feldman's version of epistemic internalism has it that facts about justification are "settled by what goes on inside of cognitive beings". That is, the location of mental states is supposed to be what makes them relevantly internal. They are internal to each

subject whose justificatory status with which we are concerned. However, if these states are no longer “inside” of the intuitive boundaries of subjects, it is not clear that they are relevantly internal. But one might object that, taking cognitive externalism into account, the mentalist can still make her case: surely, if claims like that in Clark & Chalmers are correct, then ‘cognitive being’ simply picks out “larger” objects than we originally thought. However, the comparisons Conee & Feldman make between mentalism and other forms of supervenience-driven internalism in philosophy indicate that they take ‘cognitive being’ to pick out individuals as they are canonically delineated—*viz.*, by boundaries of the body. According to these comparisons, internalists of all stripes want to rule out environmental and social factors (*op cit* 57). But these are just the sorts of factors that cognitive externalists want to include (e.g., Otto’s notebook—or skilled interaction with the environment, as in Noë 2004). Finally, not to put too fine a point on it, Conee & Feldman explicitly state that ‘internal’ means “within the bounds of an individual body”. “The relevant notion of what is internal is primarily spatial—the interiors of two people are the same in their intrinsic qualities” (*op cit* 82). Mentalism is not consistent with cognitive externalism.

Further, it seems that if Conee & Feldman were to take cognitive externalism on-board, they would change what appears to be the order of explanation between the boundaries of subjects and justification-conducive factors which their theory initially embraces. What I mean is this. Internalists in general—and Conee & Feldman in particular—take what counts for justification to be *constrained* in some regard by the nature or abilities of individuals. Accessibilists consider the reflective capacities of individuals to be the constraint on what counts for justification. That is, if a particular agent cannot—at least in principle<sup>77</sup>—access some factor reflectively (as it has been defined earlier), then that factor cannot serve as a justification for the relevant agent’s beliefs<sup>78</sup>. Mentalists, on the other hand, take something like location to be the relevant constraint. Here, being internal is—as we have already seen—cash out in terms of being “inside” a relevant agent. Now, of course, not everything inside an agent is relevant to justification (neither blots of mustard nor

---

<sup>77</sup>Different internalist accounts will differ with respect to the strength of the access constraint placed on the subject.

<sup>78</sup>This point applies equally well to both doxastic and propositional justification.

underdone potatoes need apply); but all factors relevant to justification will be “inside” the agent. However, if the mentalist accepts cognitive externalism, then she lets the spatiotemporal spread of factors relevant to justification place constraints on the nature of the individual. This is to say something like, “What are the factors relevant to justification? Let us extend the individual so that it encompasses those things. Now mentalism is true!” Can the mentalist do this?

In the debate about mental content there is an analogous argument made about the content internalist’s case. That is, people have argued that the internalist should not tie herself to a particular “size” of individual subject. If this sort of argument can work for the content internalist, why can it not also work for the mentalist? Let us look at the argument in the case of content internalism from Gary Bartlett (2008).

Bartlett argues that cognitive externalism, of necessity, does not refute—nor generally cause trouble for—content internalism. Any appearance that cognitive externalism does spell trouble for the content internalist is due, says Bartlett, to a faulty characterization of content internalism. However, he takes this faulty characterization to be the one which has been taken up by defenders of content internalism and its antagonists alike. He dubs this formulation of content internalism “INTERNALISM<sub>T</sub>” (the subscripted “T” stands for ‘traditional’).

*(INTERNALISM<sub>T</sub>):* Each of a subject’s mental states at time *t* supervenes on the subject’s internal physical state at *t* (2008: 164)

This formulation limits the supervenience base for a subject’s mental content to the extent of her physical body. If INTERNALISM<sub>T</sub> is false, then its negation—EXTERNALISM<sub>T</sub>—is true. Therefore, EXTERNALISM<sub>T</sub> is true if at least one of a subject’s mental states at *t* supervenes on a physical state (at *λ*) external to the relevant subject. That is, as Bartlett puts it, the skin becomes “. . . a crucial *mental* boundary that the internalist must guard at all costs” [original emphasis] (*ibid*). If the content internalist must defend the boundary of the skin, then cognitive externalism—if true—will refute content internalism, says Bartlett. However, he argues that the content internalist can dispense with what has been pejoratively

referred to as the “magical membrane” (Hurley 2010) intuition about the skin and thereby remain neutral as to the debate between cognitive internalists and externalists. Thus, INTERNALISM<sub>T</sub> must be the wrong formulation of internalism.

Bartlett makes his argument by way of considering Daniel Dennett’s (1981) “Where am I?” case. Schematically, the case involves a subject’s brain being removed and envatted, with all functional links between the brain and the subject’s body replicated by radio transmission. So the subject functions just as she did when her brain was in her skull, but the information loop which initiates her actions is extended across the distance between the agent and her brain. Further, by assumption, the subject’s mental states supervene on her neural states. The question which arises, then, is whether the distance between the subject’s body and her brain causes any trouble for the internalist. After all, this case appears to engender a situation in which the supervenience base for a subject’s mental content (i.e., her neural states) is outside of the physical extent of the subject (i.e., her body). Bartlett claims that it does not, and this because—though the spatial extent of the factors causally responsible for the subject’s mental contents has changed—the factors themselves have not changed. Citing Ted Levine (1999)<sup>79</sup>, Bartlett claims that what we should learn about internalism is that, “[it] has no necessary commitment to the mind’s “causal basis” (as [Levine] calls it) being limited to any particular spatiotemporal region” (Bartlett 2008: 172). Hence, for the content internalist, ‘internal’ means “part of the causal basis of”—however extended that causal basis might be<sup>80</sup>. Specifically, Bartlett assumes for the purposes of argument that cognitive externalism is true of the Dennett case, and claims that a refutation of content internalism is not in the offing.

However, these cases do refute INTERNALISM<sub>T</sub>, Bartlett claims. The defender of INTERNALISM<sub>T</sub> cannot, claims Bartlett, make her position consistent with the Dennett case by claiming that the subject as adverted to in her formulation expands to include all

---

<sup>79</sup> Bartlett notes that Levine is concerned with internalism with regard to the supervenience base for conscious experience rather than mental content.

<sup>80</sup> Bartlett adduces two modified versions of Dennett’s case to support his point, though he thinks it follows clearly enough from Dennett’s original version. Because I am not concerned here to dispute Bartlett’s claim about the consistency of content internalism with cognitive externalism, I will concede the argument to him without considering his further cases.

factors causally responsible for that subject's belief. Referring to the case of Otto's notebook, Bartlett says that this means that the defender of INTERNALISM<sub>T</sub> cannot claim that the notebook is part of Otto himself, and hence part of the subject whose innards she wishes to render as the supervenience base for Otto's mental content. Bartlett cashes this attempted defence out as an endorsement of the following principle.

(\*) Necessarily, if an object X carries some of a subject S's beliefs, then X is a part of S (2008: 175)

However, this principle is far from obviously true, says Bartlett. Indeed, Bartlett cites a cognitive externalist who denies this principle—*viz.*, Rob Wilson (2004). As we saw in Chapter 1 (INTERNAL SECTION REFERENCE), Wilson's route to cognitive externalism is via a refined account of the realisation relation. Importantly, he holds fixed the bodily individual as the subject of mental state attributions, arguing to cognitive externalism by claiming that mental states attributed to bodily individuals can be widely—or even radically widely—realised in virtue of portions of the total or core realisation laying outside the bodily subject of whom we predicate the relevant mental properties. Given the fact that a cognitive externalist can coherently reject (\*), Bartlett takes it that citing (\*) is not sufficient to argue for the compatibility of INTERNALISM<sub>T</sub> with cognitive externalism. As such, he considers the dialectical burden to be on the supporter of INTERNALISM<sub>T</sub>. More to the point, Bartlett says the following.

If we want to decide whether (\*) is true, we will have to do some independent investigation into the issue of the spatial extent of subjects. But if we must do *some* conceptual spadework in order to reformulate internalism, why dig into the concept of the *subject* in the hopes of thereby revealing something about the *mind*, when we could dig into the concept of the mind directly? (2008: 176)

Ultimately, Bartlett returns to Levine to give a reformulation of content internalism which is consistent with cognitive externalism. Such a formulation is supposed to recognize that what is relevant to the mental content of some subject's beliefs is the intrinsic features of that



subject's mind, where the mind of a subject is not necessarily circumscribed by the physical boundaries of the subject. Bartlett derives a so-called "causal basis internalism" from Levine.

'Causal basis' internalism (INTERNALISM<sub>C</sub>): Each of a subject's mental states at time  $t$  supervenes on the causal basis of the subject's mind at  $t$  (2008: 180).

The causal basis of a subject's mind is taken to be something like the aggregate of the core realisers of all her mental states. And, as we have seen, in cases of what Wilson (2004) calls radically wide realisation, the core realisers of some of the subject's mental states can stand beyond the physical boundaries of that subject. Hence, that which is part of (or, within) the causal basis of subject's mind can fail to be a part of (or, within) the physical body of that subject. Thus, so long as all of the factors relevant to the content of a subject's mental states are part of the causal basis of her mind, INTERNALISM<sub>C</sub>, can be true while either of cognitive internalism or cognitive externalism is true. Cognitive externalism does not entail the falsity of content internalism construed as INTERNALISM<sub>C</sub>.

Assuming that Bartlett's argument works for content internalism, can it be adjusted to work for mentalism? After all, Conee & Feldman take their claim about justification to be underpinned by the same notion of internality which content internalists use. Might my argument against mentalism on the basis of cognitive externalism rest on a flaw similar to that which Bartlett attributes arguments against content internalism made on the same basis? Let us go through this matter slowly.

The mentalist claims that the justification for a subject's beliefs supervenes on the mental states of that subject. She takes her thesis of justification to be an internalist one *because* of its claim that justification supervenes on the mental. Before taking Bartlett's arguments into account, we can attempt to reconstruct her claim in the following way.

- (1) A subject  $S$ 's justification for her beliefs at  $t$  supervenes on her mental states at  $t$
- (2) A subject  $S$ 's mental states at  $t$  supervene on her internal physical state at  $t$

- (3) If  $X$  at  $t$  supervenes on the internal physical state of  $S$  at  $t$ —and  $Y$  at  $t$  supervenes on  $X$  at  $t$ —then  $Y$  at  $t$  supervenes on the internal physical state of  $S$  at  $t$
- (4) Therefore,  $S$ 's justification for her beliefs at  $t$  supervenes on her internal physical state at  $t$

Premise (1) is just the supervenience principle, (S), of the formal statement of mentalism. Premise (2) is what Bartlett identifies as INTERNALISM<sub>T</sub>. Premise (3) follows from the transitivity of the supervenience relation. The conclusion, (4), follows straightforwardly from the replacement of 'X' and 'Y' in (3) by 'a subject S's mental states' and 'a subject S's justification for her beliefs', respectively. (Note that this reconstruction construes the mentalist's appeal to mental states as an appeal to the content of those states. However, it is not clear that (S) limits the supervenience base of a subject's justification for her beliefs to mental states with the sort of content at issue in the content internalism-externalism debate. I highlight this point only to push it aside for the moment. We will come back to it later in the chapter.) If this is the way to reconstruct the mentalist's reasoning, then—insisting on (1)—the failure of (2) will imply a failure of (4). That is, the falsity of INTERNALISM<sub>T</sub> means that the supervenience base of the mental states of a subject will extend beyond the bodily boundaries of that subject. But in this case, EXTERNALISM<sub>T</sub> would be true and the following alternate argument ensues, on the assumption that the mentalist sticks to (1).

- (1) A subject  $S$ 's justification for her beliefs at  $t$  supervenes on her mental states at  $t$
- (2\*) A subject's mental states at  $t$  supervene on her internal physical state at  $t$  and some physical state at  $t$  external to  $S$
- (3\*) If  $X$  at  $t$  supervenes on the internal physical state of  $S$  at  $t$  and some physical state at  $t$  external to  $S$ —and  $Y$  at  $t$  supervenes on  $X$  at  $t$ —then  $Y$  at  $t$  supervenes on the internal physical state of  $S$  at  $t$  and some physical state at  $t$  external to  $S$

(4\*) Therefore, S's justification for her beliefs at  $t$  supervenes on her internal physical state at  $t$  and some physical state at  $t$  external to S.

This is essentially the structure of the argument I have made against mentalism in this chapter. On this alternative, it is not clear why mentalism is in any sense an internalist thesis. Originally, its internalist status was supposed to be inherited from the internalist status of mental states, which status is undermined by the truth of (2\*). EXTERNALISM<sub>T</sub> (which is equivalent to the negation of INTERNALISM<sub>T</sub>) is equivalent to (2\*). But INTERNALISM<sub>T</sub> is false (and hence both EXTERNALISM<sub>T</sub> and (2\*) true) if cognitive externalism is true, according to Bartlett. Hence, the truth of cognitive externalism undermines mentalism's status as an internalist thesis, as long as internalism about mental states is best understood as INTERNALISM<sub>T</sub>. Of course, Bartlett argues that this is the wrong understanding of content internalism. He gestures towards something like INTERNALISM<sub>C</sub> instead. How would we reconstruct mentalism given INTERNALISM<sub>C</sub>?

- (1) A subject S's justification for her beliefs at  $t$  supervenes on her mental states at  $t$
- (2\*\*) A subject S's mental states at  $t$  supervene on the causal basis of S's mind at  $t$
- (3\*\*) If  $X$  at  $t$  supervenes on the causal basis of S's mind at  $t$ —and  $Y$  at  $t$  supervenes on  $X$  at  $t$ —then  $Y$  at  $t$  supervenes on the causal basis of S's mind at  $t$
- (4\*\*) Therefore, S's justification for her beliefs at  $t$  supervenes on the causal basis for her mind at  $t$

Again, (1) is just the mentalist's supervenience principle (S). (2\*\*) is the improved articulation of content internalism which Bartlett favours, INTERNALISM<sub>C</sub>. The rest follows from parallel reasoning to that found in the reconstruction of mentalism based on INTERNALISM<sub>T</sub>. Does this capture the mentalist position? To assess this, it is important to

note why Bartlett takes INTERNALISM<sub>C</sub> to be consistent with cognitive externalism. Here is Bartlett.

For example, the reason Otto's case creates no trouble for [content] internalism is that (according to [cognitive externalism]) the entries in his notebook play a causal role we usually think of as belonging to mental states. If [cognitive externalism] is true of Otto's case, then his notebook is included in the causal basis of his mind just because certain of its states play a characteristically mental role in his activities (2008: 181).

The idea is this. A subject's mental states are causally efficacious with respect to her behaviour<sup>81</sup>. In order to be efficacious in this way, the content of mental states must not depend on factors outside the causal basis of said states. However, if cognitive externalism is true, then the information in Otto's notebook is part of the causal basis for his behaviour. (Recall that this is the extended functionalist motivation for cognitive externalism outlined in Chapter 1.) Thus, according to INTERNALISM<sub>C</sub>, the information in the notebook is part of the supervenience base of the mental content of Otto's mental states *even if* it is not a part of Otto construed as a bodily individual.

To make a parallel argument in the case of mentalism, it would be required that the mentalist took a subject's mental states to be causally efficacious with respect to the (doxastic) justification of her beliefs. However, to do this is to endorse a causal account of the epistemic basing relation<sup>82</sup>. The mentalist cannot rest her thesis on such an account of the basing relation. To do so would be to exclude from the epistemic internalist camp any account which has a non-causal account of the basing relation. However, some accounts of justification traditionally labelled as internalist promulgate doxastic accounts of the epistemic basing relation. That is, such accounts say that a belief that *p* is held by a subject for a reason, *r*, when the subject recognizes that *r* is a good reason to believe that *p*. Here recognition is

---

<sup>81</sup> Bartlett does not say whether he means behavior described narrowly (e.g., S extended her forward and tightened her grip) or broadly (e.g., S reached for the glass of water). This is often taken to make a difference as to what factors are relevant to a subject's behavior. As Bartlett does not make anything of this issue, neither will I.

<sup>82</sup> The epistemic basing relation is that relation which obtains between a fact, *x*, and a belief, that *p*, just in case a subject believes that *p* in virtue of *x*.

construed as a higher-order belief—a belief about another belief’s epistemic status. This is just the sort of account of epistemic basing we would expect from the access internalist sort of theory outlined in Chapter 1. However, recall that the mentalist *does not* want to exclude access internalist theories from counting as epistemic internalist positions. As we noted, access internalism is an ultra-strong version of mentalism, according to the mentalist. Hence, the mentalist cannot avail herself of a modified version of Bartlett’s argument to avoid the problems for her position which I claim arise from considering cognitive externalism. The mentalist is stuck with thinking of internalism about mental states in terms of INTERNALISM<sub>T</sub>. In which case, Bartlett-style arguments deliver the result that cognitive externalism poses a problem for the mentalist.

Note that cognitive externalism does not have to obtain in the actual world for the threat to mentalism to arise: it need only be the case that in some possible world, cognitive externalism obtains. Even if it is agreed that, in the actual world, none of the mental states to which the mentalist appeals extends beyond the bounds of relevant individuals, this would only secure that mentalism is *contingently* an internalist thesis. However, Conee & Feldman intend to state something *necessary* about the nature of justification. The status of matters of fact will be cold comfort to the mentalist.

Here is an example. Arl has a justified belief that it is raining outside. This belief is justified by his mental states, including the visual perception of non-frozen precipitation. For the mentalist, the perception of precipitation is something like the activity of the perceptual and sensory processing areas of the brain in response to some stimuli. Importantly, Arl need not know or believe anything about these mental states for them to justify his belief concerning the rain. Nor should any concern about the individuation conditions of the content of the beliefs arise. (Recall Conee has already given us a story about counterpart propositions to deal with content externalism.) These perceptual and sensory states do their justificatory work merely by their presence within Arl. These states are Arl’s evidence for the belief that it is raining. Intuitively, any mental duplicate of Arl, Vat-Arl, will have the same evidence as Arl. This is to say that Conee & Feldman accept the New Evil Genius (NEG) intuition. Recall the formulation of this intuition previously cited.

(NEG) The extent to which  $S$  is justified at  $t$  in believing that  $p$  is just the same as the extent to which  $S$ 's envatted duplicate is justified at  $t$  in believing that  $p$ . (Pritchard 2011: 365).

Conee & Feldman do accept that an envatted subject can have justified beliefs (2004: 60-61). Hilary is an unfortunate soul who has—after a long period of normal, embodied life in the actual world—had his brain removed and placed in a vat by an evil genius. Hilary's brain is stimulated in the vat so that it seems to him his normal life has continued unabated. When Hilary has a vivid recollection of breakfast, induced by the evil genius, Conee & Feldman state, "Although . . . Hilary's breakfast belief is false and its basis is abnormal, the belief . . . is well-justified in [this case]." Of course, this in itself does not show that Conee & Feldman accept NEG. It is consistent with Hilary being well-justified in his breakfast beliefs when envatted that he would be *better* justified if embodied. However, Conee & Feldman later indicate acceptance of NEG.

Since brains in vats or victims of deceptive demons adjust their subjective probabilities in light of their experiences in just the way in which normal people do, their experiences have the same (internalist) evidential value as the experiences of normal people (*op cit* 289).

Hence, if Vat-Arl and Arl have the same evidence, then they are justified in believing the same propositions and to the same extent. Neural duplication engenders mental duplication<sup>83</sup>. Sameness of mental state implies sameness of evidence, which in turn implies sameness of justification.

In the previous chapter, I suggested that the mentalist faced problems duplicating cases consistent with cognitive externalism which support the NEG intuition. I argued that a cognitive externalist way of individuating mental states by the actual objects with which a subject interacts, coupled with the acceptance of NEG, will leave the justificational similarity of duplicates unexplained. However, let us now assume that a mentalist can reproduce

---

<sup>83</sup>This claim is subject to qualification based on how one comes down in the content internalism-externalism debate. We have already explored Conee's thoughts on this problem, and I think it can safely be bracketed for my purposes.

mental duplication under the assumption of cognitive externalism. Even making this concession, cognitive externalism will not sit well with mentalism.

Consider an alteration of the Arl/Vat-Arl scenario above. Suppose Arl has a counterpart, Carl, who inhabits a world where visual perception and/or sensation are correctly characterized not merely as the states of his perceptual and sensory processing areas, but by dynamic interaction between Carl, his perceptual and sensory equipment (e.g., eyes and visual processing brain areas) and his surroundings (à la Noë 2004). Now take a mental duplicate of Carl, Carl\*. By near parallel reasoning from the case of Arl and Vat-Arl, Carl and Carl\* are justificational duplicates. The reasoning is near parallel because the setup of the world Carl and Carl\* inhabit rules out NEG as stated. If visual and perceptual states require there to be interaction with the environment, envatted scenarios will not predict mental similarity between normal agents and their envatted duplicates. That is, neural duplication will not effect mental duplication. However, the same result can be obtained by substituting envatting with something like a *Truman Show* scenario<sup>84</sup>. Carl and Carl\* will not differ by the former being embodied and the latter envatted, but instead by the former leading a normal life and the latter leading a scripted one. When Carl sees non-frozen precipitation coming from high above, he will form the justified belief that it is raining. Likewise, when Carl\* sees non-frozen precipitation coming from high above, he will form the justified belief that it is raining (even though the situation is actually cleverly counterfeited by the show's technical staff). This picture should be acceptable given merely the letter of mentalism.

The difference between the Arl scenario and the Carl scenario is that the mental is characterized as partly external to the body of the relevant agents in the latter, rather than internal to them as in the former. It is an assumption beyond brutling (S) and (M) that only states internal to bodily individuals will count as mental. However, there is a coherent position which rejects such an assumption. Now it cannot be the location of the mental which captures what is relevantly internal for epistemic internalism. But, if the location of mental states does not capture what is relevantly internal for epistemic internalism—and if

---

<sup>84</sup> See Hurley 1998b, chp.8, for a detailed discussion of why vehicle externalism allows duplicability thought experiments.

access requirements do not do so either, *ex hypothesi*—the mentalist lacks something she wants. She has no account of what is relevantly internal for epistemic internalism, which is to fail with regard to task (b) as outlined in Section 1. Moreover, it seems that the mentalist cannot be happy with (S) and (M) as stated because these principles are consistent with *epistemic* externalism, if the mental is allowed to range outside of the bodily individual. When the environment—natural, ecological or social—is allowed into the factors constitutive of the mental, mentalism becomes externalism in disguise. Or, at least there is now room for an externalist to take (S) and (M) on board and append her own extra criteria that she takes to be necessary for the presence of justification. In any case, recall that should mentalism fail to exclude some form of externalism, this is a failure by Conee's (and presumably Feldman's) own lights and a failure with respect to task (c), outlined in Section 1. However, mentalism only excludes all forms of reliabilism if it also stipulates the internality of the mental, which we have seen it cannot do.

This is a problem for the mentalist. It is the stipulation that mental states are internal to canonically delineated individuals which makes mentalism an internalist thesis, but (S) and (M) which make it an epistemological thesis. However, if the epistemological portions of mentalism are consistent with both epistemic internalism and externalism, mentalism as an interesting thesis becomes a position in the philosophy of mind and cognitive science rather than a position in epistemology. Specifically, the interesting claim mentalism makes is that the mental states relevant to justification must reside within the boundary of the canonically delineated individual.

### §3: Possible Mentalist Responses

How can the mentalist respond? If she is like Conee, she will not insist that her philosophical notion of perception and sensation supersedes what any scientific inquiry might discover. It was just such claims which led Conee to argue that access internalist positions burden themselves with unnecessary commitments (2007: 56). However, if what cognitive externalists claim is true, then the best scientific explanation of the mental might involve factors outside the canonical individual. In some cases, perhaps such explanations already exist. However, recall that even the possibility of cognitive externalism is a threat to



mentalism as ur-internalism. Hence, even if there are currently no mental phenomena best explained in a cognitive externalist fashion, the mentalist cannot rest easy. To be safe, she must deny the possibility of cognitive externalism, claiming that there are no worlds in which it obtains. Perhaps this can be done, but it is not clear how. Moreover, even opponents of cognitive externalism within philosophy of mind and cognitive science tend to agree that there is nothing obviously incoherent about the cognitive externalist thesis (Adams & Aizawa 2009: 25, 39, 47, 48 [see fn 32]; Rupert 2009: 10).

The mentalist might retreat to a more restrictive account of the mental, one on which the only truly mental states are those within the canonical individual. For this manoeuvre not to be *ad hoc*, however, there must be an independent motivation for it. To complicate matters for the mentalist, cognitive externalist arguments are given for more states than just dispositional beliefs. Many of the mental states which are argued to be external by cognitive externalists are often just those which are intuitively relevant to justification—*viz.*, beliefs, perceptual states, sensory states, etc. Thus for the mentalist to deploy a more restrictive account of the mental—one which bars appeal to possibly externalist-friendly states—requires claiming that many states which intuitively count as mental aren't really mental states!

But perhaps this interpretation is uncharitable. Instead of giving a more restrictive account of the mental, perhaps the mentalist means to appeal to a subset of the mental states—*viz.*, the ones relevant to justification. Conee says as much. “[Mentalism] implies nothing about whether any given mental states justify, or what propositions the justifying ones justify. The specifics are the job of specific internalist theories” (2007: 63). However, even interpreting the mentalist's position this way does not save her from having to expel beliefs from the supervenience base of justification. Surely, any theory of justification which appeals to mental states at all must at least classify beliefs as justificationally relevant mental states. Therefore, mentalism does imply something about the justification-conferring ability of beliefs at least. But as long as the mentalist allows dispositional beliefs into the supervenience base of justification, the considerations from Clark & Chalmers about Otto gain purchase, and the mentalist is stuck with the problem the possibility of cognitive externalism poses.

From here, the mentalist might attempt to restrict herself to only occurrent beliefs (and occurrent versions of any other state the dispositional varieties of which are cognitive externalist-friendly). This is what Clark & Chalmers suggest is the only principled way to escape the consequences of their argument, and this strategy is expanded upon by Brie Gertler (2007a). However, consider what motivates this move. Here are two quotes from Gertler.

Still, for one who thinks that *introspectibility* is crucial to our basic concept of the mind, this point will cast doubt on C&C's conclusion that the mind extends into the world. If one can introspect only the non-extended parts of the mind, then why count the external factors as truly part of the mind? (*op cit* 195 [original emphasis])

On this view, the mind is made up entirely of occurrent states and conscious processes. These include beliefs or desires that are now being entertained, conscious thoughts, emotions, and sensations, and conscious cognitive processes (*op cit* 203).

In philosophy of mind, this move makes facts of first-personal accessibility the mark of the mental. The analogous move in epistemology would be to make facts of accessibility the mark of justificatory potency. This cannot be what motivates the mentalist. Claiming that only occurrent states and conscious processes confer justification is just what the accessibilist does. Such commitments are precisely what mentalism is meant to avoid. She wishes to avoid them not only because she thinks that they are too committing, but also because they prevent her from accomplishing task (a) as outlined in Section 1. Her position needs to be distinct from accessibilism.

Because the mentalist eschews access requirements, there appear to be no well-motivated conditions distinguishing evidential mental states from non-evidential ones. Even if there were a non-trivial thesis separating the evidential from non-evidential mental states, and which claimed that the former were internal, this would be an argument about the nature of mental states rather than one about the nature of justification. However, mentalism is supposed to be about the nature of justification.

Perhaps all the mentalist wants to say is that justification supervenes on narrow content<sup>85</sup>. When understood this way, cognitive externalism is orthogonal to mentalism (on the assumption that cognitive externalism and content externalism are also orthogonal). However, it is not clear that this prevents mentalism from collapsing into access internalism, given that arguments from narrow content are sometimes motivated by the intuition that the contents of an agent's thoughts should be introspectively accessible to her (e.g., Loar 1988). Similarly, arguments in favour of narrow content from the intuition that an agent should be able, in principle, to discover when she is irrational, smack of access internalism (*ibid*). Taking either of these routes would prevent the mentalist from accomplishing task (a).

Finally, the causal arguments given in support of narrow content (Fodor 1987) seem to give no aid to the mentalist. If she were to appeal to this motivation for narrow content, the only apparent reason would be that she took justifiers to confer positive epistemic status on beliefs by causing those beliefs. That is, the mentalist would be giving a causal account of the epistemic basing relation. However, the mentalist should not want to take on the extra commitment of defending a causal account of the epistemic basing relation in order to avoid the problem posed by cognitive externalism. The causal account faces the same serious difficulties which afflict causal accounts in general in philosophy—*viz.*, deviant causal chains (see, for example, Plantinga 1993). More importantly, it seems that mentalism should leave internalists free to pursue doxastic accounts of the basing relation as well, in order that strong accessibilist theories still count as internalism<sup>86</sup>. Recall that mentalism is supposed to classify as internalist all epistemic positions which intuitively merit the distinction. This is one of the vaunted advantages of mentalism over accessibilism. Giving this up is to fail with respect to task (d).

In short, the burden is on the mentalist opting for the narrow content strategy to provide a new, principled motivation for adopting the strategy. She cannot appeal to narrow content merely because it helps her avoid the problem posed by cognitive externalism. Further, none of the usual motivations for appealing to narrow content are available to the

---

<sup>85</sup> Thanks to Trent Dougherty for pointing out this response on behalf of the mentalist.

mentalist without compromising her claim that mentalism is both distinct from, and less committing than, accessibilism. More generally, the success of this narrow content strategy is predicated upon the success of an account which makes narrow content intelligible, and the intelligibility of narrow content is a lively debate in its own right. The mentalist should not want to stake the success of her position on another controversy. Further, recall our discussion of Bartlett's paper earlier in the chapter. Here we saw that the mentalist was stuck with a construal of internalism about the content of mental states which, according to Bartlett, is undermined by the truth of cognitive externalism. Thus, even setting aside the particular worries just adumbrated, there are general considerations which tell against a narrow-content strategy for the mentalist.

Finally, consider a dialectical point. As far as the arguments in this paper go, the mentalist is appealing to narrow content in response to the threat from cognitive externalism. Faced with cognitive externalism (either as a thesis about the actual world or some possible world), the move to narrow content may already be blocked. Numerous philosophers argue that content externalism does not entail cognitive externalism (e.g., Rupert 2004; Hurley 2010; Clark & Chalmers 1998). However, some argue that cognitive externalism does *entail* content externalism. Take Robert Wilson (2004), for example.

Locational externalism [cognitive externalism] is a stronger view than taxonomic externalism [content externalism] and entails it unless there is a plausible, nonquestion-begging [sic] way to individuate mental states independent of their total realizations (179).

Even should claims like Wilson's be shown ultimately to fail, it seems that the mentalist should not have to take on such strong philosophical commitments outside of the epistemological domain.

However, even granting the success of a narrow content strategy for the mentalist, the resultant position would be significantly different from the mentalism put forth by Conee

---

<sup>86</sup> Also, note the parenthetical in Conee & Feldman: "Actually, one of us doubts that causation is necessary for basing, but we can set this doubt aside for [discussing the generality and sorting problems]" (2004: 164).

and Feldman. Put simply, mentalism as originally stated claims that justification supervenes on the mental life of individuals. Appealing to both occurrent and dispositional mental states, events and conditions is to appeal to the mental broadly construed. It is not clear that it will make sense to attribute the sort of content under consideration to all mental states, processes, events and conditions. And as Conee makes clear, it is only the appeal to the *broadly* mental which correctly classifies as epistemic internalist all positions which intuitively merit the distinction. Thus, the narrow content strategy forfeits one of the vaunted benefits of mentalism as Conee and Feldman conceive it, and prevents the mentalist from accomplishing task (d).

#### §4: Summary

Mentalism is touted by its proponents as epistemic internalism *par excellence*. However, I have argued that the possibility of cognitive externalism obtaining in some possible world renders mentalism only contingently internalist. Moreover, there appear to be no ways to obviate this problem which do not involve *ad hoc* manoeuvres, do not surrender the mentalist's fortunes to the success of other contentious philosophical positions, or do not give up on the particular benefits gained by accepting mentalism in the first place. Mentalism is not epistemic *ur*-internalism. More to the point, mentalism does present something that the epistemic externalist is bound to reject. If (S) and (M) are consistent with epistemic externalism as I have claimed, then it is open to an externalist theorist to adopt these principles. But, if the epistemic externalist does not have to reject mentalism, then my argument to the effect that the epistemic externalist is forced to accept cognitive externalism is undermined. Recall that this chapter has been an extended discussion of an objection to the second premise of my argument from the previous chapter. I will reproduce the argument here.

(1) Epistemic externalism entails the rejection of epistemic internalism.

[*Definition*]

(2) Mentalism is a form of epistemic internalism. [*Premise*]

(3) Therefore, epistemic externalism entails the rejection of mentalism. [*From 1 & 2*]

(4) An account of justification which appeals to cognitive abilities, where those abilities are exhaustively constituted by factors internal to the individual of whom said abilities are predicated, is consistent with mentalism. [*Premise*]

(5) Cognitive internalism is the claim that cognitive abilities are exhausted by factors internal to the individual of whom said abilities are predicated.

[*Definition*]

(6) Therefore, an account of justification which appeals to cognitive abilities and is consistent with mentalism implies cognitive internalism. [*From 4 & 5*]

(7) An epistemic externalist account of justification must appeal to cognitive abilities. [*Premise*]

(8) If an account of justification which appeals to cognitive abilities and is consistent with mentalism implies cognitive internalism, then epistemic externalism must reject cognitive internalism. [*From 3 and 7*]

(9) Therefore, epistemic externalism must reject cognitive internalism. [*From 6 & 8*]

(10) The rejection of cognitive internalism entails the acceptance of cognitive externalism. [*Definition*]

(11) Therefore, epistemic externalism implies cognitive externalism. [*From 9 & 10*]

(12: C) Therefore, an epistemic externalist account of justification must appeal to cognitive abilities as envisioned by cognitive externalism. [*From 7 & 11*]

By assuming (2) and taking it together with (1)—which is definitive of epistemic externalism—I derived (3), which claimed that the epistemic externalist was required to reject mentalism. Importantly, (3) is then used to derive (8), which when taken together with (6) led to (9). In turn, (9) taken together with the definition of cognitive externalism led to (11). Finally, (11) was crucial in deriving the ultimate conclusion of the argument.

Considering that (3) does so much work in my argument, and that further (3) is clearly dependent upon (2), a successful objection to (2) saps the foundations of my argument. In this chapter, I take myself to have presented just such an objection.

But the effort to demonstrate a commitment on the part of epistemic externalism to cognitive externalism is not yet undone. In the next (and final) chapter of this work, I will argue that the epistemic externalist ought to reject mentalism anyway. The argument provided for this conclusion will, like its vanquished predecessor, provide a way to demonstrate a commitment by the epistemic externalist to cognitive externalism.

## Chapter 6

### From Epistemic Anti-Individualism to Cognitive Externalism

#### *Introduction*

In Chapter 3, we saw and immediately dismissed one argument for a commitment to cognitive externalism on the part of the epistemic externalist. In Chapter 5, the second (and more serious) argument for such a commitment from Chapter 4 was refuted. This argument failed because it relied on the assumption that mentalism was a form of epistemic internalism, which assumption we have seen was unwarranted. The argument constructed in Chapter 3—half-hearted as it was—failed because it equivocated between two senses of ‘depend’: *viz.*, a merely individuating sense and a constitutive sense.

The grounds on which the argument of Chapter 4 tried to demonstrate the epistemic externalist’s commitment to cognitive externalism were more specific than the grounds used in the argument of Chapter 3. There, we attempted to link epistemic and cognitive externalism merely via the notion of ability used by epistemic externalists. In Chapter 4, we added to the grounds of our appeal the epistemic externalist’s obligation to deny epistemic internalism. The argument I will construct in this chapter will adduce yet more specific grounds. In addition to the appeal to cognitive abilities and the obligation to deny epistemic internalism, I will now point to the arguments in favour of *epistemic anti-individualism* to make a case for the commitment to cognitive externalism on the part of the epistemic externalist.



The contrast between epistemic individualism and epistemic anti-individualism is due to Sandy Goldberg (2007). In what follows, I will not present any novel arguments for epistemic anti-individualism. Instead, I give a sketch of Goldberg's arguments where the details are relevant to constructing my argument concerning the commitments of epistemic externalism. On the assumption that Goldberg's arguments are successful, I will mount a final attempt to demonstrate the epistemic externalist's commitment to cognitive externalism. This argument will be in contravention of Goldberg's explicit claim (2007: 237, FN 34) that the position he argues for *does not* involve any commitment to cognitive externalism. We will discuss why Goldberg claims that no such commitment ensues in due course. First, we must understand the distinction between individualism and anti-individualism in epistemology.

The terms 'individualism' and 'anti-individualism' appear first in the philosophy of mind debate over mental content in the classic papers of Burge (1979; 1986), in which he defends anti-individualism. There the idea is that the property of *having a thought with the content that p*<sup>87</sup> depends not only on how things are with the individual of whom the property is predicated, but the community of other individuals in which the relevant individual is situated. Anti-individualism is the denial of individualism, which claims that the property of *having a thought with the content that p* depends only on how things are with the individual of whom the property is predicated. We briefly discussed Burge's primary example in Chapter 3, when broaching Clark & Chalmers' distinction between passive and active externalism.

Here is Burge's example. Given a pair of subjects who are physical duplicates of each other, if one (Fred) is embedded in a community where 'arthritis' denotes a disease of the joints, and the other (twin Fred) is embedded in a community where 'arthritis' denotes a disease of the joints *and* bones, the two think different thoughts when they think something which they would formulate as "That person has arthritis". The first subject, Fred, possesses the concept 'arthritis' and hence his thoughts of the form "That person has arthritis" concern arthritis. This is the case even if Fred (mistakenly) believes that arthritis is an ailment of the joints and bones, as he is a member of (and, importantly, deferent to) a speech

---

<sup>87</sup> I take this formulation of the property debated in the literature on mental content from Brie Gertler (2007b).

community in which ‘arthritis’ refers to an ailment of the joints only. Fred’s membership in (and deference to) the community standard meaning of ‘arthritis’ shape the contents of his thoughts, and the meanings of his utterances featuring ‘arthritis’ even when he has certain mistaken beliefs about the phenomenon.

Twin-Fred, on the other hand, as a member of the community which uses ‘arthritis’ to refer to an ailment of the joints *and* bones, does *not* possess the concept ‘arthritis’. Hence, his thoughts of the form “That person has arthritis” do not concern arthritis. Rather, Twin-Fred has the concept ‘tharthritis’ and his thoughts concern an ailment which affects the joints and bones—*viz.*, tharthritis. This is not because of some behavioural or physical feature of Twin-Fred which distinguishes him from Fred. Indeed, *ex hypothesi*, there are no such differences. Instead, Twin-Fred’s thoughts and utterances have different contents and meanings in virtue of facts about the community of other individuals in which he is embedded. This is the upshot of Burge’s anti-individualism as it concerns mental content and semantics.

Goldberg, in the first part of his 2007 monograph (*Anti-Individualism: Mind and Language, Knowledge and Justification*)<sup>88</sup>, deals with anti-individualism as it concerns semantics and philosophy of mind. These issues do not concern us here, so we will move directly to Goldberg’s discussion of anti-individualism in epistemology. Like anti-individualism about mental content, anti-individualism in epistemology is the claim that certain interesting properties of individuals depend on facts about individuals *other than* the one of whom the properties of interest are predicated. For Goldberg, the properties of interest are knowledge, warrant and justification. Thus, anti-individualism in epistemology is the claim that “ascription of knowledge [,warrant] and justification to a subject *S* sometimes depends on factors pertaining to the *cognitive lives of subjects other than S*” (134) [original emphasis]<sup>89</sup>. This thesis is posed as the rejection of epistemic individualism, which claims that the ascription of epistemic statuses to an individual depends only on factors of *that* individual’s cognitive life.

---

<sup>88</sup> Unless otherwise noted, all citations of Goldberg will be from the aforementioned book.

<sup>89</sup> The argument Goldberg presents is specifically an account of the epistemic statuses associated with testimony-based belief. Hence, it is the thesis that “one’s social environment [is] relevant to the *epistemic statuses* enjoyed by the beliefs one forms through accepting the word of one’s co-linguals” (*ibid*) [original emphasis].

Immediately, it is important to note that anti-individualism entails the rejection of mentalism. Recall, the characteristic supervenience thesis of mentalism.

(J) The justificatory status of a person's doxastic attitudes strongly supervenes on the person's occurrent and dispositional mental states, events and conditions (Conee & Feldman 2004: 56).

In (S), it is explicitly stated that an individual's justificatory status supervenes only on *that individual's* properties. This is exactly what the individualist in philosophy of mind says about mental content. Mentalism, then, is an epistemic individualist thesis. Hence, any thesis committed to epistemic anti-individualism will have to deny mentalism. Likewise, an epistemic anti-individualist will plausibly be committed to denying accessibilism. Recall the characteristic accessibilist thesis.

*Simple Internalism* (SI) Whether one is justified in believing P supervenes on facts which one is in a position to know by reflection alone (Pryor 2001: 104).

On the assumption that some factors in an individual's social environment are not amenable to access by reflection (specifically, the cognitive goings-on of other people), the epistemic anti-individualist is obliged to reject (SI) insofar as she allows the factors in an individual's social environment justificatory potency.

So far, the distinction between epistemic individualism and anti-individualism recapitulates the internalism-externalism distinction in epistemology. However, Goldberg notes that the former distinction actually cuts across the latter. According to Goldberg, epistemic externalists can be epistemic individualists.

Then we can call an [epistemic] externalism *individualistic* when its focus on external factors is restricted to the functioning of the agent's own cognitive system and its relations to her non-social environment. (That is, an epistemically externalist position is individualistic when it does not include factors pertaining to the cognitive lives of subjects other than the one currently under epistemic assessment) (135 – 36) [original emphasis].

Hence, a commitment to epistemic anti-individualism involves rejecting all epistemic internalist positions as well as some epistemic externalist ones.

Immediately, an opportunity to amend Chapter 4's argument for a commitment to cognitive externalism on the part of the epistemic externalist presents itself. The flaw in that argument was that there were strong considerations in favour of rejecting one of its premises. That premise, (2), concerned the status of mentalism as an internalist thesis. Deriving a commitment to cognitive externalism on the part of the epistemic externalist hinged on her obligation to reject what was consistent with her epistemic internalist opponent's position on the epistemic status at issue. Had mentalism turned out to present an epistemic internalist thesis, the argument of Chapter 4 would have gone through. However, as we saw in Chapter 5, mentalism does not present an epistemic internalist thesis.

Now consider epistemic anti-individualism. As a form of epistemic externalism, it still obliges its proponents to reject epistemic internalism. But the epistemic anti-individualist also must reject certain forms of epistemic externalism, as we have seen. Hence, if the argument of Chapter 4 could be repaired by centring it around the distinction between epistemic individualism and anti-individualism—rather than the epistemic internalist-externalist distinction—a new version of (2) would be immune to objection. As we have already noted, the mentalist supervenience thesis is explicitly an individual supervenience thesis. Specifically, it is an individual supervenience thesis making claims about one of the very epistemic statuses the anti-individualist wants to give an account of—*viz.*, justification. Hence, the mentalist and the anti-individualist concerning justification are at odds in the same way I presented the mentalist and the epistemic externalist concerning justification as being. Thus, it is plausible that an argument of roughly the same structure as the one in Chapter 4 should be good. In the next section, I will construct a revised argument. Both the argument and the way I present it will be largely familiar from Chapter 4.

### *§2: A Way from Epistemic Externalism to Cognitive Externalism—Revisited*

Now we are in a position to try once more to formulate an argument to the effect that the epistemic externalist is committed to cognitive externalism. Strictly speaking, this argument affects only anti-individualist epistemic externalists. However, I take Goldberg's arguments

in his 2007 book aim at forcing us into just such an epistemic externalism. My argument will assume the success of Goldberg's. Because of the similarity between this argument and the earlier argument from Chapter 4, I will number it similarly. Any premise which has been changed will be indicated with a '\*' beside the number which indexes it. Here is the first group of premises.

- (1\*) Epistemic anti-individualism entails the rejection of epistemic individualism. [*Definition*]
- (2\*) Mentalism is a form of epistemic individualism. [*Definition*]
- (3\*) Therefore, epistemic anti-individualism entails the rejection of mentalism. [*From 1\* & 2\**]

Epistemic anti-individualism is defined as the rejection of epistemic individualism (Goldberg 2007: §5.1). Hence, (1\*) is a definition. We have already discussed (2\*) earlier in this chapter. Because the characteristic mentalist supervenience thesis explicitly ties the supervening properties of epistemic interest to a subvenient base of properties comprising only facts about the individual of whom the interesting epistemic properties are being predicated, mentalism is an epistemic individualist thesis. Thus, (2\*) is also a definition. (3\*) follows straightforwardly from (1\*) and (2\*). Now, on to the second group of premises.

- (4) An account of justification which appeals to cognitive abilities, where those abilities are exhaustively constituted by factors internal to the individual of whom said abilities are predicated, is consistent with mentalism. [*Premise*]
- (5) Cognitive internalism is the claim that cognitive abilities are exhausted by factors internal to the individual of whom said abilities are predicated. [*Definition*]
- (6) Therefore, an account of justification which appeals to cognitive abilities and is consistent with mentalism implies cognitive internalism. [*From 4 & 5*]

Each of these premises is unchanged from Chapter 4. Thus, the arguments for them are also unchanged. Recall that (4) was motivated first by noting that Linda Zagzebski's account of justification is an internalist account which makes an appeal to abilities. From her account, we formulated a general accessibilist position which appealed to cognitive abilities.

Subsequently, though we saw there was no extant formulation of mentalism which appealed to cognitive ability, we discussed what a mentalist appeal to cognitive ability would be like, and concluded that such an appeal was coherent. Importantly, any possible mentalist appeal to cognitive abilities would have to construe those abilities in the manner indicated by (4).

Given this, and the definitional (5), (6) follows easily. Let us move on to the next group of premises.

(7\*) An epistemic anti-individualist account of justification must appeal to cognitive abilities. [*Premise*]

(8\*) If an account of justification which appeals to cognitive abilities and is consistent with mentalism implies cognitive internalism, then epistemic anti-individualism must reject cognitive internalism. [*From 3\* and 7\**]

(9\*) Therefore, epistemic anti-individualism must reject cognitive internalism. [*From 6 & 8\**]

(10) The rejection of cognitive internalism entails the acceptance of cognitive externalism. [*Definition*]

(11\*) Therefore, epistemic anti-individualism implies cognitive externalism. [*From 9\* & 10*]

(7\*) is still supported by the arguments of Pritchard et al. (2010) and Pritchard (forthcoming). (8\*) is an instance of (3\*), when taken together with (7\*). The motivation for (8\*) is the same as for (8). While there are many things upon which an epistemic individualist and an epistemic anti-individualist can agree, they cannot agree on substantive features of the account of the epistemic status at dispute between them. In this case, that status is justification. On the assumption of (7\*), an appeal to abilities is a substantive feature of the anti-individualist's account of justification.

Given (6) and (8\*), we can conclude that the epistemic anti-individualist must reject cognitive internalism and thus we get (9\*). That is, the epistemic anti-individualist concerning justification must reject those things which her opponent invokes in giving an opposed account of justification. (10), unchanged from Chapter 4, remains a definition and licenses the move from (9\*) to (11\*). However, just as in Chapter 4, (11\*) is not quite what the argument is after. Thus, taking (11\*) together with (7\*), we get the conclusion of our repaired argument.

(12\*: C) Therefore, an epistemic anti-individualist account of justification must appeal to cognitive abilities as envisioned by cognitive externalism.

[From 7\* & 11\*]

One might look askance at this argument. After all, what I have done so far is just substitute every instance of ‘epistemic internalism’ and ‘epistemic externalism’ for ‘epistemic individualism’ and ‘epistemic anti-individualism’, respectively. Moving words around is easy. However, because epistemic anti-individualism and epistemic externalism are not identical theses, might not this sort of substitution invalidate the argument? I do not think the argument is invalidated generally by the substitutions. But recall that in Chapter 4, there were two objections to the original argument’s validity centring around (4) and (8). Once again, (4) can be objected to here, as can the modified (8\*). However, it might be thought that I will have to give different answers to these objections. If so, it will be because the epistemic anti-individualist rejects mentalism *qua individualism*, rather than *qua internalism*. Let us look to these renewed objections.

### §3: *Objections and Replies*

I have changed the distinction around which my argument for a commitment to cognitive externalism on the part of epistemic externalists centres. Does the switch to talking in terms of epistemic individualism and anti-individualism affect what I can say about objections to (4) and (8\*)?

*§3.1: Objection to (4)*

The objection to (4) was that, for it to function as I need it to in my argument, a further assumption must be made. This charge can be renewed with respect to the repaired version of the argument as well. Specifically, the complaint is that I need it to be the case that mentalism is not only consistent with cognitive internalism, but also inconsistent with cognitive externalism. If mentalism were consistent with both, then my current argument would suggest that epistemic anti-individualism must reject both cognitive internalism and cognitive externalism. But this is absurd.

Part of the rebuttal I gave in Chapter 4 to this objection remains feasible. With respect to Conee & Feldman's particular formulation of mentalism, there is an ensconced commitment to cognitive internalism. In numerous places, Conee & Feldman indicate that they take the mental states upon which justification is said to supervene to be internal to bodily individual (2004: 57; 82). Nothing new need be said here.

However, what about mentalism more generally? The further part of the original objection to (4) asked whether a mentalist could divest herself of Conee & Feldman's commitment to cognitive internalism. My further response was to point out that taking cognitive externalism on board would create tension with general epistemic internalist intuitions which any mentalist would want to preserve. Specifically, I argued that allowing in environmental and social factors was anathema to the internalist intuition (pithily encapsulated by the New Evil Genius case) that justifiers should do their justificatory work non-contingently. It was not obvious that the mentalist would be able to reconstruct NEG cases after accepting cognitive externalism.

In the current dialectical situation, however, the mentalist's aspirations to epistemic internalism have been quashed. More exactly, in Chapter 5 I argued that mentalism is only a form of epistemic internalism if it sacrifices that which distinguished it from accessibilism. Indeed, we would not have need of a revised argument if mentalism had successfully usurped accessibilism as epistemic ur-internalism. This being the case, it can be objected that it is inappropriate to continue arguing that mentalism is inconsistent with cognitive externalism on the basis of mentalism's epistemic internalist commitments. As was just noted, mentalism is only an internalist thesis if it collapses into accessibilism. However, in



Chapter 4 I argued that no commitment to cognitive externalism followed from the denial of accessibilism. But in the current dialectical situation, arguing that mentalism *qua internalism* is inconsistent with cognitive externalism is the same as arguing that accessibilism is inconsistent with cognitive externalism. Because I have argued that accessibilism is not inconsistent with cognitive externalism, if I am to demonstrate that mentalism cannot free itself of Conee & Feldman's commitment to cognitive internalism, I cannot adduce any considerations related to epistemic internalist intuitions. Instead, I need to show that epistemic individualism is inconsistent with cognitive externalism.

Let us remind ourselves of the disagreement between epistemic individualists and their anti-individualist counterparts. Suppose we want to know about the epistemic statuses of Fred's beliefs. According to the epistemic individualist, the only factors relevant to our investigation of whether or not some of Fred's beliefs are justified (for example) are those pertaining to Fred's cognitive life. That is, it does not matter what is going on, cognitively speaking, with Jethro or anyone else in Fred's social environment. Fred's beliefs, the functioning of his cognitive processes and what-have-you, are that which fully determine whether Fred's beliefs are justified. On the other hand, the epistemic anti-individualist claims that whether or not Fred's beliefs are justified may depend on factors pertaining to Jethro's cognitive life. Throughout his book, Goldberg argues for anti-individualism via cases of testimony-based beliefs. In such cases, Goldberg says that a hearer demonstrates "epistemic reliance" (2007: 16) upon the testimony of others and that the testimony of others "epistemically supports" (*op cit* 33) the hearer's testimony-based beliefs. Hence, if Fred comes to believe that it is raining outside because Jethro tells him that it is raining outside, whether or not Fred's belief that it is raining outside is justified, will depend on facts about Jethro's cognitive life. (For instance, Jethro needs to be honestly asserting the proposition in question. He also needs to be reliable with respect to discerning the state of the weather.) (See notes 78 and 79 below for details on the requirements on testimony-based belief.)

What do I need to do to demonstrate the incompatibility of epistemic individualism with cognitive externalism? Given the discussion above, I need to show that cognitive externalism allows an appeal to the cognitive lives of others which epistemic individualism

rules out. That is, it must be consistent with cognitive externalism that the cognitive goings-on of some individual in part constitute the cognition of *another individual*.

Such cases are allowed for by some of the cognitive externalist positions we examined in Chapter 2. In fact, Clark & Chalmers suggest examples where what goes on in one person's head partly constitutes the cognition of another person.

What about socially-extended cognition? Could my mental states be partly constituted by the states of other thinkers? We see no reason why not, in principle. In an unusually interdependent couple, it is entirely possible that one partner's beliefs will play the same sort of role for the other as the notebook plays for Otto. What is central is a high degree of trust, reliance and accessibility . . . In other cases, one's beliefs might naturally be seen to be embodied in one's secretary, one's accountant, or one's collaborator in similar way (1998: 17 – 18).

Rob Wilson's account of extended cognition likewise allows for the sort of socially-extended cognition. If part of the core realiser of one of Fred's cognitive states is one of Jethro's cognitive states, then that state of Fred is radically widely realised, according to Wilson. Moreover, these are the sorts of state which we saw Wilson's account used as examples of cognitive externalist states.

Given that we are concerned with the "broad church" notion of cognitive externalism, these possibilities are sufficient to demonstrate that epistemic individualism is inconsistent with some form of cognitive externalism. This is all I need for the purposes of my argument. Thus, I have met the objection to (4).

### §3.2: *Objection to (8\*)*

In Chapter 4, I brought up a possible objection to (8)—*viz.*, that it was illicit to treat (8) as an instance of (3). A similar objection can be brought against (8\*). In the current argument, I treat (3\*) as an instance of (8\*). But perhaps this is an illicit move. One might object, once again, that any implication from mentalism to cognitive internalism is irrelevant. What is at issue between the mentalist (*qua individualist*) and the anti-individualist is not whether abilities

are broadly or narrowly constituted, but whether any of the epistemic statuses of one agent's beliefs can be affected by the cognitive lives of other agents.

In response to the similar objection made to the argument of Chapter 4, I drew on our lengthy exploration of the notion 'ability' from Chapter 3. I argued that the apparent agreement upon the nature and possession conditions of ability between the mentalist and epistemic externalist was an illusion. Given this, I proceeded to discuss the different options respectively available to robust and modest versions of virtue epistemology for making a distinctively externalist appeal to abilities. Ultimately, on the basis of the epistemic twin earth case from Kallestrup & Pritchard, I concluded that accounts of epistemic statuses which appealed only to an ability condition seemed to have no option but to adopt some form of cognitive externalism to do all of the epistemic work required of their theory and remain distinctively externalist (in the epistemological sense).

The problem posed by the version of the objection at hand now appears more easily solved. In response to the objection to premise (4) in the previous section, we noted that numerous forms of cognitive externalism allow for socially-extended cognition: i.e., instances wherein one agent's cognition is in part constituted by what goes on in another agent's head. Given this, to differ regarding whether or not one individual's epistemic statuses may be affected by factors pertaining to the cognitive lives of others may just be to differ with respect to whether cognition is broadly or narrowly constituted. Whether this is the case hangs on the nature of the dependence of one agent's epistemic statuses on the cognitive lives of others. However, as we saw in the argument we dismissed in Chapter 3, it is easy to equivocate between different senses of 'depend'. To make good on the response to the objection presented here, I will need to discuss which type of dependence is at issue in the anti-individualist's thesis. Thus, we will proceed, leaving the objection to (8\*) unresolved for the moment.

§4: Case Study—Goldberg’s Epistemic Anti-Individualism

At the end of his 2007 book, Goldberg claims that his arguments lead towards an “active anti-individualism” (237) concerning the epistemic statuses of testimony-based belief<sup>90</sup>. Specifically, he concludes this on the basis of considering the consumption of testimony by young children. As Goldberg puts it, “We might then say that the processing relevant to an epistemic evaluation of a very young child’s testimonial beliefs is *‘extended’ into the social world itself*” (*ibid*) [original emphasis]. This is claimed because young children are apparently not able to meet on their own the conditions Goldberg lays out for justified acceptance of testimony or knowledgeable testimony-based belief<sup>91</sup>. In particular, young children are supposed to lack the ability to discriminate reliable testimony (or testifiers) from unreliable testimony (or testifiers). Insofar as we can attribute justified acceptance of testimony and knowledgeable testimony-based belief to children, we can do so only because of work done by the adults responsible for raising the children. (We shall stick with the term ‘parents’ to refer to these adults, as Goldberg does, for simplicity.) Dutiful parents will construct their child’s epistemic environment in such a way as to ensure their children only interact with reliable testifiers (211). Thus, the epistemic statuses of a young child’s testimony-based beliefs are dependent on parents’ discrimination amongst testifiers and testimony.

---

<sup>90</sup> For Goldberg, a belief counts as testimony-based when the hearer comes to believe some proposition on the basis of the content of an utterance by a speaker, where the content of the utterance is the same as the content of the relevant testimony-based belief formed by the hearer (2007: §1.2). Imagine Fred says—with chattering teeth and shivering limb—“I am cold” to Jethro, who subsequently forms the belief that Fred is cold because of the chattering and shivering exhibited during the utterance. Jethro’s belief would not be a testimony-based one, even though it was acquired because of something Fred said. This belief is not testimony-based even if Jethro understood Fred’s utterance and Fred intended to convey that he was cold to Jethro by uttering “I am cold”, because Jethro’s belief is not held in virtue of the propositional content of Fred’s utterance. Jethro’s belief is testimony-based if he forms the belief that Fred is cold on the basis of hearing and understanding Fred’s utterance with the propositional content Fred is cold.

<sup>91</sup> In normal cases, the conditions set by Goldberg on testimony-based knowledge are: (a) the *source testimony* itself was reliable, in the sense that the testimony would have been proffered only if true; (b) the hearer’s recovery of the attested proposition was based on a *reliable comprehension process*; and (c) the hearer’s acceptance of the testimony was the upshot of a reliable capacity to *distinguish reliable attestations from unreliable ones* (2 – 3) [original emphasis]. Condition (c) is the relevant one in his argument for so-called active anti-individualism.

Immediately, Goldberg's talk about processes extending into the social world smacks of cognitive externalism. In particular, his choice to dub his anti-individualism "active" (as opposed to "passive") harkens back to Clark & Chalmers' distinction between active and passive externalism. Indeed, Goldberg acknowledges Clark & Chalmers' 1998 paper as the inspiration for this turn of phrase. However, in the same breath, Goldberg explicitly distances himself from cognitive externalism.

I borrow the idea of 'extended' cognitive processing from Clark and Chalmers 1998. The foregoing proposal can be seen to illustrate the *epistemic* dimension of extended processing. This said, the foregoing proposal does not imply the thesis of extended cognition. Granted that the adult's cognitive processing is relevant to the epistemic assessment of the child's testimonial belief, this processing need not be counted as part of the *cognitive* process that eventuates in the child's testimonial belief—any more than the broadly environmental regularities exploited by the perceptual process are part of the cognitive process that eventuates in perceptual belief (237, FN 34) [original emphasis].

In our terms, Goldberg seems to be denying that a young child's testimonial beliefs depend on the cognitive lives of their parents in the right way to engender a cognitive externalist case. Specifically, what he seems to claim by saying, ". . . this processing need not be counted as part of the *cognitive* processing that eventuates in the child's testimonial belief . . ." is that the elements of the parents' cognitive lives do not play the same sort of functional role Clark & Chalmers claim for Otto's notebook. Goldberg would thus also likely deny that any part of the core realisers for any of a young child's mental states is partly inside the heads of the child's parents. That is, Goldberg would also claim to leave untrodden Rob Wilson's path to cognitive externalism.

To clarify this somewhat, let us turn to a non-epistemic example. Goldberg invites us to imagine a teenager, Samantha, who has just gotten a driver's license (219 – 20). However, Samantha's license comes with restrictions on when and with whom she can drive. For instance, she cannot drive unless an experienced driver sits in the front passenger's seat. Samantha herself is impatient, inattentive and reckless. Thus, she is prone to ignoring traffic

signals, to speeding and to driving in inclement conditions. On the other hand, the people who might sit in the front passenger's seat while she drives (her mother, older sister or father) are all cautious and attentive while driving. Hence, whenever Samantha is driving, her passenger will correct her bad habits. Further, Samantha is simply not allowed to drive on the freeway where the traffic is denser and the other drivers are as poor as is Samantha.

What do we say about Samantha's driving—is she a safe driver? In one sense, obviously not: we have stipulated she has poor habits. That is, construed *individualistically*, Samantha is not a safe driver. However, Goldberg notes that, despite this fact, we can say that there are no (or not many) nearby possible worlds in which Samantha drives unsafely, and this because the limitations of her license ensure that she will not be in environments where her bad driving will come into effect. That is, viewed from an anti-individualist perspective, Samantha is a safe driver.

We can recast this in the distinctions made in Kallestrup & Pritchard, which we saw in Chapter 3. Imagine Samantha and her counterpart on twin earth. According to the details of Goldberg's case, for Samantha there are no nearby worlds in which she drives unsafely. In terms of Kallestrup & Pritchard's terminology, this means her global environment is one which is favourable to safe driving. (Recall that global environments fix the structure of nearby worlds.) This is the case because Samantha is required to drive with an experienced driver, who will prevent her from driving dangerously. Samantha's twin is a physical and psychological duplicate of Samantha, and thus has the same bad driving tendencies. However, suppose that on twin earth, twin Samantha's driver's license lacks the restrictions concerning with whom she can drive (or, the "experienced" drivers with whom she might drive are as bad as she). According to Goldberg, in this case there are nearby worlds in which twin Samantha drives unsafely. Thus, we should conclude that her global environment differs from that of Samantha on earth. However, recall Kallestrup & Pritchard's criteria for ability possession. "S possesses [an] ability in virtue of relevant bodily/psychological factors and mostly occupying an environment that is conducive for [S] to frequently [succeed at the goal of the relevant ability]" (7). If Samantha and her twin are bodily/psychologically identical but mostly occupy different (driving) environments—i.e., their respective global environments differ—then they do not possess the same (driving) ability.

Now apply this to the case of testimony consumption by small children. Goldberg discusses developmental psychology literature which indicates that young children are particularly credulous when it comes to accepting testimony (§8.2 - §8.5). They lack the ability to reliably discriminate reliable from unreliable testimony (or testifiers). Because they thus fail one of Goldberg's conditions of for gaining testimonial knowledge<sup>92</sup>, young children considered individualistically (i.e., as isolated agents) are not capable of acquiring testimonial knowledge. Nor are they capable of being justified in accepting testimony, given that they fail to fulfil the conditions by failing to properly discriminate reliable from unreliable testimony. However, if we include the work done by a young child's parent to ensure that their child is exposed only to reliable testimony (or testifiers), that child's credulity will not prevent them from fulfilling Goldberg's criteria. In such cases, children *will* be able to acquire knowledge from testimony and will also be able to justifiably accept testimony offered them<sup>93</sup>. Moreover, this result is supposed to be intuitively plausible. That is, if our theory of testimonial knowledge and justification barred young children from acquiring positive epistemic status for their beliefs in this way, something would seem amiss<sup>94</sup>.

The structure of the young child's (epistemic) environment is analogous to the structure of Samantha's (driving) environment. Tiny, gullible Tim is kept from liars and pranksters by his parents, and his parents themselves are reliable testifiers. Thus, Tim's global environment (where he usual forms testimony-based beliefs) is friendly to the formation of true beliefs on the basis of testimony. On a given occasion, a friend of the family, Jethro, is telling Tim neat stuff about dinosaurs. Jethro is a palaeontologist, and an honest one at that; he reliably testifies to Tim that there were dinosaurs with teeth bigger than the tallest basketball player! Thus, Tim's local environment is friendly to the formation of true beliefs based on testimony. Beside Jethro is mendacious Mordecai, who would tell

---

<sup>92</sup> Please refer back to footnote 91 of this work for Goldberg's conditions on knowledgeable testimony-based belief.

<sup>93</sup> Young children will be able to justifiably accept testimony on the further assumption of a more general thesis concerning acceptance of testimony known as "anti-reductionism".

(AR) A hearer is justified in accepting (has the epistemic right to accept; is epistemically entitled to accept) another's testimony so long as there are no undefeated good reasons *not* to accept the testimony (Goldberg 2007: 144) [original emphasis].

Tim tall tales about child-eating dinosaurs still roaming the earth today (!), if Tim's parents did not make sure that Mordecai never got to tell Tim stories. Now imagine testimonial twin earth, in which twin Tim's parents do not prevent Mordecai from fibbing to their son. Twin Tim is just as gullible as earth Tim, but the (epistemic) environment in which he forms most of his testimony-based beliefs is not friendly. That is, Tim and twin Tim differ with respect to their global environments. Hence, they differ with respect to their ability to form true beliefs on the basis of discriminating reliable from unreliable testimony.

#### §4.1: *Active Anti-Individualism?*

Having translated Goldberg's contentions about testimony-based belief in young children to Kallestrup & Pritchard's threefold distinction amongst environments, we see that the manner in which Goldberg's anti-individualism predicts differences in ability with differences in global environment squares with what we determined was the correct account of ability possession and individuation in Chapter 3. However, recall that we further concluded in Chapter 3 that the environmental dependence exhibited in each of the accounts considered was passive. An agent's immediate environs (her local environment) did not determine which ability she was deploying. Instead, *historical* facts about the situations in which she usually attempted to achieve a certain outcome (global facts) determined whether she had a particular ability. This, we saw, is the same sort of dependence externalists (or anti-individualists) about mental content appeal to in their theories. Active dependence, on the other hand, is what Clark & Chalmers' claimed for environmental factors like Otto's notebook. Cognitive externalist-style abilities require here-and-now dependence on factors with which a subject actually interacts. The contention is that this sort of active dependence engenders a constitutional relation: both bodily and extra-bodily components comprise an agent's ability. Goldberg's cases do not exhibit this latter sort of environmental dependency.

What do we make of Goldberg's claim that the account he gives of testimony-based belief in young children suggests an active anti-individualism, then? On one hand, he would seem to agree with our diagnosis using Kallestrup & Pritchard's apparatus.

---

<sup>94</sup> Goldberg (2007) explores the issue of the intuitiveness and desirability of this result (203 – 07).



[T]his [the adult’s] processing need not be counted as part of the *cognitive* process that eventuates in the child’s testimonial belief—any more than the broadly environmental regularities exploited by the perceptual process are part of the cognitive process that eventuates in perceptual belief (Goldberg 2007: 237; FN 34) [original emphasis].

However, Goldberg also indicates that he does want to argue for something more like here-and-now dependence on local facts (in Kallestrup & Pritchard’s sense).

In the case I have presented for AI-P, in contrast, adult caretakers play an *active* role in the process eventuating in the child’s acquisition of a testimonial belief, for example, by *monitoring testimony proffered in their joint presence* (*op cit* 238) [first emphasis original; second emphasis added].

AI-P<sup>95</sup> (which stands for “process anti-individualism”) states that, “[t]he process that eventuates in young children’s consumption of testimony (and hence the child’s testimonial belief) *extends to include* aspects of the child’s social environment . . .” (*op cit* 234) [emphasis added].

The picture evinced in the two previous quotations seems to be something like the following. Tim is in a room with Jethro, Mordecai and his parents. Jethro tells his story about dinosaur teeth, and Tim’s parents sit aside quietly. However, when Mordecai says child-eating dinosaurs still live today, Tim’s parents will intervene. They could say or do any number of things to affect this intervention. Perhaps, “Mordecai, what are you talking about? You don’t know anything about dinosaurs!” or, “Don’t listen to Mordecai, Timmy. He’s a big fat liar.”<sup>96</sup> Tim trusts his parents, and thus does not come to believe that there are child-eating dinosaurs roaming the land. Notice how we can apply Clark & Chalmers’ tests for cognitive extension to this case. Recall the criteria for an external resource being properly related to an agent to extend her cognition, which we saw in Chapter 2.

---

<sup>95</sup> Goldberg contrasts with his earlier anti-individualist theses concerning knowledge (AI-K), warrant (AI-W) (141) and justification (AI-J) (150).

<sup>96</sup> I take it that this is the sort of intervention Goldberg has in mind when, in a variant of his “Milk case”, a child’s mother uses “proactive monitoring” to prevent the child from consuming unreliable testimony (2007: 235).

- 1) That the resource be reliably available and typically invoked
- 2) That any information retrieved be more or less reliably endorsed. It should usually be subject to critical scrutiny . . . It should be deemed about as trustworthy as [the deliverance of a biological counterpart resource]
- 3) That the information in the resource should be easily accessible as and when required (Clark 2010: 46).

The parents Goldberg imagines are dutiful and scrupulous. Thus, they are usually around tiny Tim to monitor his testimony intake. We can easily imagine the other part of condition (1) being satisfied, as Tim learns to anticipate his parent's contributions in testimonial situations. Condition (2) is satisfied by Tim's stipulated credulity. Finally, condition (3) is satisfied by the dutiful attendance of Tim's parents as with condition (1).

The contributions of Tim's parents also jibe with parity principle reasoning, briefly mentioned in Chapter 2.

If, as we confront some task, part of the world functions as some process which, were it done in the head, we would have no hesitation in recognizing it as a cognitive process, then that part of the world (so we claim) is part of the cognitive process" (Clark & Chalmers 1998: 8).

In Clark & Chalmers' case concerning Otto and his notebook, the idea was that Otto's consultation of his notebook was analogous to a normally-functioning individual's consultation of their biological memory. That is, were Otto's consultation of his notebook done in his head, it would be just like consulting biological memory. Hence, we would consider it as (part of) a cognitive process. Thus, Otto's consultation of his notebook should count as (part of) a cognitive process when performed outside of the head. In Tim's case, the interventions of his parents are like the entries in Otto's notebook. When someone offers testimony to Tim, his parents engage in monitoring to determine whether that testimony is good. If not, they pipe in with some sort of warning to Tim, which Tim duly heeds. If this monitoring and warning were to go on in Tim's head, we would consider it a cognitive process. Indeed, the reasons for this treatment are stronger than the reasons for the same treatment of Otto's notebook: the monitoring done in Tim's parents' heads *is already cognitive*.

More tenuously, we can see the situation Goldberg sets up as one which confirms Rob Wilson's argument for cognitive externalism. Goldberg's discussion of children's ability to monitor testimonial credibility indicates that their credulity is pervasive, but not complete (2007: 201- 02). Thus, we might (tendentiously) claim that the core realiser of the ability to monitor credibility required for the attribution of knowledgeable or justified testimony-based belief to a child is partly within the child and partly within the adults monitoring the child. Were this the case, the ability to monitor credibility would be radically widely realized. Thus, we would have a case which confirms cognitive externalism.

Note that the ability which I have been talking about is the discriminatory ability relating to distinguishing reliable from unreliable testimony. When Goldberg contends that his view is not committed to cognitive externalism, he focuses on *belief-forming processes* (2007: 237, and see FN 34; 2010: 128). As he makes clear, belief-forming processes as the epistemologist is interested in them are in some cases extended, both spatially and temporally.

Consider a case of testimony. Fred tells Jethro that it is raining outside. For Jethro to count as having a knowledgeable testimony-based belief that it is raining outside the following must obtain. First, Fred must be a reliable testifier. That is, he must both be honest and be able to properly discriminate facts about the weather. Second, Jethro must understand Fred's testimony on the basis of a reliable ability for linguistic understanding. Finally, Jethro must accept Fred's testimony on the basis of a reliable ability to discriminate reliable from unreliable testimony. According to Goldberg, the belief-forming process which issues in Jethro's belief that it is raining is comprised by the factors involved in all three of the above conditions, and this is because these are all relevant to the "reliability in formation and/or sustainment of belief" (2010: 128)<sup>97</sup>. Thus, the process issuing in Jethro's belief that it is raining is temporally extended because it depends on the reliability of Fred's discernment of the weather conditions, which occurred prior to Jethro's belief formation. Further, the belief-forming process is spatially extended because it depends on the workings of Fred's cognitive system, which is obviously spatially distinct from Jethro's cognitive system.

---

<sup>97</sup> Goldberg compares this to what the epistemic externalist says about memory-based belief (*ibid*).

However, Goldberg claims that this should not be particularly interesting to a cognitive psychologist. Instead, the cognitive psychologist is interested in discovering the “basic belief forming processes” (*op cit* 129). Further, what the psychologist is interested in does not exhaust what the epistemologist is interested in, as noted above. The epistemologist does care about these basic belief-forming processes and their reliability—but she also cares about additional factors, which are relevant to the assessment of the epistemic success of beliefs. We can interpret Goldberg as follows: where cognitive psychology and anti-individualist epistemology overlap, they can agree on cognitive internalism. That is, they can agree on the cognitive internalist construal of basic belief forming processes (e.g., visual perception, memory, etc.). Goldberg’s claim is that the anti-individualist “extends” epistemic processes, not cognitive processing.

What about my claim that Goldberg’s account implies the extension of a discriminatory capacity? Does this change the state-of-play? Recall the notion of cognitive abilities we have been working with. Cognitive abilities are the capacities to discriminate, gather, process and remember various pieces of information. This is a plausible formulation of what both cognitive scientists and epistemologists mean by cognitive abilities. Goldberg’s conditions on the acquisition of knowledgeable (and justified) testimony-based belief include a discrimination condition on the agent. Specifically, she must discriminate the state-of-affairs which engenders reliable testimony from the state-of-affairs which does not. In the case of testimony consumption in young children, Goldberg argues that children cannot satisfy this discrimination condition on their own. Subsequently, I have argued that Goldberg’s invocation of the parent’s proactive monitoring of their children’s testimony consumption meets criteria for counting as an instance of cognitive extension. Hence, whatever we say about belief-forming processes, Goldberg’s argument for epistemic anti-individualism does imply an extended discriminatory process, and thus plausibly implies an instance of extended cognition<sup>98</sup>.

---

<sup>98</sup> Goldberg considers these sorts of process to be *subcognitive*. Hence, he would likely object to my argument about discriminatory processes. “[Testimonial] belief-fixation involves (subcognitive) processes whose job it is to enable the recipient to discern cases in which trust it to be extended, from those cases in which it is not” (2007: 148). Previously, Goldberg (1999) defines subcognitive processes as those which issue in beliefs, “[producing] these beliefs in a manner that is insensitive to

There appears to be a tension, then, in Goldberg's claims concerning the consumption of testimony by young children. On the one hand, the monitoring performed by parents is described as a feature of the environment upon which a child's acquisition of knowledgeable or justified testimony-based belief (merely passively) depends. This is seen in our analysis of Goldberg's cases via Kallestrup & Pritchard's distinction among types of environment, and in Goldberg's repudiation of cognitive externalism. On the other hand, Goldberg makes much of the difference between testimony-based belief in young children and in mature consumers of testimony. The role of the cognitive lives of others in the case of mature consumers of testimony is passive, whereas the role of adults in the case of immature consumers of testimony is active. Moreover, it is active in just the sense of Clark & Chalmers. Thus, I conclude that the scenarios Goldberg uses to argue for AI-P are ones in which the relevant external factors (here, elements of other agents' cognitive lives) would pass the tests set by Clark & Chalmers for cognitive extension.

#### *§4.2: Objection to (8\*)—Revisited*

Earlier in this chapter, we left unresolved an objection to premise (8\*) of my revised argument for a commitment to cognitive externalism on the part of the epistemic externalist. The objection was that the move from (3\*) and (7\*) to (8\*) was illicit. I argued that it was easier to respond to this objection than its counterpart to the argument in Chapter 4, provided that the dependence on other's cognitive lives appealed to by the anti-individualist

---

the person's other beliefs" (177). The belief-fixing processes in the case of testimony are supposed to be subcognitive in the same ways as those processes involved in vision: i.e., they "mediate between perceptual [or testimonial] experience and perceptual [or testimonial] belief" (2007: 148). However, this distinction between cognitive processes and subcognitive processes is at odds with how *both* cognitive internalist and externalists talk. Adams & Aizawa (2008) claim that cognition occurs wherever there is a distinct sort of processing over non-derived content. Specifically, they mention the visual areas of the brain (17). Further, Rupert (2009) gives the appellation "cognitive" to any factors and processes which scores sufficiently high according to his conditional probability apparatus with respect to the production of a cognitive outcome. On the assumption that visual sub-processes will score highly with respect to the production of visual judgments and beliefs, Rupert will count these processes as cognitive. Finally, cognitive externalists often treat any enabling factor for a cognitive process as part of that process (e.g., Wilson 2004; Hurley 2010). Given the disconnect between Goldberg's cognitive-subcognitive distinction and the way both cognitive internalists and externalists talk, I will not worry about objections based on said distinction.

was of the right sort. Without isolating what sort of dependence was at issue, I claimed that we ran the risk of committing an equivocation across sense of ‘depend’, as was the case in the briefly-considered argument of Chapter 3.

In our discussion of Goldberg, I have argued (*pace* Goldberg) that the dependence on the cognitive lives of their parents exhibited by children in cases of testimony-based belief is the sort of active dependence argued for by the cognitive externalist. As such, I take the objection to (8\*) to be resolved. My final argument for a commitment to cognitive externalism on the part of the epistemic externalist goes through.

#### *§5: Summary and Conclusion*

Here is the final argument, presented as a whole.

- (1\*) Epistemic anti-individualism entails the rejection of epistemic individualism [*Definition*]
- (2\*) Mentalism is a form of epistemic individualism [*Definition*]
- (3\*) Therefore, epistemic anti-individualism entails the rejection of mentalism [*From 1\* & 2\**]
- (4) An account of justification which appeals to cognitive abilities, where those abilities are exhaustively constituted by factors internal to the individual of whom said abilities are predicated, is consistent with mentalism. [*Premise*]
- (5) Cognitive internalism is the claim that cognitive abilities are exhausted by factors internal to the individual of whom said abilities are predicated [*Definition*]
- (6) Therefore, an account of justification which appeals to cognitive abilities and is consistent with mentalism implies cognitive internalism [*From 4 & 5*]
- (7\*) An epistemic anti-individualist account of justification must appeal to cognitive abilities [*Premise*]
- (8\*) If an account of justification which appeals to cognitive abilities and is consistent with mentalism implies cognitive internalism, then epistemic anti-individualism must reject cognitive internalism [*From 3\* and 7\**]

(9\*) Therefore, epistemic anti-individualism must reject cognitive internalism

[From 6 & 8\*]

(10) The rejection of cognitive internalism entails the acceptance of cognitive externalism [Definition]

(11\*) Therefore, epistemic anti-individualism implies cognitive externalism

[From 9\* & 10]

(12\*: C) Therefore, an epistemic anti-individualist account of justification

must appeal to cognitive abilities as envisioned by cognitive externalism [From 7\* & 11\*]

This iteration of the argument divested itself of reliance upon mentalism counting as a form of epistemic internalism, thus removing a weakness. That is, (2\*) cannot be objected to where its counterpart (2) could be. Objections to (7\*) are set aside on the power of arguments made by other authors. Finally, I have given responses to possible objections against (4) and (8\*). Thus, I take the argument to go through.

As far as our discussion has gone, the argument goes through on the back of Goldberg's arguments concerning testimony-based belief in young children. This is not the only set of cases we might adduce. After all, Goldberg claims that the sort of scenarios we have imagined for young children might arise for adults as well when they are particularly gullible. If these adults have friends or partners who watch out for their epistemic well-being, AI-P is recapitulated in such cases. Further, I do not think mature consumers of testimony need be generally gullible—as young children are supposed to be—for AI-P to obtain. Even relatively well-informed and intelligent adults may fail to be reliable discriminators of reliable from unreliable testimony with respect to certain bodies of propositions. Given a sufficiently recondite piece of testimony, an adult may be vulnerable to confidence scams. Within the public discourse concerning various specialist topics, adult consumers of testimony may display a certain amount of credulity by way of deference to apparent authority. However, in such cases (especially in public political and scientific discourse) bent truths may be passed-off as straight. Even worse, experts are notorious for overestimating the domain of their authority, or deliberately trading on their status as an

expert when making dubious claims. I submit that in these cases, the various agencies (independent or otherwise) which restrict what may be said in the public domain without substantial proof, or which counteract spurious testimony when it appears, play the same role as parents do. That is, these agencies (ideally) ensure that only reliable testimony makes it into the public realm; or they counteract unreliable testimony where they find it.

More generally, the cases of epistemic anti-individualism which confirm cognitive externalism are those in which the cognitive lives of others in part constitute the abilities to which the anti-individualist appeals in her account of knowledge and testimony. Interestingly, this limits the implications of anti-individualism to *socially-extended* cognition. There is no way here for things like Otto's notebook, or the on-screen manipulation of zoids in *Tetris*, to extend cognition. Nonetheless, there are enough young children, gullible adults and non-experts watched over by dutiful guardians to provide a raft of cases to support my argument for a commitment to cognitive externalism on the part of the epistemic externalist.

Note that this is still a commitment on the part of the epistemic externalist, despite our shift of focus to epistemic anti-individualism. After all, Goldberg defines anti-individualism as a strong form of epistemic externalism. On two further assumptions, the commitment ensues for all viable forms of epistemic externalism. First, we must assume that the externalist wishes to give an externalist account of testimony-based knowledge. This assumption is eminently plausible. Second, we must assume that Goldberg's arguments regarding testimony go through. Of course, we already needed to assume this for the argument for the commitment to cognitive externalism to go through. No new burdens are added in this respect.

The targets of my argument are subject to the same limitations we noted in response to the objection to (8) in Chapter 4. That is, the argument applies only when the externalist account at issue appeals only to an ability condition (and not a further, anti-luck condition) in its explication of an epistemic status. In the case of justification, we have assumed that this applies to any externalist account. With regard to knowledge, it will apply only to so-called robust virtue approaches, which claim that only an ability condition need be added to true belief to secure knowledge. Goldberg's conditions on testimonial knowledge each seem to be ability conditions (see footnote 90 of this work). The testifier must be reliable *vis à vis*



the proposition to which she testifies, and the hearer must form a belief based on the testimony offered in virtue of both a reliable ability to understand said testimony and an ability to discriminate reliable from unreliable testimony. Perhaps the condition on the testifier need not be an ability condition. If not, then Goldberg's account of testimonial knowledge escapes the commitment I pose. I think a plausible spelling-out of the testifier's reliability will involve reliability condition, but we can here bracket this concern. In any case, justified acceptance of testimony depends on the two reliability conditions imposed on the consumer of testimony. As each of these is an ability condition, my argument gets its hooks in fully. At least externalist accounts of justification incur a commitment to cognitive externalism, and so do certain accounts of knowledge. This is a substantial result.

In terms of the confirmation conditions I delineated in the first two chapters, this is the result. Confirmation of epistemic anti-individualism (which is a strong version of epistemic externalism) does constitute confirmation of cognitive externalism. Nor have we found only one instance. Given Goldberg's arguments, a whole class of cases—*viz.*, testimony-based belief in young children (and gullible adults)—have been provided. Thus, it does not matter if the confirmation conditions given for cognitive externalism were too weak because they required only *one actual case*. The final argument provided in this chapter yields a swath of actual cases. The task of this work has come to a successful end.

### **Afterword: Coping with the Commitment**

In this work, I have examined the connections between epistemic externalism and cognitive externalism. Specifically, I explored various arguments to the effect that epistemic externalism is committed to cognitive externalism. Three arguments for this commitment were constructed; only one survived scrutiny. This latter argument invoked a particular sort of epistemic externalism called epistemic anti-individualism—*viz.*, the thesis that an individual's epistemic standing may depend on facts about the cognitive goings-on of other individuals. The sort of cases which were directly used to bear out my last argument were those concerning the consumption of testimony by young children. Given the incomplete ability of young children to discriminate reliable testimony from unreliable testimony, where these children were to be ascribed knowledgeable or justified testimony-based belief, the contributions of their parents had to be taken into account. These contributions were said by Goldberg to be active rather than passive, in the technical sense of this distinction delineated by reference to the work of Clark & Chalmers. Moreover, the parents' contributions were seen to pass the tests set by Clark & Chalmers (and perhaps those set by Rob Wilson as well) for counting as an instance of cognitive externalism. Nor was this result necessarily limited to the case of young children. Adults' consumption of testimony is actively monitored in some cases as well, where the adults in question are particularly credulous by disposition, or

the content of the relevant testimony sufficiently abstruse that most adult consumers are vulnerable to confidence scams.

Should the epistemic internalist be worried about the commitment I have pinned on her? Initially, it might appear that she should. As Goldberg (2010) puts it, there may be a “bad company” objection laying in wait. There are two general sorts of reason that a commitment to cognitive externalism might be an albatross around the neck of the epistemic externalist. I will briefly discuss these worries and give a cursory diagnosis of their severity. The proceeding comments will be neither comprehensive nor detailed. It is not my aim to demonstrate that the commitment with which I have burdened the epistemic externalist brings no problems with it—just that it brings no obviously fatal ones.

First, there are worries about the plausibility of cognitive externalism. Recall that cognitive externalism is not a thesis about the mere possibility of cases in which an individual’s cognition is in part constituted by factors outside of her. Cognitive externalism asserts the *actuality* of such cases. Thus, a commitment to cognitive externalism involves being committed to the world being a certain way. Adams & Aizawa and Rupert all dispute that any version of cognitive externalism is true of our world. That is, though they may grant the odd case of extended cognition, these critics generally think that confirming instances of cognitive externalism are not sufficiently widespread to be theoretically interesting to cognitive science. Of course, this might not worry epistemologists too much. Moreover, there are arguments against the cognitive internalist’s claims that their theories better fit empirical data. (For example, Sprevak 2010.) The point is just that any commitment to cognitive externalism is not *obviously* problematic from an empirical standpoint. While it may in general be desirable to minimise the empirical commitments one’s theory has, so long as the extant commitments do not obviously conflict with empirical fact, there should be little call for concern.

More worrying might be the philosophical problems with cognitive externalism. While critics acknowledge the possibility of extended cognition in a general sense, they often find fault with the reasoning in all extant arguments for it. Adams & Aizawa (2008) argue that all extant arguments for cognitive externalism commit some version of what they call the “coupling-constitution error”. Briefly, Adams & Aizawa claim that it is erroneous to move

from facts about the coupled relationship between two factors (regardless of how intimate) to claims that these factors jointly constitute some single system or process (*op cit* Chp. 6). Of course, proponents of cognitive externalism have their counterarguments to claims like that of Adams & Aizawa. For example, Susan Hurley (2010) speaks of the error made when levelling a coupling-constitution objection against a cognitive externalist theory. “[It is] the error of objecting that externalist give a constitutive role to external factors that are ‘merely causal’ while assuming without independent argument or criteria that the causal/constitutive distinction coincides with some internal/external boundary” (*op cit* 5). Again, it is generally preferable not to be committed to contentious philosophical theses beyond those for which one directly argues—but so long as the accounts one is committed to are not *obviously* false or incoherent, little sleep should be lost.

The problems that should most concern the epistemic externalist about a commitment to cognitive externalism are those which it might create *vis à vis* motivations for her epistemic externalism. That is, if taking on board cognitive externalism brings epistemological difficulties, the epistemic externalist is in trouble. A particularly acute problem arises in virtue of the intuition that ability is involved in knowledge and justification.

Recall that we adverted to ability to solve the direction-of-fit problem. That is, intuition militates against attributing positive epistemic status to beliefs which are true through no effort of the relevant agent. In the daemon case we explored in Chapter 2, Fred’s beliefs were made true by the intervention of a benevolent demon. When Fred formed a belief that the toast was burning, this daemon ensured that the toast was, in fact, burning. In such a case, it seems that Fred deserves no credit for his true belief. However, if Fred’s belief were true because it was formed in the right way, then he would deserve credit. If Fred formed a belief that the toast was burning because he either smelled carbonising bread or saw smoke billowing from the toaster, it would be intuitively right to assign positive epistemic status to that belief. That is, he would deserve credit for the success of these beliefs because it was his abilities (here, smelling and seeing) which were responsible for the success.

But what if Fred’s abilities are partly constituted by the operation of certain sorts of instruments, or the cognitive efforts of others? Consider Clark & Chalmers suggestion,

“[e]pistemic action [like the on-screen rotation of zoids in *Tetris*], we suggest, demands spread of epistemic credit” (1998: 8). That is, some of the credit for solving a given task should be spread outside of the canonically-delineated agent, and into the portions of the world which that agent manipulates. But if the credit given to the agent herself is reduced, is it still enough to assign positive epistemic status to her beliefs?

Duncan Pritchard (2010) addresses this issue. He identifies two different versions of the criterion which requires that an agent be creditable for the success of her beliefs in order for positive epistemic status to be properly assigned to those beliefs.

(COGA<sub>WEAK</sub>) If *S* knows that *p*, then *S*'s true belief that *p* of a reliable belief-forming process which is appropriately integrated within *S*'s cognitive character such that her cognitive success is to a significant degree creditable to her cognitive agency.

(COGA<sub>STRONG</sub>) *S* knows that *p* iff *S*'s true belief is the product of a reliable belief-forming process which is appropriately integrated within *S*'s cognitive character such that her cognitive success is primarily creditable to her cognitive agency (136 – 37)

The relevant difference between these two principles is that the former demands significant credit be attributable to the agent, while the latter demands primary credit be attributable to the agent<sup>99</sup>. According to Pritchard, it is only if we adopt the strong credibility requirement that attributing positive epistemic status to beliefs reached by cognitive externalist style abilities is problematic.

---

<sup>99</sup> It is also significant that the weak condition is in conditional form, while the strong one is in biconditional form. The biconditional formulation is characteristic of robust virtue approaches, which we have discussed elsewhere in this work. However, Pritchard notes that even weakened to a conditional formulation, the strong condition seems to conflict with cognitive externalist style abilities (140 – 141). It is this weakened formulation (termed “COGA<sub>STRONG\*</sub>”) that concerns us.

Pritchard asks us to consider a case of testimony-based knowledge. He claims that, on the strong credibility requirement, we will generally not be able to attribute knowledge to agents who form beliefs on the basis of testimony.

Crucially, however, it does not seem plausible to suppose that Jenny's [the hearer] cognitive success is primarily creditable to her cognitive agency. After all, given that the bulk of the cognitive work here was done by Jenny's informant, then it would seem that if any person is primarily creditable for Jenny's cognitive success then it is her [Jenny's informant] rather than Jenny (*op cit* 141).

Of course, we are familiar with this from Goldberg's discussion of testimony-based belief in young children. Due to a lack of ability to discriminate amongst reliable and unreliable testimony, children are not able on their own effort to fulfil all the criteria for the attribution of knowledgeable or justified testimony-based belief. However, despite this it was still deemed intuitive that knowledgeable and justified testimony-based belief should be attributed to young children. These facts led Goldberg to argue that children could fulfil the relevant criteria on the strength of their parent's contributions in addition to their own.

Pritchard uses considerations of testimony-based belief to argue for adopting  $COGA_{WEAK}$ . The idea is this. If  $COGA_{STRONG}$  were true, we would not have much testimony-based knowledge. But we do have much testimony-based knowledge. Therefore,  $COGA_{STRONG}$  is not true. Note that this argument does not anywhere mention cognitive externalism. That is, Pritchard argues we should accept  $COGA_{WEAK}$  *independent* of any considerations about the nature of cognitive abilities. This being the case, the epistemic externalist can take on cognitive externalism without causing any strife within her epistemic motivations and commitments. The spread of epistemic credit to extra-individual factors does not threaten the epistemic status of individuals' beliefs.

In this afterword, I have briefly mentioned what might immediately strike one as problems ensuing for the epistemic externalist from her commitment to cognitive externalism. First, there were problems concerning cognitive externalism itself. Second, I addressed possible conflicts between the epistemic motivations and commitments of

epistemic externalism and a commitment to cognitive externalism. A brief overview of these potential worries showed no obvious catastrophes on the horizon. Indeed, in the case of the second set of worries, it was seen that the epistemic externalist's prior commitments *vis à vis* testimony-based knowledge provided friendly ground for cognitive externalism. It appears that the epistemic externalist will be able to cope well with her commitment to cognitive externalism.

## References

- Adams, F. & Aizawa, K. 2001. "The Bounds of Cognition". *Philosophical Psychology* 14: 43 – 64.
- 2008. *The Bounds of Cognition*. Blackwell Publishing.
- Alston, W. 1996. *Epistemic Justification: Essays in the Theory of Knowledge*. Cornell University Press.
- Anderson, J.R. 1995. *Learning and Memory: An Integrated Approach*. Wiley.
- Armstrong, D. 1973. *Belief, Truth and Knowledge*. Cambridge University Press.
- Austin, J.L. 1956. "Ifs and Cans". *Proceeding of the British Academy* 42: 107 – 32.
- Bartlett, G. 2008. "Whither Internalism? How Should Internalists Respond to the Extended Mind Hypothesis?". *Metaphilosophy* 39 (2): 163 – 84.
- BonJour, L. 1980. "Externalist Theories of Empirical Knowledge" in Sosa, E., Kim, J., Fantl, J. & McGrath, M. (eds.) *Epistemology: An Anthology* (Second Edition): 363 – 78.
- 1992. "Internalism/Externalism" in Dancy, J. & Sosa, E. (eds.) *A Companion to Epistemology* 132- 36. Blackwell.
- Brown, J. 2004. *Anti-Individualism and Knowledge*. MIT Press.
- 2007. "Externalism is Mind and Epistemology" in Goldberg, S. (ed.) *Internalism and Externalism in Semantics and Epistemology*. Oxford University Press.
- Burge, T. 1979. "Individualism and the Mental". *Midwest Studies in Philosophy* 4 (1): 73 – 122.
- 1986. "Individualism and Psychology". *Philosophical Review* 95: 3 – 45.
- 2003. "Perceptual Entitlement". *Philosophy and Phenomenological Research* 67 (3): 503 – 548.
- 2010. *Origins of Objectivity*. Oxford University Press.
- Clark, A. 2008a. "Pressing the Flesh". *Philosophy and Phenomenological Research* 76 (1): 37 – 59.
- 2008b. *Supersizing the Mind: Embodiment, Action and Cognitive Extension*. Oxford University Press.
- 2010. "Coupling, Constitution and the Cognitive Kind". In Menary, R. (ed.) *The Extended Mind*. MIT Press.
- Chomsky, N. 1965. *Aspects of the Theory of Syntax*. MIT Press.



- Clark, A. & Chalmers, D. 1998. "The Extended Mind". *Analysis* 58(1): 7-19.
- Conee, E. 2007. "Externally Enhanced Internalism" in Goldberg, S. (ed.) *Internalism and Externalism in Semantics and Epistemology*. Oxford University Press.
- Conee, E & Feldman, R. 1998. "The Generality Problem for Reliabilism". *Philosophical Studies* 89: 1 – 29.
- 2004. *Evidentialism: Essays in Epistemology*. Oxford University Press.
- Dennett, D. 1981. "Where Am I?" in Dennett, D. (ed.) *Brainstorms*. Harvester Press.
- 1990. "The Myth of Original Intentionality" in Mohyeldin Said, K. A., Newton-Smith, W. H., Viale, R. & Wilkes, K.V. (eds.) *Modeling the Mind* 43 – 62. Oxford University Press.
- 1996. *Kinds of Minds*. Basic Books.
- Fodor, J. 1968. *Psychological Explanation*. Random House.
- 1987. *Psychosemantics*. MIT Press.
- 1988. *The Modularity of Mind: An Essay on Faculty Psychology*. MIT Press.
- Feldman, R. 2004. "In Search of Internalism and Externalism" in Schantz, R. (ed.) *The Externalist Challenge* 143 – 56. Walter de Gruyter.
- 2005. "Justification is Internal" in Steup, M. & Sosa, E. (eds.) *Contemporary Debates in Epistemology*. Blackwell.
- Franz, V. H., Gegenfurtner, K. R., Bühlhoff, H. H. & Fahle, M. 2000. "Grasping Visual Illusions: No Evidence for a Dissociation Between Perception and Action". *Psychological Science* 11 (1): 20 – 25.
- Fumerton, R. 2007. "What and About What is Internalism?" in Goldberg, S. (ed.) *Internalism and Externalism in Semantics and Epistemology*. Oxford University Press.
- Gertler, B. 2007a. "Overextending the Mind?" in Shapiro, L. (ed.) *Arguing About the Mind*. Routledge.
- 2007b. "Content Externalism and the Epistemic Conception of the Self" in Sosa, E. & Villanueva, E. (eds.) *The Metaphysics of Epistemology*. Philosophical Issues, Vol. 17. Blackwell.
- Gibbons, J. 2006. "Access Externalism". *Mind* 115 (457): 19 – 39.

- Goldberg, S. 1999. "The Psychology and Epistemology of Self-Knowledge". *Synthese* 188 (2): 165 – 99.
- 2007. *Anti-Individualism: Mind and Language, Knowledge and Justification*. Cambridge University Press.
- 2010. *Relying on Others*. Oxford University Press.
- Goldman, A. 1986. *Epistemology and Cognitive Science*.
- 1999. "Internalism Exposed". *Journal of Philosophy* 96: 271 – 93.
- Goodale, M. A. & Milner, A. D. 1982. "Separate Visual Pathways for Perceptions and Action". *Trends in Neuroscience* 15 (1): 20 – 25.
- Greco, J. 1999. "Agent Reliabilism". *Philosophical Perspectives* 13 (s13): 273 – 96.
- 2005. "Justification is Not Internal". In Steup, M. & Sosa, E. (eds.) *Contemporary Debates in Epistemology*. Blackwell.
- 2006. "Virtue, Luck and the Pyrrhonian Problematic". *Philosophical Studies* 130 (1): 9 – 34.
- 2007. "The Nature of Ability and the Purpose of Knowledge". in Sosa, E. & Villanueva, E. (eds.) *The Metaphysics of Epistemology*. Philosophical Issues, Vol. 17. Blackwell.
- 2009. "Knowledge and Success from Ability?". *Philosophical Studies* 142 (1).
- Hagueland, J. 1998. "Mind Embodied and Embedded" in Hagueland, J. (ed.) *Having Thought* 207 – 37. Harvard University Press.
- Honoré, A. M. 1964. "Can and Can't". *Mind* Vol. 73, No. 292: 463 – 79.
- Hurley, S. 1998a. "Vehicles, contents, conceptual structure, and externalism". *Analysis* 58(1): 1-6.
- 1998b. *Consciousness in Action*. Harvard University Press.
- 2010. "Varieties of Externalism" in Menary, R. (ed.) *The Extended Mind*. Ashgate.
- Kallestrup, J. & Pritchard, D. Manuscript. "Virtue Epistemology and Epistemic Twin Earth".

- Kaplan, D. 1989. "Demonstratives" in Almog, J., Perry, J. & Wettstein, H. (eds.) *Themes from Kaplan* 481 – 504. Oxford University Press.
- Kelp, C. Forthcoming. "Extended Cognition and Robust Virtue Epistemology". *Erkenntnis*.
- Kim, J. 1993. *Supervenience and Mind*. Cambridge University Press.
- Kirsch & Maligo. 1994. "On Distinguishing Epistemic from Pragmatic Action". *Cognitive Science* 18: 513 – 49.
- Levine, T. 1999. "Philosophy as Massage: Seeking Cognitive Relief for Conscious Tension". *Philosophical Topics* 26: 159 – 78.
- Loar, B. 1988. "Social Content and Psychological Content" in Grimm, R. and Merrill, D. (eds.) *Contents of Thought*. Arizona University Press.
- Marr, D. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Freeman.
- McDowell, J. 1995. "Knowledge and the Internal". *Philosophy and Phenomenological Research* 55 (4): 877 – 93.
- Menary, R. 2007. *Cognitive Integration: Mind and Cognition Unbounded*. Palgrave Macmillan.
- Millar, A. 2009. "What is it that Cognitive Abilities Are Abilities to Do?". *Acta Analytica* 24: 223 – 36.
- Mishkin, M &. Ungerleider, L. G. 1982. "Contribution of Striate Inputs to Visuospatial Functions of the Parieto-preoccipital Cortex in Monkeys". *Behavioural Brain Research* 6 (1): 57 – 77.
- Noë, A. 2004. *Action and Perception*. MIT Press.
- 2009. *Out of Our Heads: Why You Are Not Your Brain, and Other Lessons from the Biology of Consciousness*.
- Plantinga, A. 1993. *Warrant: The Current Debate*. Oxford University Press.
- Prichard, H.A. 1950. *Knowledge and Perception*. Clarendon Press.
- Pritchard, D. 2007. *Epistemic Luck*.
- 2010. "Cognitive Ability and the Extended Mind Thesis". *Synthese* 175: 133 -51.

- 2011. “Evidentialism, Internalism, Disjunctivism” in Dougherty, T. (ed.), 362 – 92. *Evidentialism and its Discontents*. Oxford University Press.
- Forthcoming. “Anti-Luck Virtue Epistemology”. *The Journal of Philosophy*.
- Pritchard, D., Millar, A. & Haddock, A. 2010. *The Nature and Value of Knowledge*. Oxford University Press.
- Pryor, J. 2001. “Highlights of Recent Epistemology”. *British Journal for the Philosophy of Science* 52: 95-124.
- Putnam, H. 1975a. “Minds and Machines” in *Mind, Language and Reality* 362 – 85. Cambridge University Press.
- 1975b. “The Meaning of Meaning” in *Mind, Language and Reality* 215- 71. Cambridge University Press.
- Quine, W. V. O. 1969. “Epistemology Naturalized” in *Ontological Relativity and Other Essays*. Columbia University Press.
- Rowlands, M. 2009. “Enactivism and the Extended Mind”. *Topoi* 28 (1).
- Rupert, R. 2009. *Cognitive Systems and the Extended Mind*. Oxford University Press.
- 2010. “Systems, Functions and Intrinsic Natures: On Adams and Aizawa’s *The Bounds of Cognition*”. *Philosophical Psychology* 23 (1): 113 – 23.
- Searle, J. 1984. *Minds, Brains and Science*. Harvard University Press.
- Sosa, E. 2007. *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume 1*. Oxford University Press.
- Sprevak, M. 2007. “Chinese Rooms and Program Portability”. *British Journal for the Philosophy of Science* 58: 755 – 76.
- 2010. “Inference to the Hypothesis of Extended Cognition”. *Studies in History and Philosophy of Science*, 41, 353 – 62.
- Steup, M. 2006. “The Analysis of Knowledge” in *The Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/entries/knowledge-analysis/>
- Stitch, S. & Warfield, T. 1994. *Mental Representation: A Reader*. Blackwell.
- Thompson, E. 2007. *Mind in Life: Biology, Phenomenology and the Sciences of Mind*.
- Tye, M. 2009. *Consciousness Revisited: Materialism Without Phenomenal Concepts*. MIT Press.

- Vaesan, K. Forthcoming. "Knowledge Without Credit, Exhibit 4: Extended Cognition".  
*Synthese*.
- Vahid, H. 2005. *Epistemic Justification and the Skeptical Challenge*. Palgrave Macmillan.
- 2011. "Externalism/Internalism" in Bernecker, S. & Pritchard, D. (eds.)  
*The Routledge Companion to Epistemology* 144 – 55. Routledge.
- Van Gelder, T. 1995. "What Might Cognition Be, If Not Computation?" *Journal of  
Philosophy* 81: 345 – 81.
- Wedgwood, R. 2002. "Internalism Explained". *Philosophy and Phenomenological Research* 66: 349  
– 69.
- 2010. "The Moral Evil Demons" in Feldman, R. & Warfield, T. (eds.)  
*Disagreement* 216 – 46. Oxford University Press.
- Williamson, T. 2000. *Knowledge and Its Limits*. Oxford University Press.
- Wilson, R. 2004. *Boundaries of the Mind: The Individual in the Fragile Sciences*. Cambridge  
University Press.
- Zagzebski, L. 1996. *Virtues of the Mind: An Inquiry Into the Nature of Virtue and the Ethical  
Foundations of Knowledge*. Cambridge University Press.