



# 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.

# Is a subpersonal epistemology possible?

Re-evaluating cognitive integration for extended cognition

Hadeel Naeem

Doctoral Thesis in Philosophy  
The University of Edinburgh  
November 2020

*To Irteza*

# Contents

<b>Abstract</b>	<b>iv</b>
<b>Lay Summary</b>	<b>vi</b>
<b>Acknowledgements</b>	<b>viii</b>
<b>Introduction</b>	<b>1</b>
Overview . . . . .	1
Motivation for the subject . . . . .	6
Chapter descriptions . . . . .	8
<b>Chapter 1 The two integrations</b>	<b>12</b>
Metaphysical cognitive integration . . . . .	13
The glue and trust criteria . . . . .	15
What is a univocal view? . . . . .	17
The ability intuition and extended cognition . . . . .	20
Epistemic integration . . . . .	27
Example of a univocal view . . . . .	32
Conclusion . . . . .	38
<b>Chapter 2 The univocal views and the epistemic hygiene dilemma</b>	<b>40</b>
Problems with the glue and trust conditions . . . . .	41
Continuous reciprocal causation . . . . .	44
More univocal views . . . . .	48
The hygiene problem . . . . .	51
A subpersonal epistemology . . . . .	56
No need for a subpersonal epistemology . . . . .	60
Conclusion . . . . .	63
<b>Chapter 3 Immediate epistemic integration</b>	<b>66</b>
Clark's immediate integration . . . . .	67
Deep functional integration . . . . .	71

Epistemic integration at birth and later . . . . .	75
The differences in integrations . . . . .	80
Where do we stand on the univocal view? . . . . .	83
Conclusion . . . . .	84
<b>Chapter 4 Subpersonal epistemic integration</b>	<b>86</b>
Predictive processing . . . . .	87
Weak and subpersonal epistemic integration . . . . .	92
Subpersonal level: doxastic and subdoxastic states . . . . .	97
Problematic ubiquitousness . . . . .	99
Are these two integrations different? . . . . .	102
Conclusion . . . . .	108
<b>Chapter 5 Understanding epistemic responsibility via defeaters</b>	<b>111</b>
What are defeaters? . . . . .	113
Types of defeaters . . . . .	115
The no-defeater condition . . . . .	120
Subpersonal defeasibility via reflective defeaters . . . . .	125
Subpersonal defeasibility via subpersonal defeaters . . . . .	130
Conclusion . . . . .	134
<b>Chapter 6 Knowledge bloat and demarcating the epistemic agent</b>	<b>135</b>
Alexa . . . . .	136
Does Alexa extend my cognition? . . . . .	138
Does Alexa extend my mind? . . . . .	140
Forms of believing . . . . .	143
Attributing beliefs . . . . .	147
Double and triple extension . . . . .	151
The knowledge bloat problem . . . . .	155
Demarcating the epistemic agent . . . . .	161
Conclusion . . . . .	166
<b>Conclusion</b>	<b>167</b>
Areas for improvement . . . . .	171



## Abstract

Virtue reliabilism provides an account of epistemic integration that explains how a reliable-belief forming process can become a knowledge-conducive ability of one's cognitive character. The univocal view suggests that this epistemic integration can also explain how an external process can extend one's cognition into the environment. Andy Clark finds a problem with the univocal view. He claims that cognitive extension is a wholly subpersonal affair, whereas the epistemic integration that virtue reliabilism puts forward requires personal-level agential involvement. To adjust the univocal view, Clark recommends a subpersonal epistemic integration account that also paves way for a wholly subpersonal epistemology. Accordingly, an epistemic agent can take responsibility for her reliable belief-forming process by way of entirely subpersonal mechanisms.

The aim of this thesis is to argue against a subpersonal epistemology and the need for it. First, I bring into question the conditions that motivate extended cognition: the so-called 'glue and trust' requirements and the functional parity principle. Neither of these conditions demands that extension should be understood in entirely subpersonal terms. On the contrary, the glue and trust conditions suggest that agents should personally and actively engage with their external vehicles to extend cognition. Further, I consider an important disparity between the two kinds of integration. Integration that prompts extension can happen immediately, but the integration that makes a reliable belief-forming process knowledge-conducive is almost always a slow process. In light of this and other similar inconsistencies between the two integration accounts, I conclude that a type of univocal view is difficult to defend. And, since that type of univocal view is the main reason for a subpersonal epistemology, the need for it does not arise. Next, I locate Clark's main motivation for his subpersonal epistemology in his account of the predictive brain. While the predictive brain provides useful insight into how an external process might extend one's cognition, I show it runs into problems when it tries to account for epistemic integration. Then, I explore how the concept of epistemic defeaters informs epistemic integration. Agents have to meet a specific no-defeater condition to be sensitive to the reliability of their belief-forming processes and to employ

them responsibly. In an entirely subpersonal epistemology, defeasibility theory has no means to explain how an agent is rendered sensitive to the reliability of her process. Subpersonal mechanisms can become sensitive to the new process, but that does not allow the agent to employ said process responsibly. Finally, I use an AI case, based on Amazon Alexa, to argue that all accounts of epistemic integration ought to explicitly describe how one's subpersonal mechanisms link to one's whole person, even if the link is weak and indirect. Without this connection between one's subpersonal mechanisms and one's person, it can become difficult to ascribe beliefs to the agent. This becomes apparent when Alexa leads to ever-increasing layers of cognitive, mental, and agential extension.

## Lay Summary

The extended cognition thesis is the thought-provoking idea that our cognition can extend into the environment, depending on how we employ some of the resources around us. For example, we may engage with our phones in the same way that we employ our innate cognitive faculties of perception and memory. Accordingly, if our engagement with our phones is on par with how we employ our innate cognitive faculties then the former should be considered just as cognitive as the latter. With this revolutionary idea, comes another one. Can we use these external resources in a way that gives us knowledge? After all, some of our cognitive processes are concerned with forming beliefs, and if our belief-forming cognitive processes were to extend into the environment they may be in the position to provide us with knowledge (or extended knowledge). My dissertation is concerned with this debate about extended knowledge. Primarily, I examine how a process extends my cognition and how it develops into a process that can give me knowledge, i.e. how can I achieve extended knowledge?

According to the recent literature on this debate, I can be involved in the process of acquiring extended knowledge in two ways. According to the first possibility (let's call it A), I should employ the said resource in a seamless, passive, and non-conscious way, which is in tune with how I invoke my innate neural faculties. This is necessary for the cognitive process to extend. And, for this resource to generate knowledge for me, I should employ it seamlessly and non-consciously but in a way that I can indirectly monitor if there was anything wrong with the said resource. So, I should not be actively or consciously employing the resource but I should have conscious access to it, for crucial times.

The second view (call it B) disagrees with the last part where I have conscious access to how I am employing my process. It concurs that for a process to extend my cognition I will have to employ it passively and seamlessly. However, for this process to produce knowledge for me, I do not have to be in the position where I can become aware of my process's malfunction. In fact, on the second view, I can sometimes acquire knowledge from an external resource even if I have no conscious access to my employment of the said

external resource. What is important is that parts of my cognitive system can appropriately exchange information with the said external resource and that can sometimes be sufficient to attribute knowledge to me.

I take issue with proposal B. I understand knowledge as a kind of success that is creditable to me when I have worked for it in some sense. Proposal B suggests that knowledge can be credited based on entirely consciously inaccessible mechanisms that occur in my cognitive system. I conclude that proposal A aptly describes the role I play in employing my resource responsibly and why extended knowledge should be creditable to me.

## Acknowledgements

I want to begin by thanking my principal supervisor, Duncan Pritchard, for reading through the many drafts I sent his way in these four years. His invaluable feedback and advice have helped me shape my ideas and polish my arguments. Without Duncan's help, this PhD thesis would not be in the shape it is now. I would also like to thank my secondary supervisors, Orestis Palermos and Mog Stapleton. Discussions with Orestis inspired me to select this topic at the end of my first year. He has also given detailed and extremely useful feedback on most of my chapters. Mog has been an encouraging supervisor and has pushed me to write better philosophy. If the thesis makes sense, it is because Mog nudged me to explain my ideas better. I would also like to thank my new principal supervisor, Nick Treanor, for helping me submit my thesis.

I am grateful for my time at the University of Edinburgh and the city of Edinburgh. It has been one of the most exciting chapters of my life. I am pleased that I was able to attend several workshops and seminars on different areas in philosophy at the University of Edinburgh and other institutes, during my PhD. I am also glad that I received the opportunity to present at, and organise, some of these workshops and seminars. Some of these events led to engaging discussions on my work with Sandy Goldberg, John Greco, Rob Rupert, and Ernest Sosa.

I thank the Higher Education Commission of Pakistan for their grant in support of my Phd.

I would also like to acknowledge people close to my heart who rooted for me at every step and cheered me on even in the worst of times. I thank my ami and baba for all their love and support, and my siblings, Rawail and Faiz, for always being there. They both provided feedback on some of my chapters. Faiz even shares my interest in this topic; he and I have had many passionate debates on some of the arguments I discuss in my thesis. I want to thank Khadijah for reading through my chapters and pointing out some seriously embarrassing mistakes. I also want to thank Talha and Nabeel for their encouragement. I am grateful for my lovely Edinburgh friends, Kate Webb and Rie Iizuka. They made it easier for me to live so far away from home.

Above all, I am grateful to Julian Hauser for giving me comprehensive comments on all my chapters, for building a work-from-home environment with me, for introducing me to coding for academic writing, for providing technical help, for feeding me this last year, for consoling me through the hard times, and for being a wonderful partner.

I am dedicating this work to Irteza. He took care of the rest of my world so I could take time out for myself– for a PhD in Philosophy.

# Introduction

## Overview

This thesis is concerned with the epistemological ramifications of the *extended cognition thesis* (Clark and Chalmers 1998), which is the revolutionary idea that cognition is not all in the head. More specifically, it suggests that we can employ some features of our environment in a way that engaging with them becomes a genuine part of our cognitive process. For example, an Alzheimer's patient may rely on her notebook in a way that her process of employing the notebook to store her beliefs may be considered her extended cognitive process. Some of this cognition, that can extend beyond our skin and skull boundary, may be involved in generating epistemic states such as knowledge. This is the hypothesis of *extended knowledge*.

Duncan Pritchard (2010) brings support for extended knowledge via an account of knowledge that is based in virtue epistemology. In his understanding, such an account can be useful to grasp the concept of knowledge in cases of extended cognition. According to this view, knowledge ought to be understood as a cognitive success that results from at least significant manifestation of one's cognitive ability. Our cognitive abilities are knowledge-conducive, reliable belief-forming cognitive processes. What makes these reliable processes knowledge-conducive is the fact that they are appropriately integrated into our cognitive systems. These reliable belief-forming processes do not necessarily need to belong to our neural apparatus; they may be processes that reside in the environment. The key factor is that they are epistemically integrated into our cognitive system. For example, an Alzheimer's patient can acquire extended knowledge by recruiting her notebook (i.e. retrieve stored data from the notebook), as long as her resource is appropriately integrated into her cognitive system.

Epistemic cognitive integration, therefore, explains how a process can appropriately integrate a reliable belief-forming process into one's cognitive system. It does not matter if the said belief-producing process is non-neural or lies outside one's body. When a process epistemically integrates into the agent's

cognitive system, she can take epistemic responsibility for this process. And, therefore, the agent can employ it in a way that manifests her cognitive agency.

We have two accounts of epistemic integration that fit well with the extended cognition thesis. The first is the *weak epistemic integration* that Pritchard proposes as a snug fit for extended cognition (2010). On this view, a reliable-belief forming process can integrate into the agent's cognitive system if it has developed 'cooperative interactions' (Greco 2010, 154) with other processes in the agent's cognitive system. The epistemic agent is able to monitor the reliability of her target process in the background in virtue of the interconnectedness of her cognitive system. Consequently, if her target belief-forming process were to malfunction, she would become aware of the problem. This is how the agent epistemically integrates her belief-forming process, on Pritchard's account.

Andy Clark, one of the main proponents of the extended cognition thesis, finds the weak epistemic integration account too active to be consistent with the extended cognition thesis (2015). He argues that our cognition can only be extended when we use external vehicles as seamlessly as we employ our innate neural cognitive processes. He calls this critique the *epistemic hygiene dilemma*. Accordingly, extended cognition requires that the target process is employed and *metaphysically integrated* into one's cognitive system without any agential involvement whatsoever. Whereas, weak epistemic integration requires that the agent should be in a position where she is involved in the epistemic integration on a (somewhat) personal level. That is, the agent can become aware of the failing reliability of her process. This, for Clark, is not compatible with the kind of metaphysical integration that extends our cognition. Accordingly, Clark calls for a more passive version of epistemic integration. This brings us to the second account of epistemic integration.

The second account, which I call *subpersonal epistemic integration*, is based on Clark's predictive brain model (2013, 2015). He proposes that, in cases of extended cognition, an entirely subpersonal exchange between our cognitive system and the target process can bring the latter (epistemically) on-board. There is no need for the agent to even be in the position to become aware of her malfunctioning process; it can all be done on a subpersonal level.

My thesis first questions the need for a wholly subpersonal epistemology and, then, I present a collection of arguments against this proposal. To argue for the first part of my thesis I bring attention to the worry that Clark has against the weak epistemic integration account. He finds it active in the sense that it involves the agent on a personal level. I argue that if we are not looking for a symmetric entailment between epistemic and metaphysical integration, in accounts of extended cognition, then there are no grounds to reject the weak epistemic integration approach based on the epistemic hygiene dilemma. The need only arises if we want to prove that the same conditions that allow metaphysical integration of a process can also allow epistemic integration of the said process. Why should we make this demand when all cognitive processes are not also epistemic?

In the rest of my thesis, I examine Clark's subpersonal epistemology and its implications. One of the objections that I bring against a fully subpersonal epistemology, based in the predictive account, relates to how Clark describes the predictive brain model. According to his account, an operation of *precision estimation* assigns reliabilities to the candidate processes on the basis of which they are integrated (and employed) or ignored by one's cognitive system. This is the process that characterises subpersonal epistemic integration. He further adds that this precision estimation mechanism is ubiquitous in the sense that it occurs at all levels of the brain's functional hierarchy. This entails that the process of epistemic integration does not only occur on the level of belief-formation but also takes places on subdoxastic levels that have nothing to do with beliefs. Subdoxastic states are associated with information that operates below the level concerned with beliefs, for example, wavelength information that your receptors detect. Clark claims that the brain's functional level that deals with wavelength information or other consciously inaccessible information is also subject to the operation of precision estimation— that is, for Clark, subpersonal epistemic integration.

Another disagreement I level against Clark's account is that it is not sufficiently different from Pritchard's weak epistemic integration account. The one thing that distinguishes Clark's epistemic integration from Pritchard's is that the former is fully subpersonal and not linked to a personal scenario. Via

the Temp case that Clark sites as an example of his predictive processing agent, I demonstrate that his epistemic integration is linked to only one personal level scenario. This is the same personal level account that Pritchard asserts; that is, the agent will become aware of any counterevidence to the reliability of her integrated process and act accordingly. When the integrated process of Clark's predictive processing agent begins to malfunction, the agent does become aware of the counterevidence to the reliability of her process. I suggest that Clark should dissociate his entirely subpersonal epistemology from that one personal level event in order to make the case that his account is sufficiently different from Pritchard's. After all, a subpersonal epistemic integration implies that any number of possibilities can take place at the personal level. Then, why is there only one personal level scenario possible when Clark's subpersonally integrated process malfunctions?

I also consider a subpersonal defeasibility theory for Clark's epistemic integration approach and evaluate how well it informs epistemic responsibility. A defeasibility theory provides an account of how an epistemic agent deals with defeaters, which are propositions the agent may be aware of or unaware of that attack the truth of her belief or its source. I devise a subpersonal defeasibility account to fit Clark's subpersonal epistemic integration account and find it unsatisfactory in explaining how an agent can responsibly employ her integrated process.

Implications of an entirely subpersonal epistemology are dire as well. I put forward two worries that a wholly subpersonal epistemology, in cases of extended cognition, will need to address. Firstly, there is the issue of *knowledge bloat* or the pernicious leaking of knowledge in the environment. Secondly, and relatedly, a fully subpersonal epistemic integration can make it difficult to *demarcate the epistemic agency*.

The knowledge bloat problem that I am going to elaborate is different from the cognitive bloat worry that is often used to critique extended cognition in general. According to the latter, if we do not have an appropriate understanding of how an external vehicle becomes a part of our cognitive system, we run the risk of assigning the 'extended cognitive process' label to far too many external resources. For instance, if the criteria for cognitive extension is

reliable availability and trust, then perhaps google on our phones (that is reliably available to us and we can trust most of its information) can also be considered our extended cognitive process. Cognitive bloat, therefore, makes it look like cognition is leaking into the environment uncontrollably. This cognitive bloat may lead to a knowledge bloat situation, i.e. we can know everything we can look up on our Google phones.

I formulate my knowledge bloat worry differently. Consider an advanced AI system that is present in my home, my phone, smartwatch, smart car, etc. Also, suppose that the AI meets the necessary requirements for extended cognition, and appropriately extends my cognition in the environment. Now, it is the subpersonal exchange of information between the AI and my cognitive system that will determine if a belief is being realised. If the subpersonal exchange is appropriate I may have an extended belief, in this case, that is being realised by the non-neural structure of the AI and neural apparatus of my cognitive system. Here, I ask if knowledge can also be determined in terms of mere subpersonal exchange of information between the AI and my cognitive system.

Furthermore, I explore the possibility of multiple layers of extension. For instance, the advanced AI, I have just discussed, may be capable of further extending my cognition into other advanced technologies, and the latter may be able to extend my cognition another fold, and so on. After all, what is necessary for cognitive extension is how my cognitive system and these technologies interchange information. So, I might have several extended beliefs depending on how information is subpersonally being exchanged amongst these technologies. I may have one belief that is being realised by my cognitive process and device A, another that is being realised by my cognitive process, device A, and device B, and so on. The problem is, if knowledge (like belief) is also determined on the basis of subpersonal exchange between processes, then there may be a lot of cases of knowledge in these layers of extension. Some of this knowledge may result from the employment of technologies on the third, fourth, or fifth layer of extension. I may have no idea that I even employ these resources, and, yet, I may be said to know beliefs that are realised in these advanced vehicles. Generally put, knowledge will overextend and leak perniciously into the environment if understood in terms of a wholly subpersonal integration.

On a related note, a solely subpersonal epistemology will also have to address how it demarcates the epistemic agency. The problem begins in the domain of beliefs. If a belief is partially realised by my cognitive system and partially by resources A and B, then where is the cognitive agency that the belief ought to be ascribed to? Similarly, if this belief is knowledge, then where is the epistemic agency that this knowledge is creditable to? This, I find, is a more important question for virtue epistemologists. They understand knowledge as a cognitive success that is at least significantly creditable to the manifestation of one's cognitive agency. However, if this success is acquired on the basis of wholly subpersonal contributions of different processes and advanced external vehicles, then who or what is it creditable to and why?

If we relinquish the pursuit of a completely subpersonal understanding of epistemic integration, we do not fall prey to the problems of knowledge bloat and the delineation of epistemic agency. Further, if we take Pritchard's route to describe extended knowledge, we can establish that the knowledge bloat problem does not arise, and we can successfully demarcate the epistemic agency. The reason for this is that the weak epistemic integration account (unlike its wholly subpersonal counterpart) explicitly links the subpersonal goings-on to the epistemic agent. The epistemic agent, in order to epistemically integrate a candidate process, will have to develop counterfactual sensitivity to the reliability of the said process. Processes that do not link to the personal level of the agent in some way cannot be employed to generate knowledge. And, epistemic agents (in cases of extension) are subjects who are sensitive to their process's reliability.

## Motivation for the subject

My motivation for this subject comes from virtue epistemology, which I believe presents a good understanding of knowledge. Knowledge is valuable and we should regard it as a cognitive success that is creditable to the agent when the said agent manifests relevant cognitive ability towards this goal. Virtue epistemology provides support for the extended cognition thesis. It has the concept of cognitive ability. One's cognitive ability is a knowledge-conducive reliable belief-forming processes that may (partially) reside outside of our skin

and skull. Greco (1999) describes our cognitive abilities as our innate cognitive faculties as well as our acquired habits of thoughts. The latter may include our engagement with resources in the environment or technologies that we learn to rely on for our cognitive productivity. This understanding of cognitive ability can help elucidate and provide support to the extended cognition thesis. After all, if we can incorporate external processes that can provide us with knowledge, we can, most likely, also make sense of how external processes function as our cognitive processes.

However, Clark's description of how extended knowledge comes about, interferes with the 'virtue' in virtue epistemology. The 'virtue' in virtue epistemology applies to the ability intuition. Knowledge, more specifically, is a cognitive success that is achieved by manifesting a virtue, i.e. a cognitive ability. Clark's proposal detaches the said cognitive ability from the agent exercising it and only properly links it to her subpersonal mechanisms. But, for knowledge to be (even only significantly) creditable to the epistemic agent, it needs to be firmly established that her cognitive abilities are linked to her person and that the agent has played a part.

Another motivation for this topic came from the possibility of 'neuromedia' that Pritchard (2018a) and Lynch (2014) discuss. These are technologies that will integrate so seamlessly in our cognitive system that the agent will learn to employ them and rely on them the same way she does on her innate cognitive faculties. As the name implies, neuromedia are cognitive enhancements that can be fitted inside the agent's brain. Imagine if Google (or another AI assistant like Alexa or Siri) was implanted in our brains and we could recruit it the way we use our innate faculty of perception. We would have far more propositional knowledge than seems possible right now.

This is clearly not the way we know things right now, but virtue epistemology, and specifically the weak ability condition on knowledge, has an explanation of how employing neuromedia in this way can give us knowledge. We need a clear conception of what makes neuromedia our cognitive ability or how we can employ it in an epistemically responsible way. For the resultant knowledge to be creditable to us, we need to employ neuromedia in a way that manifests our cognitive ability. Is it possible for us to acquire knowledge

based solely on the subpersonal interaction between other processes in our cognitive systems and the neuromedia? What are the implications of conceiving knowledge this way? With these technologies being not only possible but right around the corner, these are pressing questions that we need to address. These are some of the questions and the overall direction that motivated me to pick up this topic for my research.

## Chapter descriptions

In chapter 1, I introduce the concepts required to affiliate with this debate. First, I unpack the metaphysical cognitive integration account that outlines what cognition looks like. I discuss Pritchard's weak epistemic integration and support it with Greco's concept of cooperative interaction between the target process and other processes in our cognitive system. And, I describe that such an account allows the agent to passively monitor the reliability of her process, allowing her to fulfil the epistemic responsibility condition. Furthermore, I compare the two integrations and evaluate the relationship between them. I come to the conclusion that there is some overlap between their conditions, but it is not a complete overlap. I argue for the latter by bringing forward cases that suggest that we can find epistemically integrated processes that have not undergone metaphysical integration in terms of glue and trust.

I begin chapter 2 by discussing problems that concern the glue and trust condition. I suggest the solution lies in considering complementarity style arguments for extended cognition. I elaborate and defend one such argument from Palermos. This argument paves way for more discussion on the relationship between conditions required for metaphysical integration and those that lead to epistemic integration. I take a stance on this matter, here, and defend partial overlap of these conditions (i.e. asymmetric univocal view). This chapter also introduces Clark's epistemic hygiene dilemma. Accordingly, he proposed that Pritchard's weak ability intuition is too active to accommodate metaphysical integration that allows cognitive extension. In lieu of Pritchard's epistemic integration, Clark puts forward a wholly subpersonal epistemic integration. I elucidate how his view defends a complete overlap between the conditions of the

two integrations (metaphysical and epistemic) and argue that there are many challenges against such an epistemic integration view.

The next chapter examines a specific claim about Clark's concept of metaphysical and epistemic cognitive integration. Clark envisions that advanced technologies will soon be able to immediately metaphysically and epistemically integrate into our cognitive system. In chapter 3, I analyse the notion of immediate integration. I find this concept inconsistent with Pritchard's weak epistemic integration. Moreover, this chapter questions the need for a subpersonal epistemic integration that advocates for a complete overlap between the conditions required for metaphysical and epistemic integration.

In Chapter 4, I base Clark's subpersonal epistemic integration in his predictive processing model of the brain. I argue that the foundation of this epistemic integration account is weak. As per the predictive brain model, there is a 'precision estimation' mechanism that occurs at all levels of the brain's functional hierarchy. It occurs at the lowest level of neural interactions and it occurs at the higher levels in which mental states are realised. The function of precision estimation units (a group of our cognitive processes) is to assign reliability weightings to target processes, on the basis of which the said processes can integrate into one's cognitive system. On Clark's account, this process allows for the minimal epistemic integration of a new process, in cases of extended cognition. However, since this mechanism occurs at all levels of the brain's functional hierarchy, it even occurs at levels that are not concerned with belief-formation. I argue, at this point, that epistemic integration is far too ubiquitous on Clark's view. Furthermore, with the help of a case, Clark shows how subpersonal epistemic integration takes place in a predictive processing agent. I analyse Clark's case and find that his predictive processing agent has only one possible personal level scenario, which is the same scenario that weak epistemic integration describes. That is, the agent becomes aware of a problem with her malfunctioning process. Whereas, on a wholly subpersonal integration any different number of scenarios should be possible on the personal level. I suggest, on this note, that subpersonal epistemic integration is wedded to a personal scenario that it shares with weak epistemic integration. This indicates that the former may not necessarily be an account that is, in any substantial

way, different from weak epistemic integration.

Chapter 5 discusses epistemic responsibility in terms of defeasibility theory. This chapter provides a thorough understanding of how weak epistemic integration deals with defeaters and also attempts to devise a subpersonal defeasibility for Clark's subpersonal epistemic integration. Defeaters, in general, are propositions, conscious to us or not conscious to us, that attack the truth of our belief or its source. How an agent deals with defeaters against her beliefs or the reliability of her source can tell us a lot about how responsible an agent is in employing her process in a particular way. So, when defeasibility theory is applied to epistemic integration, we get a more rigorous picture of how epistemic responsibility requirement is met. In this chapter, I suggest that weak epistemic integration has an effective understanding of how an agent deals with defeaters, and therein lies how this account meets the epistemic responsibility condition on knowledge. The idea is that when a process integrates via weak epistemic integration, the agent first becomes aware of most of what is required to monitor the reliability of her process and that allows her to responsibly employ such a process. There are two no-defeater conditions at play here. The agent ought to meet the no-normative-defeater condition, which requires that there should not be any defeaters (against the reliability of her source) that the agent ought to be aware of. Secondly, the no-conscious-defeater requirement should also be met. It requires that the agent is not aware of any defeaters against the reliability of her process. When these two conditions are met, the agent is able to responsibly employ her candidate process. The rest of this chapter attempts to devise a similar (in terms of no-defeater conditions) concept of epistemic responsibility for Clark's epistemic integration.

Chapter 6 elaborates how belief-formation can be understood in terms of a wholly subpersonal exchange of information between different processes. According to the extended cognition and the extended mind thesis, my mental state can supervene on physical states in my cognitive system and in an AI system (like Alexa) outside my head, if there is appropriate information exchange between the two systems. I think it is harmless to conceive of belief-formation this way. However, when we try to think of knowledge the same way we run into problems. Knowledge cannot be made sense of in terms of information exchange

between two systems. A personal level involvement is required to determine whose knowledge we are dealing with and how to delineate the epistemic agency. Secondly, a solely subpersonal integration like the one that Clark proposes, brings forward a new kind of knowledge bloat problem. If knowledge is to be understood only in terms of subpersonal exchange of information then we can imagine other external mechanisms that have the necessary subpersonal information exchange with Alexa and my cognitive system to further extend my cognition and mental states. Similarly, there can be another process that further extends my cognition the same way. Eventually, we will be in the position to have knowledge of a belief that is being realised by our cognitive system and several other advanced technologies, some of which we have never consciously employed. This, I call the knowledge bloat worry. The knowledge bloat and the problem of demarcating the epistemic agency are worrisome for a wholly subpersonal epistemology. I conclude by demonstrating how the weak epistemic integration account can successfully counter these worries.

## Chapter 1 The two integrations

In this chapter, I introduce the vocabulary and set the initial stage for the main debates of this dissertation. I begin by presenting the two cognitive integration accounts: metaphysical cognitive integration and epistemic integration. The first integration account is important for the thesis of extended cognition. It will help us understand what makes a process cognitive and, therefore, how our cognition may extend into the environment. The latter, epistemic integration, is fundamental to our virtue based understanding of knowledge. It describes how our reliable belief-forming processes can become knowledge-conducive cognitive abilities for us.

The extended cognition thesis is motivated by the *functional parity principle* and the *glue and trust criteria*. In these theories, I will look for the conditions for metaphysical cognitive integration. Similarly, in order to single out requirements for epistemic integration, I will examine Pritchard's *ability intuition* and Greco's notion of cognitive integration.

Further, I will compare the two integration accounts and the conditions I have for them to determine a relationship between them and evaluate how consistent they are with one another. My deduction, in light of Pritchard and Palermos' arguments, will be that these two integrations are a snug fit and have a univocal relationship.

The univocal view suggests that either the conditions for these integrations fully overlap (symmetric entailment) or partially overlap (asymmetric entailment). The former is the view that the same conditions that mark extended cognition can explain how one's reliable belief-producing mechanism becomes knowledge-conducive. I consider some evidence in favour of this view and also some problems that it can fall prey to. On the asymmetric view, conditions for metaphysical cognitive integration are understood as a subset of conditions for epistemic integration. I will provide support for the asymmetric univocal account.

Another important observation I make is that it is not very helpful to understand cognitive extension in terms of the glue and trust criteria. They do not provide an appropriate description of what a cognitive process is like, and

a more thorough understanding is required to evaluate the two univocal views—symmetric or asymmetric.

In terms of the bigger picture, this chapter introduces the background debate that will bring Clark's subpersonal epistemology (that explains extended cognition) into the scene. But, it is only in the next chapter that I explain how that happens, and then in further chapters, I investigate issues related to an entirely subpersonal epistemology.

## Metaphysical cognitive integration

The thesis of extended cognition states that our cognition is not always entirely in our heads (Clark and Chalmers 1998). Sometimes we use external vehicles or artefacts in a way that extends our cognition outwith our skin and skull boundary. Part of our cognitive process goes on inside the external vehicle, extending our cognitive process into the environment.

Consider the Otto-Inga example (Clark and Chalmers 1998). Otto is an Alzheimer's patient. He uses his notebook to record his memories as well as important everyday information that he comes across. He uses his notebook the way Inga uses her biological memory. At a conference, where they are both present, the speaker mentions an art exhibition at MoMA (Museum of Modern Art). Inga recalls from her biological memory that MoMA is on 53rd street. Otto locates MoMA's address in his notebook that he always keeps on him. He depends on his notebook the same way and for the same things that Inga relies on her biological memory for. Both of them make their way towards MoMA for the exhibition.

One of the foremost motivators for the extended cognition theory is the *functional parity principle*. According to it, as long as a process plays the same functional role as one played by our innate processes,<sup>1</sup> it should be treated

---

<sup>1</sup>I use the term innate to distinguish processes that, in some sense, humans are born with, for example perceptual faculties and to compare them to acquired processes like arithmetic skills. I understand that there is a debate in philosophy of mind and philosophy of cognitive science that requires a more concrete concept of what innateness is (see Gross and Rey 2012), and such a view can conflict with cognitive extension, in a way that proponents of extension will have to commit to a concept of innateness. In the way I use the term 'innate,' I do not

as a cognitive process.<sup>2</sup> In the Otto-Inga case, Otto's use of his notebook is functionally equivalent to how Inga relies on her biological memory. The principle of functional parity states that we should not treat Otto's process as any less cognitive just because a part of it exists outside of his head. Clark and Chalmers' describe the principle in the following way: 'If, as we confront some task, a part of the world functions as a process which, were it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world is (so we claim) part of the cognitive process. Cognitive processes ain't (all) in the head!' (1998, 8).

If we do not abide by the principle of functional equivalence or the parity principle, we run the risk of committing bio-prejudice (Carter and Kallestrup 2019). This means that we are giving unwarranted weight to location and neural constitution when all we need to be looking at is the function of the process and perhaps the behaviour that it brings about. Otto and Inga are both able to reach MoMA by relying on their respective processes. The parity principle emphasises that 'Extended processes should not have to meet higher standards merely because they are extended' (Sprevak 2009).<sup>3</sup>

The functional role of a process is often understood in terms of how the process is employed by one's cognitive agency. So, if Otto's agency employs  

---

presuppose a sharp distinction that this debate urges for. My intention is to merely distinguish between obviously external or acquired processes and the processes that we are in some sense born with (but may be triggered by external factors).

<sup>2</sup>The functional parity principle is motivated by the understanding that we ought to look towards functionalism to make sense of cognition. See Wheeler (2010) for a defence of functionalism.

<sup>3</sup>Rupert (2004) argues that the parity principle does not make a case for extended cognition. He discusses all the differences between a biological memory and a notebook and how these two resources are invoked very differently. In response, consider Menary (2010)'s argument that the parity principle alone is only an intuition pump for the extended cognition thesis, not an argument in its support, *per se*. Menary argues that arguments for the extended cognition thesis come from complementarity style views, which I will discuss in the next chapter. Rowlands (1999), Wilson (2004), and Menary (2007) discuss how the external artefact ought not be compared to the neural process. It is the practice of employing the artefact that should be compared to how we recruit our biological cognitive faculties. More emphasis on this debate can be found in Menary (2012), where he compares artefact extension with enculturated extension.

his process of operating his notebook the same way Inga employs her memory faculty, then both these processes can be considered functionally equivalent. One way to understand how one's agency employs a process is to invoke the concept of *cognitive integration*. The first to discuss cognitive integration in this context is Richard Menary (2006, 2007, 2009, 2012). He discusses the thesis of cognitive extension in terms of the integration of an external vehicle with an internal one. Accordingly, a process (of manipulating an external artefact, for instance) should be considered cognitive if it is appropriately integrated into one's cognitive system. In order to distinguish this cognitive integration from another kind of integration that I will discuss later, I will call this one *metaphysical cognitive integration* (henceforth MCI) following Carter and Kallestrup (2019, 5). Metaphysical cognitive integration, therefore, defines how an external vehicle, or our practice of manipulating an external artefact, becomes a part of our cognitive process. In what follows, I will discuss and analyse how MCI can be acquired.

Cognitive integration can be understood in terms of the functional parity principle. That is, if we come across a new process that exists outside our skin boundary we need to consider if it is integrated in our cognitive system like our innate processes. If the answer is affirmative, then it ought to be considered a cognitive process, no matter where it is located. In this case, then, it can be argued that if Otto's practice of engaging with his notebook is integrated in his cognitive system in the same way that Inga's biological memory is integrated in hers, then we should consider the notebook a part of Otto's cognitive system, just as we consider Inga's biological memory hers.<sup>4</sup> In the next section, I will discuss how conditions for metaphysical integration can be met (it will require looking at the conditions that our innate processes meet).

## The glue and trust criteria

The important questions to ask here are: How do we know that a process has appropriately integrated into our cognitive system? What are the conditions

---

<sup>4</sup>Note that it is not the artefact, the notebook itself, that integrates in his cognitive system, but Otto's practice of engaging with his notebook that integrates in his cognitive system. This is in line with the discussion from Menary (2012).

required to satisfy MCI? The answers lie in further examining the Otto case. Otto's reliance on his notebook presents some features that can help sketch an account for MCI. Such an account can help us identify cognition even if it lies external to our skin and skull boundary. Otto's process of manipulating his notebook has three important features called Clark's 'glue and trust' conditions (2010a, 84). They are:

1. That the resource be reliably available and typically invoked.
2. That any information thus retrieved be more or less automatically endorsed. It should not usually be subject to critical scrutiny. It should be deemed about as trustworthy as something retrieved clearly from biological memory.
3. That information contained in the resource should be easily accessible as and when required. (Clark 2010b, 47)

The glue and trust criteria require that for an external process to count as our cognitive process it should be in some sense glued to us and we should deem it trustworthy.

These three requirements are informed by the parity principle. Our innate processes are also glued to us, in a manner of speaking, and we can trust them completely. Otto's notebook is present to him in this same 'glue and trust' way that our innate neural processes are present to us.

Consider the individual requirements and how they portray our reliance on our neural processes. The first 'glue and trust' condition requires that the resource is present at all times that it is needed, so it can be typically employed. Otto carries his notebook all the time and is never without it, just like we always have our innate processes available to us when we need them. The second condition is about how Otto trusts the information from his notebook, just like we endorse the information that our cognitive faculties give us in typical conditions. The third criterion demands that the employment of the resource should be easily accessible for the agent. If our Alzheimer's subject was differently-abled in that he could not read what was in the notebook then the information of the notebook would not be accessible to him. So, it is not enough that we have the resource with us and can trust it but also that we can

access the output that the resource provides. Overall, if there is an external process that we can access anytime we need to, and whose information we can access, endorse and trust without scrutiny, then we are most likely using it the way we use our innate cognitive faculties. Such an external process will most likely meet the MCI requirements, and, therefore, it should be considered cognitive.

### What is a univocal view?

Our cognitive processes may be epistemic or non-epistemic. The former are processes that are directed towards epistemic goals, i.e. getting a belief right or acquiring knowledge. Not all cognitive processes are epistemic. Coordinating myself through the house and preparing a meal for myself would involve cognition that is not oriented towards epistemic goals.<sup>5</sup> By epistemic goals, I mean the aim of getting beliefs right. Even belief-production can include cognitive processes that are not epistemic, like the movement I make towards an epistemic goal.

Extended cognitive processes can also either be epistemic or non-epistemic. The extended epistemic processes will be external mechanisms that we rely on to fulfil our epistemic goals. Otto seems conscientious enough to want to get beliefs right and to acquire knowledge from his reliance on his notebook. It can be inferred then that Otto's extended process is not just a cognitive process, but an extended epistemic process as well. And, if his extended epistemic process is able to form a true belief that is knowledge, then Otto has what we may call, *extended knowledge*. It is this concept of extended epistemology that urges us to compare the integration that marks cognition with the integration that allows belief-producing processes to become knowledge-conducive.

We understand cognition in terms of how our cognitive agency recruits a cognitive process, i.e. in accordance with the MCI conditions. Just like this, epistemic processes can be understood in terms of how our cognitive agency employs them. To give an account of how our agency employs an epistemic

---

<sup>5</sup>It may involve some epistemic processes as well like forming beliefs about where things are, but there are also cognitive processes involved that are not epistemic, nor concerned with forming beliefs, e.g. paying attention to the surroundings.

process is to provide conditions for 'epistemic cognitive integration' (ECI) (Carter and Kallestrup 2019, 5). I will discuss a detailed account of ECI in the coming sections. In this section, I will focus on explaining the need to compare the two integration accounts.

Comparing the two integrations can help us make sense of extended epistemic processes that can generate extended knowledge for us. These processes fill some ECI conditions that are consistent with the MCI account that describes extended cognition. MCI describes what makes an external process become a cognitive process by identifying what is needed for cognition. Hence, it is essentially an understanding of what cognition looks like.<sup>6</sup> Further, to pursue extended epistemology, we need an ECI account that is in line with MCI, which should not be very hard to find because all our neural epistemic processes are all, also, cognitive processes. This means that they satisfy some kind of MCI as well as ECI. Now, it is also possible that these two integrations are not just consistent but that their conditions overlap in some way. So, a relationship between ECI and MCI already exists, and we need to find what it is. After all, it should be an essential aspect of a neural epistemic process that it is also a cognitive process. This means that on the course of understanding ECI, we may be able to make sense of MCI as well, and vice versa. In other words, the way our agency employs an epistemic process can involve an understanding of how an agency recruits a cognitive process. So, by theorising conditions for ECI, we should also be able to get an understanding of the MCI conditions. This is called a *univocal view* (Carter and Kallestrup, 2019).

The univocal view states that there is a strong correlation between the conditions required for MCI and the ones that necessitate ECI. Carter and Kallestrup describe the univocal view as, 'an external artifact E is metaphysically cognitively integrated if, and only if, it is epistemically cognitively integrated' (Carter and Kallestrup 2019, 5). This is a strong interpretation of the univocal view and suggests that these two integrations have a *symmetric entailment*. According to the symmetric view (or strong univocal view), the conditions for

---

<sup>6</sup>Note that I have only expressed the MCI in terms of being glued to or trusting an external process. This may not be the only way or the best way to explain cognition. I will discuss this point in detail later.

ECI and MCI are the same. If your process fulfils the MCI criteria then it is also epistemically integrated, and *vice versa*. There is also a weaker univocal view or the *asymmetric entailment* between the two integrations. This proposal is that there is a significant overlap between the conditions for epistemic integration and the conditions for metaphysical cognitive integration or that the MCI requirements are a subset of ECI condition. I will discuss both of these accounts in detail later and find that asymmetric approach better describes the relationship between the two integration accounts.

An important aspect of the univocal view is that it gives an epistemological understanding of a philosophy of mind concept, and, in return, a philosophy of mind theory supports an epistemological account. So, both accounts are being strengthened in some way. It also brings together two concepts of cognitive integration: one that can explain cognition and the other that can explain knowledge-generating processes.<sup>7</sup>

In contrast to the univocal view is the understanding that these two kinds of integrations are very different from each other, and hence their merger is not as simple, or it is not possible at all. A principal difference between these two integrations is that they have completely different aims. Epistemic integration endeavours to explain what makes a process an epistemic process, one that is likely to generate knowledge. Whereas, metaphysical cognitive integration describes how one can employ an external resource as a cognitive process. It helps to identify what cognition looks like. So, while the univocal view suggests that understanding the former set of conditions can give us an idea about the latter, and *vice versa*; it may not necessarily be the case. It can also be that none of the conditions for ECI overlap with conditions for MCI. The process might need to meet an entirely different set of requirements to be considered cognitive. We can call this the *non-univocal view*.

So far I have only discussed what a univocal view looks like. To evaluate its feasibility, I will need to unpack an example of such a view, and for that, we first need to settle the question of what ECI conditions best explain a knowledge-conducive process. Then, I need to make sure that the ECI account

---

<sup>7</sup>Some variations of the univocal view have received support from Pritchard (2010), Palermos (2011), and Palermos (2014b).

is consistent with MCI, and can allow extended cognition. I will not be able to accomplish all these tasks in this chapter but I will present literature that is required to evaluate the univocal view.

## The ability intuition and extended cognition

It is time to flesh out the details of the virtue epistemology account that lends support to the extended cognition thesis. It will be clear by the end of this section how epistemic integration can shed light on the theory of extended cognition.

An internalist epistemology demands that the agent have access to whatever it is that justifies her belief. So, if it is the reliability of a belief-forming process that justifies a belief, then internalism requires that the agent should have access to the reliability of her process. Epistemic externalism, in contrast, would not demand that the agent have access to the reliability of her belief-forming process. A reliable process will form reliable beliefs even if we have no awareness of the reliability.<sup>8</sup>

Process reliabilism is an externalist epistemology. According to it, only beliefs formed by a reliable-belief forming process can be in the running for knowledge, where reliability means that the process forms far more true beliefs than false. Since process reliabilism is an externalist epistemology, it does not require the agent to have access to the reliability of her process to justifiably hold the belief from this process or have knowledge of it.

Process reliabilism, however, runs into a problem because it cannot describe how Norman, the clairvoyant does not have knowledge despite employing a reliable process for his beliefs (Bonjour 1980; Bonjour 1985). Norman has an odd but reliable faculty that tells him when the president is in New York. However, he has no clue that he has such a faculty or that it is a reliable one. Let's say that the president is in New York and Norman, sitting at home, forms the belief via his clairvoyant faculty that the president is in

---

<sup>8</sup>For a detailed understanding of epistemic externalism, see Goldman (1986). For discussion on active epistemic externalism from a virtue epistemology standpoint, see Carter and Palermos (2015). A defence of internalism, i.e. the need for reflective access to the reasons for one's belief being true, can be found in Chisholm (1977) and Pryor (2001).

New York. Does Norman have knowledge? Intuition suggests that Norman does not have knowledge. Evidently, a reliable process alone is not sufficient for knowledge, and, therefore, process reliabilism is missing something.

Virtue reliabilism points out what process reliabilism is lacking.<sup>9</sup> The issue at hand is that Norman has no connection to the reliability of his process. He has no access to the reliability of his belief-forming mechanism. He may hear in the news that the president is in New York and find out that this particular belief of his was true, but this information does not give him a link to the reliability of his process. Things will be different if Norman appropriately relies on his clairvoyance faculty for a long time. Let's suppose, that Norman forms beliefs via his clairvoyance faculty all the time and that he has done so since a very young age. He has formed beliefs about where the president is and they have always come out to be true. It seems that when understood in this way, Norman's reliance on his faculty is not odd. What has changed in all these years is that Norman has found a connection to the reliability of his process.<sup>10</sup>

Another way that Norman can connect to his reliability is if he finds direct reflective access to it. For instance, if a doctor who deals in odd faculties explained to Norman that his mechanism was reliable and why it was so. This strong awareness condition is sufficient to connect to the reliability of one's process, but it is not necessary. As explained above, there is a weaker route that Norman can take to connect to the reliability of his process.

The missing connection to one's reliability that Norman demonstrates can be referred to as a kind of 'subjective justification' (Greco 2010, 166). If Norman had a sort of subjective justification to the reliability of his process he

---

<sup>9</sup>Another form of virtue epistemology is virtue responsibilism. Virtue responsibilists argue that merely forming true beliefs via reliable cognitive faculties does not suffice for knowledge. Instead, a knower ought to manifest intellectual virtues, that are character traits an agent maintains through training and self-control. Examples of these virtues are open-mindedness, epistemic courage, epistemic humility, etc. For a thorough comparison of virtue reliabilism and virtue responsibilism, see Pritchard (2016a). For a discussion on extended epistemology in terms of extended intellectual virtues, see Pritchard (2018a) and (2018c).

<sup>10</sup>Greco (1999) explains this in terms of the problem of strange and fleeting processes. Some processes, even though they are reliable, cannot give us knowledge because they are either strange or fleeting. It is important that an agent gets to know her process in a way that it is no longer strange or fleeting for her.

could be in the position to have knowledge. It is important not to confuse this externalist understanding of subjective justification with the internalist concept of having reflective access to the reliability of one's process. The idea is not that Norman should have access to the reliability of his process, but that he should have some sort of indirect connection to it, (which is not a direct awareness condition) which is in keeping with the externalism of reliabilism.

An important point to note about the subjective justification requirement is that it allows the agent to responsibly employ her reliable process. Initially, Norman's case is not a case of knowledge because he is not employing his reliable clairvoyance in an epistemically responsible way. This is why the result of his reliable process is not attributed to him. He has to take responsibility for his process so that the belief that results from his use of this faculty can be his.<sup>11</sup> So, subjective justification of this sort allows Norman to meet the responsibility requirement on knowledge. The concept of subjective justification or epistemic responsibility paves way for virtue reliabilism. And, the example of Norman suggests that we should give up process reliabilism in favour of virtue reliabilism.

Generally, virtue reliabilism is the position that an epistemic agent ought to responsibly recruit her reliable belief-forming process to acquire knowledge. More specifically, it advocates that the agent's belief-producing process should be (1) reliable, and (2) it should meet the subjective justification requirement. Pritchard suggests that this requirement be understood in terms of the *ability intuition* on knowledge (2010). According to this intuition, knowledge is a kind of cognitive success (true belief) that is achieved by manifesting a virtue, called the cognitive ability. Just like success in archery comes from manifesting one's archery ability, so does knowledge come from exhibiting one's cognitive ability.<sup>12</sup>

What is cognitive ability? An agent's cognitive ability is her reliable belief-forming process that manifests her cognitive agency, by connecting her agency to the reliability of her process (fulfilling the subjective justification requirement) (Pritchard 2010). But, for our purpose of comparing integration

---

<sup>11</sup>I discuss epistemic responsibility in terms of how beliefs are attributed to us in the last chapter of this dissertation.

<sup>12</sup>This view of knowledge as a cognitive achievement that results from one's manifestation of cognitive ability has also been furthered by the following: (Sosa 1991; Greco 2010, 2012; Pritchard 2007, 2009a, 2009b, 2010).

accounts, it is important to understand cognitive ability in terms of cognitive integration. Accordingly, one's cognitive ability is one's reliable belief-producing mechanism that is properly integrated into one's cognitive character. And, one's cognitive character is made of one's reliable, dispositional and stable, natural as well as acquired cognitive faculties (Greco 1999).

Pritchard discusses that to understand cognitive ability only in terms of a reliable belief-forming process is too weak. It makes us run into problems similar to Norman the Clairvoyant case. He has a reliable process but his process does not manifest his cognitive agency. In other words, his reliable belief-forming process does not exhibit cognitive ability on his part. How then should we understand cognitive ability? The answer, as I discussed above, is located in how the reliable belief-forming process is integrated in the agent's cognitive character. It ought to be integrated such that if knowledge is produced as a result of its employment, this knowledge can be at least significantly creditable to the manifestation of one's cognitive ability.

Before I discuss how one's cognitive ability arises from epistemic integration, let me first introduce two possible relationships between cognitive ability and knowledge. The credit account of knowledge, that virtue reliabilists subscribe to, suggests that knowledge is a kind of cognitive success that is creditable to the manifestation of one's cognitive agency.<sup>13</sup> On the strong cognitive agency account, one's cognitive success can be knowledge only if it is primarily creditable to one's cognitive agency, whereas on the weak cognitive agency account, one's cognitive success need only be significantly creditable to one's agency. What this means is that on the strong agency account, that one's cognitive agency plays a primary role to bring about one's cognitive success. On the other hand, the weak account only requires that one's cognitive success is brought about a significant (not primary) causal role that one's cognitive ability plays.<sup>14</sup> This account acknowledges that other factors in the environment also affect one's cognitive success.

---

<sup>13</sup>For Palermos' discussion on credit account of knowledge, see (2016). Also, see Baehr (2012) for a discussion of credit theories of knowledge in terms of the value of knowledge debate.

<sup>14</sup>For a comparison of weak and strong accounts of agency, also see Pritchard (2008); Pritchard (2009a); Pritchard, Millar, and Haddock (2010)

Let me explain these accounts in terms of the Norman case. In order to satisfy the strong agency account, Norman's cognitive success ought to be primarily creditable to his cognitive ability. And according to the weak agency account, his success only needs to be significantly creditable to the manifestation of his ability. How can he satisfy these requirements? Note that if he meets the strong requirement he automatically fulfils the weak condition. Pritchard asserts that what would suffice for the strong agency account is that Norman become aware of the reliability of his process and what makes it reliable (2010). Of course, such a condition is not necessary to meet the ability intuition. This is why, on the weak account, it suffices that Norman is sensitive or responsive to counterevidence to the reliability of his clairvoyant faculty.

These two accounts to satisfy the ability intuition also provide insight into how one's process should be epistemically integrated into one's cognitive character, and therefore it describes the role one's cognitive character plays in bringing about their cognitive success. Pritchard describes these two accounts of ability in the following way:

(COGAWEAK) If S knows that p, then S's true belief that p is the product 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.

(COGASTRONG) S knows that p iff S's true belief that p 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 (2010).

Both these accounts can explain how one's reliable belief-forming mechanism can become integrated into one's cognitive character. On the strong account, the internalist condition of gaining access to the reliability of one's belief-forming process will do the trick. Consider Alvin who has a serendipitous brain lesion (Plantinga 1993). The lesion is reliable in that it allows Alvin to form true beliefs about the weather outside. On the strong account, Alvin's process will integrate in his cognitive character if he gains reflective access to

the reliability of his process. Let's say that Alvin comes across a newspaper article about these kinds of lesions and reads all about their reliability and how to use them. Now, the cognitive success or true belief that our epistemic agent forms by deploying his lesion will be primarily creditable to him. Pritchard writes, 'Alvin has in this way integrated this belief-forming process within his cognitive character to a sufficient degree that his cognitive success is now primarily creditable to his cognitive agency, rather than being creditable to something external to his cognitive agency' (2010, 226).<sup>15</sup>

According to the weak agency account, an epistemic agent can exhibit cognitive ability by being only significantly (not primarily) involved in the process that results in the true belief.<sup>16</sup> Let's reconsider Alvin's case but bring down Alvin's involvement towards his cognitive success. Perhaps Alvin did not read about his lesion in the newspaper and, therefore, does not have direct access to (the existence of) his process or its reliability. However, Alvin has used his lesion for 10 years now. He uses it every day to form true weather beliefs. When the lesion tells him it will rain, it does rain. And, when it tells him it will snow, it does snow. By employing his process for a while, he develops sensitivity to the reliability of his process. His process will be integrated in his cognitive character according to these weaker grounds. He may not have direct access to the reliability of his lesion, but he has gained a sort of sensitivity or responsiveness to its reliability.<sup>17</sup>

---

<sup>15</sup>Pritchard argues against the strong agency account by discussing *epistemic dependence*. According to this concept, knowledge is both positively and negatively influenced by factors that are outwith one's cognitive agency. Positive dependence can be observed when external factors allow an agent to acquire knowledge even though her own cognitive agency is not sufficient to achieve that result. External factors may also keep one from acquiring knowledge, despite the agent manifesting a high degree of cognitive agency Pritchard (2016b).

<sup>16</sup>Note here that Kelp (2013) argues that the strong version of the ability condition (robust virtue epistemology) will have no problem accommodating the extended cognition thesis. It depends on how Otto uses his notebook and if his cognitive success can be primarily creditable to him. If his wife is helping him with how he relies on his notebook, then his cognitive success is not creditable to him.

<sup>17</sup>The sensitivity that is being discussed here is the agent's counterfactual sensitivity to her process's reliability, which means that if her process was unreliable she would become aware of it. This sensitivity is unrelated to the 'sensitivity principle' in epistemology that discusses

The weak agency account proposes that the subjective justification requirement can be met by minimal agency involvement. Even without direct access to the reliability of our processes, we can still use them as our cognitive abilities.<sup>18</sup> What is important is that we are (or become) responsive to the reliabilities of our belief-forming mechanisms. In the Alvin case, our agent had relied on his odd faculty for so long that he had become sensitive to its reliability. We can be sure that Alvin has gained sensitivity or responsiveness to his process's reliability if he is in the state where he can become aware of any problem with the reliability of his process. This clause measures the degree of sensitivity that the agent must acquire. Evidently, the level of sensitivity ought to be such that the agent becomes aware of any counterevidence to her process's reliability.

Both weak and strong agency accounts can be used to explain how a belief-forming process that lies outside one's skin and skull boundary can become a part of one's cognitive system. What needs to be accomplished is that the epistemic agent can take responsibility for the said process by integrating it into her cognitive character. Alvin's lesion is not a naturally occurring faculty, of course. And, even though it is present inside his brain, it is still strange and anomalous. Nevertheless, Alvin can take responsibility for this process and integrate it into his cognitive system by either acquiring access or sensitivity to the reliability of his process. This understanding gives us the hint that the ability intuition is compatible with, and can help understand, the extended cognition thesis.<sup>19</sup>

The way integration is defined in the strong agency account is not in line with the functional parity principle. The parity principle motivates extended cognition by showing that we employ some external resources the same way we employ our innate faculties. We do not employ our innate processes like perception and memory by gaining access to their reliabilities. We rely on them automatically and seamlessly. Hence, our dependence on our neural resources is more in line with the weak agency account, where there is no direct access to

---

attempts to describe an anti-luck condition on knowledge (Pritchard 2007).

<sup>18</sup>The concept of cognitive abilities can be traced back to Sosa (1988, 1993).

<sup>19</sup>Roberts (2012) discusses how the agent has to meet the epistemic responsibility requirement for her extended belief-forming process to develop into an extended epistemic process.

the reliability of one's process. This is why I agree with Pritchard's argument that the weak agency account that does not demand direct reflective access to the reliability of one's process is in line with extended cognition because it conforms to the functional equivalence principle.

In proposing that the weak ECI account is in line with functional parity, Pritchard is gesturing towards the univocal view. His argument is that when an external process develops into a cognitive ability, by integrating into the agent's cognitive character via the weak ECI route, it does so in a way similar to how our innate processes integrate into our cognitive character. So, in a way, satisfying the weak epistemic integration conditions takes care of fulfilling the functional parity requirement, which is also what MCI is supposed to describe. This means that Pritchard's argument is a step in the direction of the univocal view; that is, in meeting the weak ECI requirements, the agency takes care of whatever the MCI requirements are.

So far, I have not singled-out the requirements for ECI. It should be clear from my discussion, though, that these conditions will be based in Pritchard's weak agency account. In what follows, I will introduce Greco's concept of epistemic integration to derive the ECI conditions that can further a univocal view.

## Epistemic integration

In Pritchard's weak agency account, we find how an agent can take responsibility for a belief-forming process (and make it hers) even if she does not have access to the reliability of the process. It seems that the belief-forming process ought to be reliable and should, via the weak route, *integrate* into one's cognitive agency. John Greco provides a cognitive science picture to this epistemology. He is the first to discuss epistemic integration in terms of epistemology, and he provides an informative picture of what integration of a process would look like in the head. In this section, I will elucidate Greco's account of epistemic integration and attempt to list the conditions for epistemic integration.

The weak and the strong agency accounts provide us with an understanding of epistemic integration, which means they describe how

one can make a certain belief-forming process theirs and take responsibility for the said process. Once a process epistemically integrates into the agent's cognitive character, the agent can meet the subjective justification requirement and the epistemic responsibility condition. The way a process integrates into one's cognitive system provides an understanding of how cognitive agency is manifested. A higher degree of agency is manifested if a process is integrated on the basis of direct access to her process's reliability. And, a lower degree of agency is manifested when the agent only develops a sensitivity to the reliability of her process.

First, let's reconsider what a cognitive character is. This will also tell us what a new process has to be like to become a part of one's cognitive character. An agent's cognitive character is a collection of her dispositional and stable, natural and acquired cognitive abilities. They can also include: 'acquired skills of perception and acquired methods of inquiry, including those involving highly specialised training or even advanced technology' (Greco 1999, 287).<sup>20</sup> The definition tells us that a target process may be innate or acquired, neural or non-neural, but in order to integrate it has to be a *reliable and stable disposition* of the agent's cognitive character (Sosa 1991; Greco 1999).

What are stable dispositions of a cognitive character? I adopt new epistemological habits from time to time, but, then, I often lose them. These are not the kinds of dispositions that Sosa and Greco are referring to. Sosa proposes that cognitive abilities are virtues in the sense that vision is the virtue of eyes and hearing is the virtue of ears (Sosa 1991). These are not features adopted on a whim. They are stable dispositions of these organs. Similarly, our cognitive character has stable dispositions that generate reliable beliefs, our cognitive abilities.

One way to form cognitive abilities is to acquire a 'perspective on reliability' of one's process. But, Greco writes, 'it would be a mistake, however, to think that such a perspective is always needed to make reliability and success internal to agency' (2010, 142). What is necessary and sufficient for epistemic integration is that the new process forms cooperative interactions with other

---

<sup>20</sup>For comparison, see a more restricted understanding of cognitive character in Zagzebski (1996).

processes in the agent's cognitive character. Greco states that epistemic integration is basically the function of 'cooperation and interaction' (Greco 2003, 474; 2010, 152). When a reliable belief-producing process becomes a part of one's cognitive character, it develops interconnectedness with other processes in one's cognitive character, on the basis of its output. The output from the target process is assimilated by another to form more output, that is further assimilated by other abilities. This is how processes in our cognitive characters cooperate and work together and how new processes become stable dispositions of our cognitive characters. So, for a weaker understanding of integration, it suffices that we do not have any reasons against our process's reliability. It is necessary, though, that we are conscientious in the sense that we genuinely care (even if not always reflectively) about getting beliefs right (Greco 2010).

How is it that the function of cooperation and interaction can explain how an agent can take responsibility for her process? The idea is that the cooperatively functioning cognitive system can monitor the reliability of its processes. If any process, in a cooperative system, malfunctions, the cognitive (or epistemic) loop will come to a halt and send alert signals to the agent. This background monitoring is best described in these words:

'... if one's belief-forming process cooperatively interacts with other aspects of one's cognitive system, then it can be continuously monitored in the background such that if there is something wrong with it, then the agent will be able to notice this... Otherwise—if the agent has no negative beliefs about his/her belief-forming process—he/she can be subjectively justified in employing the relevant process by default, even if he/she has absolutely no positive beliefs as to whether or why it might be reliable' (Carter, Clark, and Palermos 2018, 337).

This is why the agent does not need access to the reliability of her process. She is, indirectly, tracking the reliability of her integrated processes because of the interconnectedness of her cognitive character.

The background monitoring of (the reliabilities of) integrated processes puts the agent in a state where she can become aware of any counterevidence to

the reliability of her process. Palermos, describes this as something only humans can do, i.e. become aware of the problems with our processes on the basis of the integratedness of our systems. He writes,

‘This sense of epistemically adequate—yet unreflective—cognitive responsibility can only be achieved by agents like us, whose intellectual capacities are appropriately interconnected such that in cases where there is something wrong with the way we form our beliefs or with the beliefs themselves, we will be able to notice this and respond appropriately’ (Palermos 2014b, 1934).

This proposition is a major point of investigation for my thesis but one that I fully develop in the next few chapters. In short, on this claim, Andy Clark argues that the agent does not always gain such counterfactual sensitivity. It is possible that sometimes only her cognitive system or a part of it becomes counterfactually sensitive in this way and can monitor the reliability of the process (Clark 2015).<sup>21</sup> So, it seems we have a problem here. Clearly put, Clark states that epistemic integration only requires that the interconnectedness of a cognitive system can monitor the reliability of the process. However, Greco and Palermos assert that integration allows *epistemic agents* to become aware of their malfunctioning integrated processes. The problem is whether the agent’s position to become aware of any problems with her integrated process is a part of the ECI account or should it be discarded? I will not get further into this debate here, since all these issues will be systematically unpacked in forthcoming chapters.

An important feature of epistemic integration is that it can be present in different degrees since it is essentially the function of interaction and cooperation between the target process and one’s cognitive character. If a process is not employed frequently over a long period of time, then it will only produce output that is few and far between (Greco 2010). This output may not be enough to develop sufficient cooperation with other processes in the cognitive system. Furthermore, the degree of integration also depends on the

---

<sup>21</sup>I will discuss this critique in more detail, and place it in the overall debate, in the next chapter.

kind of process in question. Some processes, for example, a visual aid, is likely to be used more frequently over a period of time, as compared to a weather belief-forming faculty. Hence, the former will most probably integrate in a shorter period of time. So, all processes do not epistemically integrate into our agencies at the same rate or with the same degree of integrations. How do we determine, then, if the target process has integrated? The answer is that the degree of integratedness that puts the agent in the position to become aware of any problem with the reliability of her process is the degree that is sufficient for epistemic integration. In other words, when the epistemic agent has developed responsiveness to the reliability of her process, it means she has integrated (via weak ECI) the said process into her cognitive system.

Greco discusses these two aspects, side by side, in the following words:

'one aspect of cognitive integration concerns the range of outputs – if the products of a disposition are few and far between, and if they have little relation to other beliefs in the system, then the disposition is less well integrated on that account. Another aspect of cognitive integration is sensitivity to counterevidence, or to defeating evidence' (2010, 152).

He does not link the first aspect with the second one, which is what I would like to do here. I am emphasising that the second aspect informs the first one, to specify the degree of interconnectedness that is sufficient for integration. To put it another way, Greco only lists these as features of cognitive integration. That is, that integration of different processes may require different degrees of integration and that once a process is integrated, the agent is in the position to become aware of the problem if it malfunctions. What I would like to bring attention to here is that the degree of integration can be determined in terms of the position that the agent comes to be in. The degree of epistemic integration is sufficient when the agent can become aware of counterevidence to the reliability of her process.

Now, that I have explored Greco's structure of epistemic integration that informs Pritchard's weak ability intuition and the weak ECI concept, I will enlist the weak ECI conditions to evaluate a univocal view in the next section.

Palermos claims 'the only requirement is that the process be integrated into the agent's cognitive character, by engaging in cooperative interaction with the rest of the agent's cognitive system' (2014b, 1944). But, this formulation does not tell us that degree of cooperation that is sufficient for weak ECI. As an alternative, I propose a more thorough list of conditions:

- (1) The agent should be conscientious (i.e. someone who is genuinely interested in forming true beliefs).
- (2) The belief-forming process should be reliable.
- (3) Other processes and the new belief-forming process should form cooperative and interactive complementary connections (to a significant degree) that drive cognition forward
- (4) Lastly, the degree of integration will be sufficient for weak ECI when the agent is in the position to become aware of any counterevidence to the reliability of the said process.

Some clarifications are due here. I have tried best to not merely list the weak ECI conditions in terms of characteristics a belief-forming process should possess. It is important to include, in the weak ECI conditions, how an agent or her cognitive agency ought to approach the said process. Also, (1) is not an awareness condition. It does not require that the agent is aware of anything, only that the agent is interested in forming true beliefs. And, I have argued for (4) before but I should clarify that (4) is not an awareness condition either. This means that the awareness needs to be counterfactual instead of being actual. So, the interactive cognitive system should allow one to become aware of any counterevidence to the reliability of the belief. So, awareness or any active and conscious involvement from the agent is not a part of weak ECI conditions.<sup>22</sup>

### Example of a univocal view

Now that we have a set of ECI conditions as well as MCI requirements in the shape of the glue and trust criteria, we can compare the two kinds of

---

<sup>22</sup>To be clear, my account is not different from Pritchard's weak ECI, as discussed in (2010). What I am trying to provide here is merely an elaboration of the conditions, why they are added, and what purpose they serve in an epistemic integration account.

integration and observe what a univocal view looks like. Recall that the glue and trust requirements are supposed to provide an MCI account or a commonsensical view of what cognition looks like. However, these criteria are essentially a list of characteristics for the belief-forming processes to fulfil, e.g. accessibility, availability, reliability, trustworthy information, which is why they sound more like weak epistemic integration requirements. In what follows, I describe Palermos' comparison of the weak ECI structure to Clark's glue and trust requirements. After comparing these two integrations in terms of their conditions, I move on to describing and critically evaluating a univocal view.<sup>23</sup>

Recall the first glue and trust requirement that demands that the agent's resource should be readily available and typically invoked by the agent's cognitive agency. Palermos argues that this requirement is the same as the weak ECI demand that one's belief-forming process ought to be one's stable disposition. This is because our stable dispositions are always available to us and we employ them in a specific manner (Palermos 2011, 2014b).

Furthermore, he points out that the second glue and trust condition that the information from the resource should be trusted by the agency without any scrutiny is to demand that one's process should be reliable. The reliability of the process has a relationship with how the agency trusts the process. Our agencies rely trustfully on processes that give them reliable beliefs. So when the ECI requirement for a reliable process is met, the MCI requirement for the agency's trust in the information from the resource is also satisfied (Palermos 2011, 2014b).

The third glue and trust condition demands that the information from the resource should be accessible to the agent when necessary. Information from processes that are integrated into our cognitive characters is easily accessible to us. Palermos mentions that this glue and trust condition is the same as the demand that the target process should be integrated into our cognitive system, on the basis of cooperation and interaction.

Palermos' argument is that weak ECI's demands, i.e. stable dispositionality and reliability of the process and cooperative interactions between the target

---

<sup>23</sup>A more detailed version and defence of the univocal view will be introduced in the next chapter. It will also include a revised understanding of MCI requirements.

mechanism and other faculties, are the same as the glue and trust demands. Therefore,

‘the same features of a process that epistemologists deem important in order for a process to be knowledge-conducive are required by a common-sense functionalist understanding of cognition in order for a process to count as part of one’s mind. This is a promising observation’ (Palermos 2014b, 1945).<sup>24</sup>

The similarities between the requirements for weak ECI and the glue and trust conditions support a symmetric univocal view. That is, the requirements that fulfil the criteria for MCI, also meet weak ECI conditions, and *vice versa*. Even in the Otto case, we found that both integrations existed side-by-side. As Otto’s employment of his notebook meets the glue and trust conditions, it is also a cognitive ability for him that can produce knowledge for him.

However, all cognitive processes are obviously not epistemic. This is clear by the difference between the nature of belief and knowledge; knowledge is factive and involves an ‘epistemic dimension’ (Carter and Kallestrup 2019, 7). So, whatever describes epistemic processes has to have more demanding requirements than the conditions that describe cognition. The MCI conditions that indicate cognitive processes, which all epistemic processes are, should be a subset of ECI requirements. Accordingly, the relationship between the two integrations will be asymmetric, in the sense that where there is ECI there will be MCI present, but the presence of MCI may not entail ECI.<sup>25</sup> The other side of this story is that if a process meets the MCI conditions, it has met part of the ECI conditions. In other words, the MCI are a subset of the ECI conditions. This makes more sense mainly because all cognitive processes are not epistemic, but all epistemic processes are cognitive.

---

<sup>24</sup>A common sense functionalist understanding means to discuss the functional role of a mechanism or to interpret cognition in terms of functional roles that are accessible in terms of common-sense. For instance, the glue and trust conditions are a common-sense understanding of what cognition is like.

<sup>25</sup>Aizawa presents the worry of the symmetric entailment when he discusses Palermos’ univocal view in (2018).

The following example is going to show that there are serious problems with the symmetric entailment univocal view. Consider Otto\* who also has Alzheimer's.<sup>26</sup> On top of that, Otto\* has been in an accident. He was brought to the operation table after the accident, where his doctors implanted a device in his brain. This device works the same way Otto's notebook and Inga's biological memory. Otto\* is able to store his beliefs in this implant and retrieve them when they are required. This device will make his life easy and he will no longer have to carry a notebook around. The procedure goes well and Otto\* recovers from his accident with a useful implant in his brain. The only problem is that his doctors forget to tell him about this implant. Hence, Otto\* has no idea that he has such a resource or that this resource is reliable.<sup>27</sup> A few days after the surgery, Otto\* is hanging with Otto and Inga where he hears about the exhibition at MoMA. He uses his implant to retrieve MoMA's address and makes his way there.

It is evident that Otto\*'s device fulfils the glue and trust requirements. The implant is attached to his brain, which means it is readily available to him. He can typically invoke it since he is able to retrieve information about everyday details, like the location of MoMA. The information from the implant is most likely automatically endorsed and easily accessible. He can automatically endorse the information about MoMA from his implant because it is coming to him just the way our memories make our beliefs available to us. And, this information will be accessible easily to Otto\* because he carries the implant around everywhere he goes.

If the symmetric relationship between the two integrations is correct, then Otto\*'s resource should also fulfil the weak ECI requirements. This is to say, that Otto's implant should be a cognitive ability for him and should be able to give him knowledge of MoMA's whereabouts. However, concerned epistemologists, like Pritchard and Palermos, are convinced that Otto\* does not have knowledge in this case, at least not right away. He has to use his implant for a long time, and frequently, in order to epistemically integrate it into

---

<sup>26</sup>Adapted from Carter and Kallestrup's Memoraid case (2019).

<sup>27</sup>Otto\* is in the same position as Norman the Clairvoyant. He has a reliable process but he has no perspective or sensitivity to the reliability of his process.

his cognitive character. The implant should form cooperative links with other processes in his cognitive character and the density of integration should be such that Otto\* can become aware in case of a problem. In this case, when Otto\* forms the belief about MoMA's whereabouts he is not sensitive to the reliability of his process. His implant could be malfunctioning and Otto\* would have no idea. If he uses the implant for a couple of months or years, he will be able to develop counterfactual sensitivity to his implant's reliability and use it as a cognitive ability. Hence, while the glue and trust requirements can be met right away, the weak ECI conditions require time and use.

The Otto\* case shows that metaphysical integration may be fulfilled independent of the epistemic integration, which is why we see that Otto\* meets the glue and trust but fails to form a cognitive ability. And, the reason for this is that weak ECI takes longer to develop. But this is not true for all kinds of epistemic integration. Next, I will reflect on a case in which epistemic integration is achieved independent of metaphysical cognitive integration.

Otto\*\*'s doctors do not forget to tell him about the surgery he went through. They tell him about his new implant, how it works the same way our biological memory works, and that it is reliable. Since, he has access to the reliability of his implant, he is able to employ it carefully and, therefore, the implant immediately epistemically integrates into his cognitive character. He can responsibly employ his implant to form true beliefs that can be knowledge, right away. Although, now Otto\*\* has failed to meet the metaphysical cognitive integration requirement, i.e. the glue and trust. The second glue and trust requirement demands that the information from one's resource should be endorsed by the agent without any scrutiny. But, Otto\*\* has scrutinised his implant by gaining access to its reliability and employing it carefully. This means that the implant can generate knowledge for our agent but it is not an extended cognitive process of his cognitive system. If at some later time, Otto\*\* stops caring about the reliability and begins to employ his implant seamlessly, we can say that it fulfils the extension criteria then.

These modified Otto cases indicate that a symmetric entailment probably does not exist between the two integration accounts. That is, MCI and weak ECI do not always have to be present together. A process may integrate in

MCI terms even when it does not integrate in terms of weak ECI, and weak ECI may be present when MCI is not. Although, it is worth noting that we have described MCI only in terms of the glue and trust conditions. A different understanding of metaphysical cognitive integration may yield different results in this matter. After all, in the Otto\*\* case, the agent fails to meet the glue and trust conditions but meets the ECI requirement. This obviously cannot mean that his process is epistemic but not cognitive. Most likely, Otto\*\*'s process is cognitive but cannot be described as such in terms of the glue and trust conditions. What it can mean is that the glue and trust requirements are not very good at identifying cognition (or do not necessarily and sufficiently meet MCI) and that we need to come up with better criteria for MCI.<sup>28</sup> This also means that when we have a better understanding of MCI, we can reevaluate the symmetric and asymmetric entailments.

Furthermore, the Otto\*\* case suggests that there may not even be an asymmetric entailment between the two integrations, since epistemic integration can be achieved independent of metaphysical cognitive integration. However, again, this is the absurd claim that a said process is epistemic but it is not cognitive. Evidently, we ought to look for a better depiction of what cognition is like and what MCI conditions are required for it. Without such an account we will not be able to evaluate the two univocal views.

Another reason that Otto\*\* does not meet the glue and trust requirement may be that perhaps these requirements are only an indicator of *extended* cognition and not cognition *tout court*. So, the glue and trust can explain what cognition looks like when we are looking for it outside our head but it does not explain all cases of cognition. After all, we do not describe our innate, neural processes as cognitive on the basis of their availability, accessibility, etc. It is how these processes process information that makes them cognitive.<sup>29</sup> This can explain why Otto\*\*'s process is epistemic (via the strong agency route), and obviously a cognitive process that meets MCI, but it is not an extended

---

<sup>28</sup>A better account of MCI is presented by Palermos as the 3+1 thesis. This is a stronger version of his univocal view and I will deal with it in the next chapter.

<sup>29</sup>Aizawa makes the argument that trusting and being glued to an information resource should not be used to describe what makes our capacities cognitive (2018).

cognitive process. This argument also prompts us to look for a more adequate understanding of metaphysical cognitive integration that can describe cognition more appropriately.

## Conclusion

The function of this chapter was to introduce concepts and terms that are required for the main debate on subpersonal epistemology (that will be discussed more thoroughly in the next chapters). This chapter acquaints the reader with the two integration accounts: metaphysical and epistemic. The former gives insight into what is required for cognitive extension, and the latter describes what makes our cognitive processes epistemic.

I suggest, in line with Pritchard, Palermos, and Carter, that the univocal view is a good explanation of the relationship between these two integrations. But there are two univocal views to choose from. The symmetric entailment view proposes that the conditions for these two integrations fully overlap. The asymmetric entailment, on the other hand, suggests that there is some overlap in the conditions for these two integrations.

I have presented a symmetric univocal view and presented arguments against such a view. I have also discussed cases that describe how it is possible to achieve metaphysical integration without epistemic integration, and a form of epistemic integration without metaphysical integration. Although, I suggest that we look at metaphysical integration in terms of more than just the glue and trust conditions and the functional parity principle. I will discuss in the next chapter, another theory that describes metaphysical integration more appropriately for our purpose (which is to compare the two integrations and evaluate the univocal view).

This chapter also discusses my reasons for leaning towards an asymmetric entailment between the two integrations. And, it provides an overview of the debate that will continue to the end of my thesis, namely that the weak epistemic integration account is too active to be compatible with extended cognition. In my understanding, it is appropriately active. I have brought attention to the fact that integration comes in degrees and while both these

integrations incorporate interactions amongst processes, we need to specify the degree of interaction that will be sufficient for epistemic integration. Simply put, epistemic integration requires that the agent develop some sensitivity to the reliability of her integrating process.

## Chapter 2 The univocal views and the epistemic hygiene dilemma

The aim of this chapter is to introduce other aspects of the debate about the relationship between metaphysical cognitive integration that marks cognition and weak epistemic integration. In the previous chapter, I suggested the need for a better understanding of what cognition is like and mentioned that the glue and trust conditions may not be the best way to approach metaphysical cognitive integration. I begin this chapter by considering some serious problems facing the glue and trust criteria. Opponents of the extended cognition thesis have proposed that the thesis confuses coupling of an external mechanism to our cognitive process with the constitution of the said mechanism in our cognitive process. And, also, that the glue and trust conditions fail to demarcate the boundary of the extended cognitive process. As a result, it seems like all gadgets and technologies that we trust and have constant access to can extend our cognition. Proponents of the extended cognition thesis introduce complementarity (Menary 2007) style arguments in order to combat the worries that arise from the analysis of the glue and trust conditions.

I consider Palermos' complementarity style argument in detail and explain metaphysical cognitive integration in terms of it. According to this view, the on-going reciprocal exchange between an external process and processes in one's cognitive system allows the external process to couple with one's cognitive system in a way that the coupling is a kind of constitution. The two systems come together as a single unit to perform functions and affect other processes in the cognitive system.

With this revised understanding of metaphysical cognitive integration (MCI henceforth), I revisit the relationship between MCI and epistemic cognitive integration (ECI). In the previous chapter, I discussed an argument from Palermos that indicates a symmetric relationship between the two integration accounts. The symmetric account shows that conditions for MCI and ECI fully overlap. In this chapter, I discuss another symmetric integration view in terms of the revised MCI account and a convincing asymmetric view as well. An

asymmetric view points out that either the MCI conditions are a subset of the ECI conditions, or at least that there is some overlap between the conditions required for these integrations.

Further, I explain Andy Clark's *epistemic hygiene dilemma*, which suggests a major, irreconcilable difference between the two integrations— that is, MCI is an entirely passive affair, whereas epistemic integration, even in its weakest form (i.e. Pritchard's weak ECI), requires active agential participation. In light of this dilemma, Clark puts forward his *subpersonal ECI* account (i.e. integration explained thoroughly in terms of parts of the agent) in lieu of the weak ECI, which I describe is a symmetric view. I present problems with Clark's proposal on account of it being a symmetric view and for proposing a thoroughly subpersonal (i.e. in terms of part of a person, instead of a whole person) concept of how an agent can responsibly acquire knowledge. Lastly, I discuss how the dilemma does not harm the asymmetric view, a path that we can still take if we acknowledge that the weak ECI account is a substantially passive form of epistemic integration.

## Problems with the glue and trust conditions

The glue and trust conditions are far from perfect in describing the MCI conditions. In this section, I discuss problems associated with understanding MCI in terms of the glue and trust conditions alone. This will pave way for introducing a better concept for metaphysical cognitive integration, which explains how a new process becomes our cognitive process.

Consider Telo (Carter and Kallestrup 2019, 2018). Telo's brain functions typically. Like most other people, he also relies on external resources to enhance some of his cognitive processes. Telo has a contact list on his phone that stores all the important numbers for him. Phone numbers of his parents, friends, office colleagues, etc. are all saved in his contact list. Telo's phone is on him all the time. He invokes it typically and trusts the phone numbers he retrieves from his digital phonebook. The information from his phone is easily accessible to him as well. When Telo wants to call someone, he can simply retrieve their number from his phone; he does not have to memorise all the numbers.

Evidently, Telo meets the glue and trust conditions because he trusts the information that he looks up on his phonebook, and he carries his phone with him at all times. This idea proposes that we should consider Telo's phonebook an extension of his cognition, like we do Otto's (from the previous chapter) notebook. Another claim that can be made is that Telo's mental states are realised partially by his cognitive system and partially by the stored information in his phonebook. Furthermore, some of these mental states, or beliefs, of Telo's may be knowledge for him. This gives birth to the idea that Telo perhaps *knows* all these phone numbers. Along the same lines, we can imagine another agent whose cognition can extend into Google, because he carries the application around in his phone everywhere he goes and trusts this information. It can be said that he, too, knows all the information that is available on Google because he carries his phone with him all the time and trusts the information that he can look up on his phone's Google application. The point is that if the glue and trust criteria are the only measure for cognition then it seems like cognition extends into the environment in a dangerously unrestrained manner. And, with it comes the possibility that knowledge also extends in an unrestrained way. Rob Rupert calls the unrestrained cognitive extension the cognitive bloat worry (Rupert 2004), and Mark Sprevak suggests that if the cognitive is the mark of the mental, then this problem further leads to the 'rampant expansion of the mind' (Sprevak 2009, 503).<sup>30</sup>

There is another worry that hints at the same conclusion, that the glue and trust conditions do not adequately explain cognitive extension. Adams and Aizawa (2008) argue that the extended cognition supporters confuse coupling with constitution and that coupling entails causal dependency, not constitution. This is called the coupling constitution fallacy. Adams and Aizawa write, 'it simply does not follow from the fact that process X is in some way causally connected to a cognitive process that X is thereby part of that cognitive process' (2008, 91).<sup>31</sup> The argument is that, in extended cognition literature, the terms

---

<sup>30</sup>Farkas (2012) discusses a similar problem in terms of audiotapes of a complete history of Europe.

<sup>31</sup>For a more detailed discussion of the debate on coupling-constitution fallacy, see Ross and Ladyman (2010).

integration, coupling, manipulation, etc. are used to describe how an external mechanism causally affects a cognitive process, but the conclusion drawn is that the causal dependency makes the mechanism a part of the cognitive process in some way. But, simply, causally affecting a cognitive process is not the same as constituting a cognitive process. For example, just because Otto carries his notebook around or uses it in a way that causally affects his cognitive process does not mean that the mechanism of this external vehicle can become a part of his cognitive process.

The glue and trust requirements are susceptible to this argument against extended cognition, just like they are to the cognitive bloat worry. This is because they only describe an external resource in terms of its availability, accessibility, and trustworthiness to one's agency. They present no reason towards the view that the mechanism of the external resource merge with our cognitive process and become a new, evolved cognitive process. These worries urge proponents of cognitive extension to bring forward a new kind of motivator for cognitive extension, one that explains how an external mechanism can merge into our cognitive process. I will describe these style of arguments, shortly, in more detail.

There is one last issue with the glue and trust criteria that I would like to examine in this section. As I mentioned in the previous chapter, the glue and trust requirements are expressed in terms of adjectives— accessible, available, endorsable, etc. The adjective describe the qualities that a process should possess in order to qualify as a cognitive process. It can be said that the glue and trust conditions are like a *selection criteria* for processes. They do not sketch a picture of how cognitive integration actually occurs, or how our cognitive agency employs an external mechanism in a way that the latter can become a part of our process. They merely express the attributes that the agent ought to look for in a process. In this way, the glue and trust conditions suggest that the role of cognitive agency in employing a process is an active and scrutinising role. To put it another way, the language used to describe the glue and trust conditions makes it sound like the agency plays an active role in recruiting its cognitive processes. However, consider the second glue and trust condition. It requires 'That any information thus retrieved be more or

less automatically endorsed. It should not usually be subject to critical scrutiny. It should be deemed about as trustworthy as something retrieved clearly from biological memory' (Clark 2010b, 47). This criterion clearly demands that the role of agency should be passive and unscrutinising. Why, then, are the glue and trust conditions formulated in a way that makes them sound like a selection criteria? In my understanding, the glue and trust requirements seem confusing and somewhat self-contradictory in this way.<sup>32</sup>

## Continuous reciprocal causation

So, the glue and trust conditions struggle to demarcate the extended cognitive process, and they do not describe how an external mechanism can be included in one's cognitive process. Both these tasks can be taken care of with a different style of arguments, complementarity-style arguments, that motivate extended cognition. One such argument is presented by Palermos<sup>33</sup>, and it presents an improved description of metaphysical cognitive integration. I am going to elucidate this new MCI account in this section so a better comparison of the two integrations can be made.

Palermos' improved MCI account is based on the mathematical theory called, Dynamic Systems Theory (DST) (Thompson and Varela 2001; Palermos 2011, 2014b, 2016, 2017).<sup>34</sup> Accordingly, two systems, S and O, can temporally evolve into a third, combined system E, if they come together in a continuous

---

<sup>32</sup>What I am suggesting in this paragraph will make more sense once Clark's critique of the weak epistemic integration account has been discussed. Simply put, Clark finds the weak ECI approach far too active to be in line with cognitive extension. I am merely pointing out that Clark's glue and trust conditions are described in a way that makes them sound in line with the weak epistemic integration account (i.e. glue and trust conditions are as active as conditions for weak ECI are). That is, the glue and trust conditions are not as passive as he wants the weak epistemic integration account to be. I will get into the debate about how active and how passive these conditions are, later in the chapter.

<sup>33</sup>It is important to note here that the main themes of this argument were first presented by Clark (2008), but their current shape in literature can be credited to Palermos.

<sup>34</sup>For my purposes, I do need to describe the theoretical basis for Dynamic Systems Theory. I am interested in using the explanation that Palermos relies on to describe CRC. That is, the evolution of two systems that occurs on the basis of a high intensity of interaction between the two systems.

and reciprocal exchange of information. On the extended cognition view, an external mechanism can have *continuous reciprocal causation* (CRC) with a process (or processes) that belongs to the agent's cognitive system, and this on-going and non-linear exchange can allow the former mechanism to fuse with the latter cognitive process (Clark 2008). Clark describes CRC as, "when some system S is both continuously affecting and simultaneously being affected by activity in some other system O" (2008, 24).<sup>35</sup> The significant feature of CRC is that it allows the cognitive loop to be driven forward, which means that the new mechanism becomes a part of the cognitive process and allows cognition to move forward. This is possible because the target process achieves the necessary degree of interaction with other (relevant) processes of the agent's cognitive system that allow them to work and produce output together as one unit.<sup>36</sup> All in all, DST motivates extended cognition in a very different way from glue and trust conditions. It explains (1) how two processes should affect each other (in terms of CRC) and (2) that this type of causation entails constitution.

Philosophers of mind (Chemero 2009; Tollefsen and Dale 2012) concur that the complementarity theory provides a better understanding of what a cognitive process is and how we can identify one. A part of one's cognitive system is attached to the whole on the basis of on-going mutual interactions

---

<sup>35</sup>Clark suggests that some inputs only affect the cognitive process to allow the processing to move forward, whereas other inputs become a part of the cognitive process and drive the process forward by producing output that can convert back into input. He writes, 'What these simple examples show is that (as we might expect) coupling alone is not enough. Sometimes, all coupling does is provide a channel allowing externally originating inputs to drive cognitive processing along. But in a wide range of the most interesting cases there is a crucially important complication. These are the cases where we confront a recognizably cognitive process, running in some agent, that creates outputs (speech, gesture, expressive movements, written words) that, recycled as inputs, drive the cognitive process along. In such cases, any intuitive ban on counting inputs as parts of mechanisms seems wrong' (2007, 185). On this note, Palermos states that Clark finds CRC only sufficient for cognition, whereas he would describe it as necessary as well (2011).

<sup>36</sup>I describe CRC in terms of the neuroscientific concept of deep functional integration in the next chapter. The discussion from Greenwood (2013) emphasises on the interaction between two systems in a way that produces a common output. This is in line with Palermos' description of how a common product from the target process and the agent's other processes can become input and produce further output and therefore drive cognition forward.

between the two. Interaction and cooperation, the element of epistemic integration, is basically at the centre of the complementarity-style arguments for extended cognition as well. As it will become clear from discussion in this chapter, CRC alone is not sufficient to understand epistemic integration. Furthermore, I would like to clarify here that I am not defending CRC as the best argument for metaphysical epistemic integration or cognitive extension, even though it has been considered such in Palermos' work. I only intend to describe the full picture of Palermos' work so I can present a relevant argument on the subject of univocal views. Later in this thesis (Chapter 6), I will present ideas that cast doubt on the assumption that CRC does a thorough job in defining extension.

Another feature of an extended cognitive process is that the external mechanism is joined via CRC, adds to the behaviour of the cognitive system, and it is, therefore, a part of it. There is an observable decline in 'behavioural competence' when the new part is removed or altered (Palermos 2011, 751). Consider the Otto example once again. Otto's cognitive system and the mechanism of his notebook can merge together in virtue of CRC and, therefore, evolve into a new process: Otto's cognitive process of manipulating his notebook's mechanism. Now, if Otto's notebook is taken away from him for any reason or if he loses it, it will affect his behaviour. For one, Otto will most probably not be able to reach the exhibition at MoMA because he will be unable to retrieve the address he desires. Also, he will fail to perform many other tasks for which he relies on the notebook.

Adding the CRC requirement to the understanding of MCI takes care of the cognitive bloat worry and the coupling-constitution fallacy. Let's take another look at the Telo case in terms of how we understand metaphysical integration now (with CRC). What we are trying to find out is if Telo's phonebook is integrated in his cognitive system via the CRC requirement. Telo's cognitive system does not have a continuous exchange of information with his phonebook. He does not employ his phonebook as consistently as Otto employs his notebook to store and retrieve information. Telo only has to use his phonebook when he needs to call a friend or acquaintance. You might think he carries his phone with him the same way Otto carries his notebook with him.

But, Telo's phonebook is with him because he carries his phone with him all the time. He does not keep his phone on him only for the phonebook; he uses it for myriad other functions. Furthermore, Telo's reliance on his phonebook is not only not continuous, it is also not reciprocal. The exchange of information between Telo's phonebook and his neural cognitive processes is not the same as Otto's notebook and some of his neural processes. For instance, Otto automatically writes the things he hears and looks them up when he needs to find out something that he has likely written down in the past. Telo does not engage in a continuous non-linear interaction with his phonebook. If Telo loses his phonebook he will need to contact another friend perhaps or refer to the yellow pages to find a phone number. He will basically employ another tool to perform the same function. If Otto loses his notebook it will be a much bigger loss, since the notebook carries the information that Otto requires to get through his everyday tasks. Otto's notebook, in this sense, is a part of his cognitive system, while Telo's phonebook is a tool for him. Also, Otto's other cognitive processes interact with the information in his notebook in an automatic way. For instance, he seamlessly approaches his notebook to recall where he lives in order to get home safely. Telo, on the other hand, reflectively interacts with his phonebook to look up a phone number that he needs. All in all, Otto's notebook enhances his behaviour in ways that Telo's phonebook does not. The latter is only an instrument for Telo.

By understanding MCI in terms of CRC we can draw a clear distinction between the Otto and the Telo case. CRC distinguishes why the way Otto uses his notebook constitutes his cognitive process (and drives the evolved cognitive process forward) but the way Telo sometimes employs his phonebook does not. Evidently, if we understand extended cognition in terms of continuous reciprocal causation then we will note that it does not confuse coupling with constitution, since it provides an explanation for how some kinds of causations can entail constitution. Telo's phonebook does not, therefore, help realise his beliefs about his friends' phone numbers. His process of engaging his phonebook is not extension of cognition.

## More univocal views

Palermos does not distinguish the two kinds of integration, and so we do not find the terms ECI and MCI in his work. He uses the word 'integration' only for epistemic integration, and his works compare epistemic integration with the conditions for extended cognition. But, since my aim is to assess the univocal views at hand, I will adhere to the ECI and MCI classification. The task at hand, in this section, is to evaluate the symmetric univocal view. This symmetric view is also from Palermos and is based on the revised conception of MCI.

The CRC requirement does not only solve the cognitive bloat and the coupling-constitution problems, but Palermos also uses it to provide two univocal views— a symmetric and an asymmetric one. In order to assess these univocal views, I should outline Palermos' description of the conditions required for epistemic integration and the ones that describe cognitive extension. Firstly, his concept of epistemic cognitive integration is that 'a belief-forming process is knowledge-conducive if it is a reliable belief-forming disposition properly integrated within the agent's cognitive character' (Palermos 2011, 756). The keywords here are *reliable*, *disposition*, and *integrated*. In terms of availability, accessibility and endorsability, the glue and trust conditions also describe a *reliable* process that is a *normal and stable disposition* of one's cognitive character.<sup>37</sup> So, the three glue and trust criteria take care of reliability and dispositionality. Integration of a new process occurs on the basis of its cooperation with other faculties in one's cognitive character. As described in the previous section, the new process forms output that is assimilated by other processes to form more output. This phenomenon is the same as the presence of CRC between a new process and other processes of the system. Palermos writes that 'in order for a belief-forming process to be appropriately integrated within one's cognitive character, the phenomenon of CRC must be manifested between the target process and one's organismic cognitive faculties' (Palermos 2011, 758). Notice, that CRC is described in terms of an on-going reciprocal exchange that allows outputs to be recycled as inputs and allows the cognitive loop to

---

<sup>37</sup>I discuss this in detail in the previous chapter, where I describe Palermos' preliminary symmetric univocal view. According to this view, the glue and trust conditions are the same as Greco's epistemic integration requirements.

be driven forward. And, this is the same way that Greco explains epistemic cognitive (weak ECI) integration. Hence, Palermos' ECI account is a 3+1 (three glue and trust conditions plus one CRC requirement) model.

Secondly, Palermos explains extended cognition (MCI) also in terms of the 3+1 criteria. He writes:

Consequently, then, apart from the 'glue and trust' criteria, in order to account for the constitutive status of external elements within one's overall cognitive mechanism, we also need the phenomenon of continuous reciprocal causation between the outer and the inner parts to take place. These 3+1 criteria seem to jointly ensure the integration of the external artifacts within one's overall cognitive mechanism, thereby overcoming the 'coupling-constitution' fallacy and the 'cognitive bloat' worry. (Palermos 2011, 756)

Hence, his claim is that the same conditions that describe cognitive extension, also suggest how one's reliable belief-forming mechanism develops into a knowledge-conducive cognitive ability. This is a symmetric univocal view.

Aizawa (2018) points out the obvious problem with this, or any, symmetric univocal view. Such a view claims that the same conditions that explain why a process is cognitive can explain why it is epistemic. As discussed in the previous chapter, it leaves no room for cognitive processes that are not epistemic. According to this description, all cognitive processes ought to be epistemic. However, all cognitive processes are not epistemic; they do not aim towards getting beliefs right. In fact, some cognition, like finding our way to the kitchen, does not require any epistemic processes. We only need to navigate towards a location. I will discuss this in more detail when I discuss Clark's univocal view in the following sections.<sup>38</sup>

In Palermos' defence, it may be that the glue and trust requirements are only practical preconditions for MCI. They are neither necessary nor sufficient for extended cognition. Palermos implies this when he suggests that 'normality

---

<sup>38</sup>Joelle Proust (2013) distinguishes epistemic agency and mental agency. She writes about how we cognise to control our emotions and to change our mood and that is not the same as agency being manifested to form knowledge (epistemic agency).

and dispositionality,' which he equates with the glue and trust conditions, are only 'practical preconditions' to integration. He writes, 'even though normality and dispositionality will, in most cases, be practical preconditions, they are neither necessary nor sufficient for a process to count as integrated into the agent's cognitive character' (Palermos 2014b, 1943). And, he clearly states that CRC can be both necessary and sufficient for extended cognition (2011, 754). Therefore, Palermos' univocal view is most likely an asymmetric one, since MCI is brought about by CRC and for ECI the 3+1 (glue and trust + CRC) phenomenon should be present. The asymmetric view will settle Aizawa's worry.

However, there is another strong gesture towards a symmetric view in Palermos' work. He does not find the glue and trust necessary or sufficient for epistemic integration either. Even epistemic integration, for Palermos, can solely be understood in terms of cooperative interactions, i.e. the presence of CRC. Once again, it seems that MCI and ECI can both be understood, necessarily and sufficiently, in terms of CRC. Consider the following continuation of the previous quote, from Palermos, that distinctly suggest that CRC or cooperative interactions are both necessary and sufficient for extended cognition as well as epistemic integration:

...even though normality and dispositionality will, in most cases, be practical preconditions, they are neither necessary nor sufficient for a process to count as integrated into the agent's cognitive character. Instead, the only requirement is that the process be integrated into the agent's cognitive character, by engaging in cooperative interaction with the rest of the agent's cognitive system. Accordingly, no matter what the practical preconditions for this interactive process to be achieved are, once it is in place, it will guarantee both that the relevant belief-forming process is a cognitive process, and that it is indeed part of the agent's cognitive character such that he can be conscientious in employing it (Palermos 2014b, 1943).

All in all, Palermos (2014b) argues for a complete overlap (i.e. a symmetric entailment between the two integrations) between the conditions

of epistemic integration and cognitive extension and a partial overlap (i.e. an asymmetric entailment).

The important thing to take from this section is that there is, in fact, significant resemblance between the virtue epistemology account of the weak ability intuition and the extended cognition hypothesis. Generally, they both provide an understanding of how an external mechanism can be incorporated into one's cognitive system, which means that there can be a complete overlap between their conditions. However, the virtue epistemology account, i.e. weak ECI, goes further to elucidate how we can include knowledge-conducive belief-forming processes in our cognitive characters, which may be more demanding than the MCI account. Palermos' work shows that there may be symmetric entailment between the two integrations, but there surely is at least an asymmetric entailment. In what follows, I will describe Clark's dilemma with the close fit between these two integrations, and I will argue that it is only a problem for the symmetric entailment (not the asymmetric view).

## The hygiene problem

Clark's *epistemic hygiene dilemma* suggests that the two integrations, ECI and MCI, may not be a very close fit (2015).<sup>39</sup> Simply put, the dilemma points out that the epistemic integration account requires more active agential involvement as compared to the metaphysical integration. To unpack the dilemma thoroughly, I will first discuss what epistemic hygiene is and what is the basis of this dilemma.

In the last chapter, I mentioned a (non-internalist) 'subjective justification' requirement that externalist epistemologies often include in their theories of justification in order to ensure that the epistemic agent is responsibly employing her reliable belief-forming process. Essentially, the agent can only responsibly employ her reliable process if she is aware of the reliability (internalist/awareness condition) or has some kind of responsiveness to the potential unreliability of a process. Both these routes can allow the agent to meet the epistemic

---

<sup>39</sup>Clark, like Palermos, does not distinguish between the two integration accounts and, therefore, does not use the terms ECI and MCI for these accounts.

responsibility requirement on knowledge. This is what Clark calls the practice of epistemic hygiene. To take care of epistemic hygiene is to fulfil the responsibility condition on knowledge. And, recall from the previous chapter, that there is a strong route to epistemic hygiene and a weak route (that weak ECI is based in).

Clark's epistemic hygiene dilemma specifically states that an agent has to consciously encounter a reliable belief-forming process, or practise some other form of active epistemic hygiene, to make it a knowledge-generating cognitive ability. And, such a conscious or active encounter of the target process is antithetical to the extended cognition and not in line with MCI.

The idea is that the active agential involvement that epistemic hygiene requires is not in line with functional parity. As you may recall, the functional parity principle informs what a cognitive process is. Accordingly, when an external mechanism is employed the same way or performs the same functional role, as the cognitive process inside our head (a neural and innate faculty) then employing the external process should also be considered cognitive. Our agency employs our cognitive processes in a seamless manner, without any active involvement. Clark writes that our biological memory 'for the most part, functions in a kind of automatic, subterranean way. It is not an object for us, we do not encounter it perceptually. Rather, it helps constitute us as the cognitive beings we are' (Clark 2015, 3762).

Consider the strong agency account that I discussed in the last chapter. If we are to integrate a belief-forming process via the strong route, we will need to gain access to the reliability of our process. This is in contrast with how we employ our innate processes; we rely on them seamlessly. Gaining access to the reliability of a process is a very active form of practising epistemic hygiene, and it is, therefore, not consistent with extension. It requires the agent to actively and consciously recruit her process in a responsible manner. We need to look for a less active form of epistemic hygiene— an ECI account that is similar to our seamless reliance on our innate processes. So far, epistemologists, like Pritchard and Palermos, are on board with the view that strong ECI requires too active an agency. Epistemic hygiene can be a very active pursuit, and it often involves awareness conditions (i.e. the same as to demand access to the reliability of one's process), which is why both Palermos and Pritchard argue that a weak

ECI is in line with extended cognition.

Weak ECI is described by the function of cooperation and interaction within one's cognitive character (Greco 2010). A target process forms cooperative interactions with other processes in one's cognitive system to integrate in terms of weak ECI. This integration occurs on the basis of the output from the target process that are assimilated by the agent's innate processes to produce more output which are further assimilated. This occurs in a way that keeps the cognitive loop moving forward. The interconnectedness that results from this flow and exchange of information allows the agent to monitor her process's reliability in the background. She does not require positive reasons or any forms of awareness to integrate her process. Her process integrates as she gains counterfactual sensitivity to its reliability in virtue of its connections with other faculties of hers.

However, Clark argues that even weak ECI cannot account for the minimal agency requirement in cognitive extension. He states:

'Thus we are told that what this kind of minimalist demand ("integration with her cognitive character') delivers is a kind of sensitivity to overall coherence such that "the agent must be able to become aware that the process is unreliable in certain circumstances" (op cit, p. 1939). I read this to mean that at times an agent who meets Palermos' minimalist condition must be able to become consciously aware that a belief-forming process (such as the notebook-process, in the case of Otto, or the thermometer-process, in the case of Temp) is unsafe...I shall question even this requirement' (2015, 3766).

The reason he finds the weak ECI account too active is because it still demands a kind of conscious or *personal-level* (associated with the whole person) engagement from the agent.

The personal-level is that which is associated with the whole person, whereas the subpersonal is concerned with only a part of the person. Things that are consciously accessible to us are available to us at a personal-level. For

example, my beliefs are a personal-level phenomenon.<sup>40</sup> I am either aware of my beliefs or I am in the position of becoming aware of them. In contrast, some phenomena are not personally or consciously accessible to us. For instance, the wavelength information that some nerves in my eyes collect and send to other nerves in my brain is not present to me at a personal-level. It is only present to me at a subpersonal level, i.e. parts of my cognitive system can interact with that kind of information.

Clark argues that even in weak ECI account the agent is involved at a personal-level. This is because weak ECI requires the agent to be in the position to become aware of any counterevidence to the reliability of her process. It treats reliability of the agent's process, or the lack thereof, as a phenomenon that the agent *can* become conscious of. This means that the state that weak ECI puts the agent in is a personal-level state and therefore requires personal engagement from the agent, even if only when the process malfunctions. Clark urges for an even weaker account, 'one that does not require any kind of conscious or personal-level engagement between the agent and the cognitive process on the part of the agent at all' (Clark 2015, 3766).

One can argue that the weak epistemic integration account does not involve the agent at a personal-level. After all, the principal tenet of the weak ECI account is that integration of a new process is based on the interconnectedness of one's cognitive system. If the new process is reliable, it will be able to join the cognitive loop on the basis of its output. Other faculties of the agent's cognitive system will be able to assimilate the output and continue the cognitive loop. And, if the target process forms less than reliable beliefs then it will not integrate. The cognitive faculties will simply reject the target process on the basis of its output. It is the interconnectedness of an agent's cognitive character that either allows or declines integration of a new process, and it is an entirely subpersonal process. Comparatively, in a strong ECI account, the person herself either accepts or rejects a belief-producing process for integration, but, in a weak ECI account, this process occurs subpersonally.

---

<sup>40</sup>We can also understand belief production at a subpersonal level but this level will not involve the concept of beliefs. On this level, we will discuss the operations that go on in parts of one's cognitive system that bring about a mental state (like belief).

This means that weak ECI should be consistent with an entirely subpersonal MCI account (and extended cognition).

The virtue reliabilists can also argue that the agent only becomes aware of something when the process is no longer integrated.<sup>41</sup> The agent does not become aware of anything as long as the target process is working fine and is properly integrated into the agent's cognitive system. She only becomes conscious (and, therefore, personally involved) when there is a lack of integration, i.e. when her process begins to malfunction. When a problem with the reliability of her process surfaces, she becomes aware and conscious of the problem. But at this point, the said process is no longer integrated, because it is no longer reliable; reliability is a necessary precondition for weak ECI. Therefore, the personal-level phenomenon (i.e. awareness) should not be linked to weak ECI when it is only concerned with the failing of weak ECI. The actual process of integration manifests subpersonally.

In my understanding, Clark does not care that the agent's employment of her resource, according to weak ECI, is unconscious; he is only concerned that it involves the agent's cognitive agency at a personal-level. In fact, he should not have used the word 'conscious' to describe weak ECI at all, because the actual process that allows integration has nothing to do with consciousness. His argument should have only focussed on the personal-level aspect of weak ECI. Weak ECI occurs on the basis of the agent's responsiveness or sensitivity to the (counterevidence to) reliability of her process. This sensitivity is a personal-level concept. It is not just a part of the person that is becoming sensitive to the reliability (or the lack thereof); it is the whole person. So, even if the process never malfunctions and the agent never becomes aware of anything, it is still the personal-level sensitivity that takes care of epistemic hygiene in a weak ECI account. Furthermore, as I mentioned in the previous chapter, there is a specific degree of interconnectedness that allows the agent counterfactual sensitivity to the reliability of her process, and only when this degree of integratedness has obtained can we say that the process is integrated via weak ECI. This is why the personal-level concept of agential sensitivity is at the heart of the weak ECI

---

<sup>41</sup>The virtue reliabilists who want to argue in favour of the snug fit between weak ECI and extended cognition will want to argue along this line.

account: it determines the degree of interaction required between a new process and other faculties in the agent's cognitive system.<sup>42</sup>

The epistemic hygiene dilemma can further a non-univocal view, i.e. the concept that the ECI and MCI are entirely different and may have no conditions in common. Neither a symmetric, nor an asymmetric entailment holds between the two integrations. One of these integrations occurs without any personal-level involvement from the agency, and the other requires personal engagement from one's agency. However, Clark's hygiene dilemma is not antithetical to the asymmetric univocal view either. It can still be that all of MCI conditions are required for ECI, but ECI also requires a little more personal-level involvement from the agency. The dilemma is indeed a problem for the symmetric view (i.e. same conditions can explain both integrations). However, Clark's own solution to the epistemic hygiene dilemma is also a symmetric univocal view, one that I unpack and analyse in the next section.

## A subpersonal epistemology

Clark's solution to the epistemic hygiene dilemma is to introduce an even weaker formulation of epistemic integration, which, he argues, is compatible with the MCI account and is in line with how epistemic agents recruit their innate cognitive processes. With this weaker epistemic integration account, Clark is able to make the claim that conditions that bring about cognitive extension and develop a reliable belief-forming process into a cognitive ability are the same.

Simply put, Clark's weaker epistemic integration account is a subpersonal epistemic integration. I am going to refer to it as subpersonal ECI, henceforth. To understand the subpersonal view, take weak ECI and, from it, remove the claim that for a process to integrate, the epistemic agent becomes counterfactually responsive to its reliability.<sup>43</sup> So, once the personal element of

---

<sup>42</sup>I would like to draw attention to a point here that might be unclear. Conflicting information does not necessarily indicate that the target process will un-integrate. It only indicates that the agent will become aware of the conflict and be able to assess if the process needs to be employed again or not.

<sup>43</sup>Subpersonal ECI is the most minimal interpretation of epistemic integration for Clark. He acknowledges the validity of the weak and strong route for epistemic integration, but finds his

weak ECI is dismissed, we are left with the basic understanding of a subpersonal ECI. Also note, that Clark's ECI account is 'a sub-personal take on core insights from virtue epistemology' (2015, 3770). He also mentions that his approach can fit with a sophisticated version of process reliabilism, like the one that Kornblith (2012) and Goldman (2012) support, but he gives no further details of this account. He does, however, name one of his sections, 'Towards a subpersonal virtue epistemology,' so I think it's fair that I treat his proposition as a version of virtue epistemology (Clark 2015).

As a version of virtue epistemology, Clark's subpersonal ECI ought to describe the agent's link to the reliability of her process, so she can responsibly recruit it. In my understanding, in lieu of the claim that the agent develops counterfactual sensitivity to the reliability of her process, he will need to add to his account a subpersonal interpretation of this claim. One such claim may be that *a (subpersonal) part of the agent's cognitive system* gains access or becomes sensitive to the reliability of the agent's belief-producing process and therefore allows the new process to integrate into her cognitive character. Hence, Clark's subpersonal ECI is almost the same as weak ECI, except that it clearly does not have any personal-level element in its explanation. The agent, herself, at a personal-level does not need to have access or responsiveness to her process's reliability to meet the responsibility requirement. This requirement can be met entirely subpersonally. This, for Clark, is the most minimal portrayal of epistemic integration and, therefore, consistent with what makes a process cognitive.

In the general context of epistemology, Clark's subpersonal virtue epistemology holds an interesting position. Epistemic hygiene or responsibility has been a significant concept for epistemologists. Most internalists argue that an epistemic agent ought to be aware of the reasons that make her beliefs true. Externalists have also acknowledged that (unlike Norman the Clairvoyant) epistemic agents ought to have some kind of link to whatever it is that makes their belief true, even though they need no awareness condition. Without such a link, epistemic agents cannot be responsible (or maintain epistemic hygiene) in forming their beliefs, which also means that their beliefs will not be creditable

---

version the weakest interpretation which is also in line with cognitive extension.

to them. Clark (2015) proposes a new way an agent can meet the responsibility requirement on knowledge— by an entirely subpersonal link to the reliability of one's process.

Consider a closer look at how Clark's view is a symmetric univocal proposition. We discussed above on how MCI, the integration that marks a cognitive process, is best understood in terms of CRC or subpersonal on-going, non-linear information exchange between the target process and other faculties in the agent's cognitive character. CRC is the degree of reciprocal interconnectedness between two systems that allows them to temporally evolve into one system or one cognitive process. Subpersonal ECI, according to Clark, can also be fully explained in terms of CRC. When the degree of non-linear interaction between a target process and other cognitive processes in the agent's cognitive system is sufficient for subpersonal ECI, the former merges into the latter, forming a process that is both cognitive and epistemic. This means that CRC (does not only explain cognition but also) allows other processes in one's cognitive system to subpersonally vouch for the reliability of the target process, so it can be epistemically integrated.

One of the problems with this symmetric view or any symmetric view is that you cannot have a cognitive process that is not also epistemic. But, clearly, such processes do exist. When I go to the kitchen to fix myself a sandwich, there is cognition involved in the movement I make. Granted that I do rely on some epistemic processes as well, for instance I will need justified beliefs about where the bread is, if I need more jam, how much jam I like on my bread, etc. But, some of the processes involved in this event will have nothing to do with forming beliefs and getting them right. For instance, I will need to navigate to the kitchen, if I want a sandwich. For that, I will pay attention to physical cues, without forming beliefs about them. Moreover, not all cognition is present to us consciously. Most things we perceive throughout the day, we do not form beliefs about them, but they play a part in our cognition. This is one reason why epistemologists might find it hard to get on board with an entirely subpersonal epistemology. Knowledge is a personal phenomenon that is creditable to the agent when she manifests some kind of personal cognitive agency. Knowledge is more valuable than (true) belief. Some cognitive processes are involved in

forming beliefs, whereas epistemic processes have the ability to form knowledge. This status of knowledge compared to beliefs indicates that the integration required to form an epistemic process ought to be more demanding than the one that forms a cognitive process.<sup>44</sup>

To fix this worry, within the realm of a subpersonal epistemology, we can adjust his view ever so slightly towards an asymmetric approach. This would mean to adjust our claim about MCI. Accordingly, we can say that CRC defines a degree of integration that allows a target mechanism to merge into our cognitive process. Whereas, for the same process to count as epistemic, it will have to at least integrate to a degree, that allows other processes in the cognitive system to (subpersonally) vouch for the reliability of the said process. This is an asymmetric representation of a univocal view, that includes subpersonal ECI as the account of epistemic integration.

But Clark's symmetric portrayal of the univocal view is the least of the worries. The major point of dissent for epistemologists will be that he is using an entirely subpersonal mechanism (part) to explain what makes an epistemic agent (whole) responsible in employing her reliable process. And, this responsible employment may result in a true belief that is knowledge, which will be (at least significantly) creditable to the agent's cognitive agency. However, the cognitive agency is manifested by only a part of the epistemic agent, and it has no link to the agent (the whole person). Of course, Clark is of the view that subpersonal cognitive agency (i.e. agency manifested by a part of the agent) is sufficient as the significant agency that knowledge can be creditable to. In other words, an

---

<sup>44</sup>Recall that I indulge in this debate in the previous chapter, where I introduce weak epistemic integration thoroughly. For weak epistemic integration, it is crucial that there is a specific level of interconnectedness between the target process and other cognitive processes in the agent's cognitive system. This level is marked by the clause that the agent will be in the position to become aware if the target process began to provide conflicting evidence. This is the description of how an epistemic process is formed (i.e. a sufficient degree of interconnectedness marked by a clause). On the other hand, according to DST, for a process to be considered cognitive it should interconnect with other processes in a way that improves the overall behavioural competence of the system. In my understanding, a lower degree of interconnectedness is required for the latter, whereas a higher degree will be required for the agent to become sensitive to the reliability of her process. I provide more detail on this matter in the next section.

agent can be credited with knowledge on the basis of the significant subpersonal agency she manifests. But, the question is, can Clark argue for this conclusion? In chapter 4, I unpack Clark's predictive brain theory in which he grounds his subpersonal ECI account and that is where I explore and settle this question.

In the remaining chapter, I will argue that there is no need for a subpersonal epistemology, and I will re-emphasise that Pritchard and Palermos' weak ECI account is sufficiently subpersonal.

## No need for a subpersonal epistemology

Firstly, the epistemic hygiene dilemma is not a threat to the asymmetric univocal view. This is the view that describes MCI conditions as a subset of what is required for epistemic integration. On this view, it can be stated that a process qualifies as cognitive if it fulfils the CRC requirement. So, an external mechanism will be absorbed into my cognitive process if it continuously and reciprocally exchanges information with it. And, this process will be considered epistemic when, on top of CRC, the integration also provides me with counterfactual sensitivity to the reliability of my process. So, in its most minimal interpretation (that is compatible with cognitive extension), ECI will be equal to MCI plus agent's link to the reliability.

I would like to review the relationship between MCI and ECI in accordance with an aspect of integration– the degree of interconnectedness. Both integrations, MCI and ECI, have interconnectedness, on-going cooperation, and complementarity at the heart of their accounts. So, reciprocal interconnections are the basic unit that both ECI and MCI require. It may be that ECI and MCI differ mainly in terms of the degree of interconnectedness and complementarity that they require. MCI requires the degree of cooperation (within the target process and other processes in one's cognitive system) that allows the target process to become a part of one's cognitive process, such that they function and produce output or affect other processes together, as a whole.<sup>45</sup> Jennifer Greenwood explains that the move from deep interconnectedness to constituting

---

<sup>45</sup>This is in line with the dynamic systems theory (DST) that I discussed in the beginning of this chapter.

a cognitive process occurs when 'the degree and complexity of functional complementarity and integration in extended cognitive systems is such that the internal and external components of the system function synchronically as a whole, producing effects resulting only from the system functioning synchronically as a whole' (Greenwood 2013, 425). Similarly, for a process to (minimally) epistemically, integrate the sufficient degree of integration may be slightly more demanding than what is required for MCI. The degree of interconnectedness needed for weak ECI may be adequate when the agent becomes counterfactually sensitive to her process's reliability. If she is not in the state to become conscious of her process's failing reliability, then her epistemic hygiene she has practised is unlikely to give her knowledge.

On this hypothetical view, there is no reason to think that the same degree of interconnectedness cannot bring about both integrations. The degree of reciprocal connections that make a process cognitive may also make it epistemic (one that allows the agent to responsibly produce true beliefs that may be knowledge). So, symmetric entailment between epistemic and metaphysical integration can hold for some processes, but because it is not the norm, we should not define the relationship between the two integration in these terms. For example, it is possible that the same degree of interconnectedness between Otto's process of manipulating his notebook and other processes in his cognitive system can make Otto's employment of his notebook a cognitive process for him and also make it epistemic. The former would require that Otto's manipulation of his notebook become a part of his cognitive character (in terms of DST), and produce output or affect other processes as one unit. And, Otto's process would become epistemic when Otto has at least developed sensitivity to the reliability of his process (i.e. his manipulation of the notebook).

We have considered one aspect of integration, namely the degree of interconnectedness. The cognitive architecture is not as simple to depend only on this one factor. There are myriad other elements to consider when comparing the interconnections that lead to these integrations, some we might not even know of yet. Primarily, integration depends on the type of target mechanism, the functional role it plays, and the system that it integrates into (Carter

and Kallestrup 2019).<sup>46</sup> A target mechanism integrating into my perceptual system will form different kinds of bonds to cooperate and interact with other processes than the one integrating into my memory system. Similarly, in another individual, contingent on what they do for a living, how their eye-sight is, etc., a target process may integrate into their perceptual system differently from how it would in me. Salma has a visual aid implanted in her brain. Let's say that Salma is an ophthalmologist, which means she has to look into other people's eyes to see if their vision is fine. Now, if Salma and I both get the same visual aid implanted in us, hers is going to integrate differently. Especially in terms of epistemic integration, since Salma has a more pressing need to form true beliefs about the patients' eyes she examines, her implant will epistemically integrate differently from someone who did not have a job so heavily reliant on their perceptual system.

Furthermore, external mechanisms that extend our cognition can play so many different functional roles that the way they integrate changes accordingly (Carter and Kallestrup 2019).<sup>47</sup> Otto's notebook process (i.e. his process of physically manipulating his notebook a certain way) will integrate differently compared to an AI assistant that a chief executive officer of a multinational company relies on. This smart assistant is programmed to pick details about all the meetings from this person's conversations and list them in a proper system. He can also manually enter his meetings. The AI, then, reminds him of the 20 different engagements he has every day in a timely way, so he is not late for any of them even if he has to drive to different places for each of them. Where Otto writes things down in his notebook and carries it around everywhere, the smart assistant manages the calendar mostly on its own. So, even though both these examples are of processes that belong to (mostly) the memory system, their different functional roles will mean that they will integrate differently.

Moreover, all these external processes will also be very different in the way they epistemically integrate. Once again, it will be the function of the process that will determine how much cooperative interaction will be required

---

<sup>46</sup>I use the generic word integration because all of this is true for both integrations.

<sup>47</sup>On this point, Carter and Kallestrup (2019) mention the futility of looking for necessary and sufficient conditions for integrations.

for the agent to become counterfactually sensitive to its reliability. The Salma case above points out how some individuals may be in positions where there is a pressing need for their process to epistemically integrate better or faster in comparison with others. Also, epistemic integration can depend on other factors like the type of the belief-forming mechanism, the agent's cognitive character, the conscientiousness of the agent, etc.

Overall, the asymmetric view grants that while cooperative interaction is something that both integrations have in common, there are still many other differences between these two integrations. The epistemic hygiene dilemma only points out one of these differences, one that highlights that epistemic integration requires a hint of more personal involvement than metaphysical integration. Hence, there is no need for a subpersonal epistemology.

The last point I want the reader to reconsider is that the process that causes weak ECI is entirely subpersonal— again, the need for an entirely subpersonal epistemology, based in a weaker understanding of integration, does not befall us. Weak ECI occurs on the basis of subpersonal complementary cooperation between a target process and other processes in one's cognitive system. It does not require the agent to be aware of anything. The personal element (i.e. counter-sensitivity to the reliability of one's process) that arises as a result of these subpersonal cooperations is, in fact, useful. It helps mark the degree of complementarity and integratedness that is sufficient for weak ECI. Hence, it is not a convincing claim that weak ECI involves the epistemic agent on a personal or conscious level when it is entirely possible for the agent to become aware of nothing throughout the weak ECI process. If the target process's reliability does not falter, she will not become aware of anything, and the target process will be considered integrated via weak ECI.

## Conclusion

So, in Palermos' work, we find a new argument that describes cognitive extension better than the glue and trust conditions. With this argument based on the dynamic systems theory, he argues how MCI and ECI both require cooperation and complementarity between the target mechanism and other

processes in the agent's cognitive system. In its basic form, this argument shows an asymmetric entailment between the two integrations. Clark, with the epistemic hygiene dilemma, emphasises that metaphysical cognitive integration is entirely passive and subpersonal, whereas even the weakest ECI, i.e. weak ECI, still requires some personal-level engagement from the agent. He answers his own dilemma by proposing a different ECI account, the wholly subpersonal ECI. This epistemic integration approach explains, in terms of the subpersonal goings-on of the agent's cognitive system, how she can responsibly produce knowledge that can be significantly creditable to her. I argue that this account is a symmetric view and ought to answer problems that all symmetric views are vulnerable to. I also discuss the aspects of a wholly subpersonal epistemic integration that virtue epistemologists will take issue with. In the end, I support the asymmetric relationship between the two integrations which is not susceptible to the epistemic hygiene dilemma. Since, it does not demand symmetry between epistemic and metaphysical integration, there is no need for a different epistemic integration account at all.

This chapter has unpacked important ideas that I need to discuss throughout the rest of my thesis. Along with the previous chapter, it has introduced the various ways that the relationship between the two integrations is outlined. The reason that we are looking in this direction at all is that bringing ECI and MCI together will help us describe external mechanisms that extend our cognition and are also able to provide us with knowledge. How these processes come about is at the centre of my examination of this topic. This chapter introduces symmetric and asymmetric proposals, of which I favour the latter.

Another important purpose of this chapter was to briefly lay out the subpersonal ECI account and touch upon the debates that are related to it. On this account, Clark has explained, entirely subpersonally, how an agent can responsibly invoke her reliable process. Chapter 4, 5, and 6 will provide more insight into Clark's account, where he is coming from, the theories he grounds his wholly subpersonal notion of epistemic integration in, and the several issues that such an approach can lead to.

The immediate next step in my thesis is to consider another major difference between Clark's account and the weak ECI account. He points out

that he can envision new technology that will integrate into our cognitive systems instantly and provide true beliefs (that may be knowledge). In other words, the epistemic integration account that Clark thinks will snugly fit with his understanding of cognitive extension is one that can happen immediately. I will argue that epistemic integration does not and cannot come about as immediately as Clark's metaphysical cognitive integration.

## Chapter 3 Immediate epistemic integration

In the previous chapter, I discussed Clark's epistemic hygiene dilemma. This dilemma brings forward the incompatibility of the integration that explains why a process is our cognitive process and the integration that makes the said process an epistemic one. Generally, the former integration (i.e. metaphysical cognitive integration or MCI) can be understood at an entirely subpersonal level, whereas the latter (i.e. epistemic integration or ECI) requires some personal-level explanation as well. Clark's own response to this dilemma is to suggest a different and more minimal account for ECI— one that can be fully understood at a subpersonal level. I have called it the subpersonal ECI account. With this proposal, Clark gives us a symmetric univocal view, which is the idea that the conditions for MCI are the same as those of ECI or that processes that meet MCI conditions, also meet the minimal ECI conditions. In other words, a target mechanism has to fulfil the same requirements to become our cognitive process (i.e. extend our cognition) and to become our cognitive ability (i.e. our reliable belief-forming process that can give us knowledge).

In this chapter, I am going to examine this symmetric view in terms of a claim that Clark makes about it. He states that new technology will soon integrate into our cognitive systems in a way that it can produce knowledge for us right away. He envisions modern external mechanisms to *immediately* metaphysically integrate into our cognitive systems, such that they extend our cognition, and also epistemically so that they can provide us knowledge. I call this concept *immediate integration*.

With the help of analogies, cases, and examples, I aim to demonstrate that epistemic integration usually does not occur as immediately as MCI or as immediately as the immediate epistemic integration that Clark mentions. I make use of Pritchard's account of the Tempo and Tempo\* case, whose implants are fitted at birth and a later stage in life, respectively.<sup>48</sup> Pritchard elaborates how weak ECI, the model of epistemic integration that fits well with extended cognition (according to Pritchard), can occur relatively swiftly when the device

---

<sup>48</sup>For a more detailed discussion of the Tempo case, consult Pritchard, Millar, and Haddock (2010).

is implanted at birth, compared to when it is fitted later in life. Here, I argue that even the epistemic integration that occurs right after birth, the one that Pritchard describes with the Tempo case, is still not as immediate as the kind of epistemic integration that Clark has in mind, neither is it as immediate as metaphysical integration. With this, I show that what Clark has in mind for minimal epistemic integration is indeed different from the weak ECI account that Pritchard and Palermos favour as one that describes extended cognitive abilities.

I begin this chapter by bringing the glue and trust conditions plus the continuous reciprocal causation (CRC) requirement, as in the 3+1 model of MCI, into question. Neither the glue and trust conditions nor the CRC is entirely compatible with Clark's claim about immediate integration. Then, I introduce the neuroscientific concept of *deep functional integration* which is another way to look at the continuous reciprocal causation (CRC) requirement. In terms of these concepts and the Tempo and Tempo\* cases, I examine if immediate epistemic integration is possible. Deep functional integration of the implants in the Tempo and Tempo\* cases provide insight into how and why epistemic integration requires time to develop, and how deep functional integration of these implants develops first.

Finally, this chapter also looks into the concept of univocal views and discusses them in terms of Clark's immediate integration claim. While the latter seems to suggest a symmetric entailment or a complete overlap between the conditions of metaphysical and epistemic integration, my discussion of the Tempo and Tempo\* cases and the deep functional integration, indicate a partial overlap of conditions between the two integrations.

## Clark's immediate integration

Clark describes immediate integration in the following way:

I believe that a piece of new cognitive technology could be so well-designed as to be immediately assimilated into our daily routines, requiring no reflective window before properly being counted as delivering knowledge. Such a well-fitted device would enable new states of knowing from the very moment it is fitted

(much as do the sensory organs of creatures that do not have the luxury, immediately after birth, of spending time 'testing them out' in a nurturing or protected environment. (Clark 2015, 3764)

Clark envisions that new technology will be able to immediately integrate into our cognitive systems, metaphysically and epistemically. A modern gadget may become our cognitive process and also our cognitive ability that can produce knowledge for us the very first or second time we employ it.

Clark points out in his quote that his concept of immediate integration is in line with functional parity. Right after birth, we employ our innate faculties in a way that they forward our cognitive loop and become a part of our cognitive system. And, we can immediately employ some of these processes so that they can generate knowledge for us. He argues that future technologies will be able to integrate this way because it is functionally equivalent to how our innate faculties integrate. That is, since our innate faculties can immediately integrate via both MCI and ECI, so can the external and artificial processes that we come about.

Although, a simple problem for the functional parity principle is that an agent's cognitive character is different when it is still developing compared to when it has fully matured. Can we even compare integration that occurs later in life to the integration of our innate processes that occurs soon after our birth (when our cognitive system is developing), when the former takes place in a mature cognitive system and the latter does not?

The best account of metaphysical cognitive integration, or how an external target mechanism becomes a part of our cognitive system, is in terms of the 3+1 model. The 3+1 model demands that we understand MCI in terms of Clark's three glue and trust criteria and the continuous reciprocal causation (CRC henceforth) requirement. Both of these concepts have been unpacked properly in the previous chapters. Here, I am only going to invoke them to indicate that these ideas do not properly support Clark's claim about immediate integration. In this section, I will mainly focus on immediate integration as a metaphysical integration claim.<sup>49</sup>

---

<sup>49</sup>I am not claiming that MCI cannot occur immediately. It is possible for surgeons to im-

Firstly, consider the glue and trust conditions. These require that the target process and its information is readily available, easily accessible, and trustworthy for the agent's cognitive agency. Clark presents these requirements in reference to the Otto-Inga Case. As you may recall, Otto has Alzheimer's, and he uses his notebook as an extension of his cognition. Otto's process of manipulating his notebook meets the glue and trust conditions of availability, accessibility and trustworthiness. I imagine, that Otto did not begin to use his notebook in this way, overnight. Over time, he learned to use his notebook a certain way, rely on it more fully, and properly trust the information from it. It is possible that his notebook was readily available from the off (that is, from the first time he used it), and the content of the notebook was easily accessible from the very beginning, but it must have taken time for Otto to begin to trust this information. So, the glue and trust are not compatible with immediate integration.

Secondly, even the CRC requirement does not provide support for immediate integration. It requires that the target process should engage in non-linear and on-going information exchange with a part of the agent's cognitive system, such that the input is recycled as output making the cognitive loop move forward (Clark 2008; Chemero 2009; Palermos 2014a, 2014b). An *on-going* information exchange between two processes is likely to take time, and it is unlikely to happen immediately.

So, it is difficult to find a clear understanding of immediate integration in terms of the glue and trust conditions and the CRC requirement. This does not mean that it is impossible to argue that MCI can occur instantly. Clark (2013, 2015, 2016a) and others (Friston 2009; Hohwy 2013) put forward a concept of cognition and perception based on the predictive brain. On this view, Clark describes metaphysical integration and a kind of epistemic integration that occur immediately. But, I need an entire chapter to deal with how Clark's integration account is grounded in his predictive processing model, I am going to unpack

---

plant a visual aid implant in my brain that immediately becomes a cognitive process for me, and helps me fulfil (non-epistemic) cognitive tasks, like navigating myself through the kitchen to fix myself a sandwich. Here, I am only pointing out that the 3+1 model does not provide a satisfactory understanding to Clark's immediate integration claim.

this properly in the next chapter. Here I will touch upon it superficially and mostly focus on analysing immediate integration in terms of the theories that have been fully unpacked in the previous chapters.

One way to grasp immediate MCI is to consider an external mechanism that forms sufficient reciprocal information exchanges with a local cognitive process, that when it produces a cognitive output for the first time (which is not epistemic), it does so as a synchronous whole. In other words, due to the subpersonal exchanges of information the output produced comes from an evolved cognitive process that includes part of one's cognitive system and the external mechanism.

Another way to understand this is to think that a component of our cognitive system has the capacity to predict if the said process will metaphysically integrate or not. If you recall, important aspects of Clark's cognition account were discussed in the previous chapter. On Clark's view, metaphysical integration is necessarily a subpersonal affair and epistemic integration can also, at times, be explained entirely subpersonally.<sup>50</sup> So, when he claims that he can envision target mechanisms that undergoes both these integrations immediately, I imagine that these mechanisms have a kind of *MCI switch*. Our cognitive agency (i.e. a part of our cognitive system), on Clark's account, can detect if the said mechanism will (in future) develop continuous and reciprocal exchange of information with the rest of the cognitive system or not. If an external mechanism has this switch, the agent's cognitive agency will immediately integrate it metaphysically. I agree with this immediate notion of metaphysical integration. What I have reservations about is its epistemic counterpart.

To grasp Clark's immediate epistemic integration, think of an *ECI switch*.<sup>51</sup> Accordingly, a part of the agent's cognitive system can detect if a certain external process (e.g. the technology Clark envisions) will be

---

<sup>50</sup>These are claims from Clark's view that I mentioned in the previous chapter. To fully unpack his integration account I have to delve into his concept of predictive brain, which I do in the next chapter. For this chapter, I am going to put Clark's subpersonal ECI account aside and focus on the weak ECI that Pritchard and Palermos support.

<sup>51</sup>This is one way to understand Clark's subpersonal epistemic integration account mentioned in the previous chapter.

reliable to employ based on the presence of this switch. If the said process that manipulates this technology has the particular ECI switch then one's cognitive agency will be able to immediately integrate it and employ it as a knowledge-conducive process. In other words, subpersonal goings-on of the cognitive system can sometimes ensure that the first time a target process is employed, it can immediately produce an epistemic output that can be knowledge. I will unpack Clark's account further in the next chapter. Here, I will endeavour to account for this immediate integration in terms of Pritchard's weak ECI account (described in the previous chapters).

## Deep functional integration

Over the years, brain imaging studies have helped scientists of the mind observe how the different parts of the brain work together to perform certain functions. A famous study, Bartels and Zeki (2004), uses functional magnetic resonance imaging (fMRI) to see brain patterns for love. People's brain activity is scanned while they are looking at photos of loved ones. These studies indicate that specific parts of the brain are not responsible for bringing about different functions, in fact, different parts of the brain interact heavily to bring about different functions. Specifically, fMRIs allow neuroscientists to observe and study brain activity patterns for different behaviours that agents execute. Since the advent of these imaging tools, it has become easier to study the impact of external as well as internal processes on one's cognitive system and the similarities between them. These fMRIs do not only map patterns that indicate the parts of the brain that are engaged in accomplishing a cognitive task, but also their intensity signifies the degree of interaction or involvement of these parts.

When different processes in the brain interact with each other in strong and intricate patterns, the phenomenon that can be observed amongst them is called *deep functional integration* (Greenwood 2013). Not all interactions and integrations amongst processes or parts of the brain can be dubbed deep functional integration. This particular phenomenon is only observed when a process interacts with a system or other processes in ways that augment the

cognitive system. Simple input of sensory signals from the process cannot generate the kind of patterns that demonstrate deep functional integration.

To unpack deep functional integration further, let's first consider the rival theory of extended cognition– *hypothesis of embedded cognition* (HEMC). Rob Rupert argues that cognition does not extend into the environment; it is, however, very heavily embedded into the environment. In light of the cognitive bloat problem that suggests that cognition may leak everywhere into our surroundings and the coupling-constitution fallacy (ideas that I discussed in detail in the previous chapter), Rupert suggests that we ought to understand cognition as organism-bound but very heavily dependent on the environment (Rupert 2004, 2009). He writes, 'Cognitive processes depend very heavily, in hitherto unexpected ways, on organismically external props and devices and on the structure of the external environment in which cognition takes place' (Rupert 2004, 393). All in all, for Rupert, cognitive mechanisms remain inside a biological body but rely strongly on the input from the environment.

Greenwood argues that the brain imaging studies in neuroscience do not point towards the HEMC– or the environment's heavy influence on our cognitive capacities. The fMRIs indicate a stronger dependence on the environment, one that suggests that the environment is a part of the cognitive architecture. In other words, deep functional integration supports the extended cognition thesis. This is because deep functional integration, or the brain activity patterns on fMRI, suggest a very high degree of complementarity and complexity of integration between the process and the agent's cognitive system. The high degree of complementarity and complexity in integration is an indication of how the external target process (does not simply influence but) augments the rest of the cognitive system, and 'the internal and external components of the system function synchronically as a whole, producing effects resulting only from the system functioning synchronically as a whole' (Greenwood 2013, 425).

Deep functional integration is the neuroscientific depiction of the CRC concept. It is the kind of integration that Palermos (2014b, 2014a) and Clark (2001a, 2008) describe as the coming together of two processes and evolving into one that functions as a whole (and move cognition forward). The high degree of complementarity and complexity indicate continuous and reciprocal

exchange of information between the process and the cognitive system, such that a new process evolves. The augmentation of the cognitive system that results from deep functional integration is in line with the third hybrid system that two dynamic systems can temporally evolve into when they are exchanging information frequently.

So, as mentioned before, Greenwood explains that sometimes the external stimuli prompt brain activity patterns that only point towards embedded cognition (HEMC), i.e. cognition relies heavily on external props. However, at other times, the external input brings about such a high degree of complementarity and interconnectedness that the external process and part of our cognitive system work as one unit. This demonstrates cognition extending into the environment and not simply a strong dependence on environmental props.

Furthermore, Greenwood brings to our attention that 'the functional integration of intraorganismal and extraorganismal components is triggered initially and then developed through processes in which functional complementarity, complexity, and integration are progressively increased' (Greenwood 2013, 425). She presents cognitive ontogenesis or human emotional developments as an elaborate example of such deep functional integration that develops over time.<sup>52</sup>

Neonates or infants develop and mature human emotions by a process in which the faculties in their brains depend heavily on the external stimuli from their caregivers. For the caregivers to influence ontogenesis in their infants, they have to first comprehend the basic reactions that their neonates give them in response to baby talk, facial expressions, etc. In the very beginning, neonates can display very basic functions like sucking to calm themselves, looking away from stimuli, and some basic form of mimicking their caregivers' expressions of emotion. Caregivers first discover what the infants are trying to communicate with these responses and the stimuli that elicit these reactions. They also respond in baby-talk and with emotion-laden responses that the infants can further mimic. Their responses are also embedded in their culture and language.

---

<sup>52</sup>Also, see Spurrett and Cowley (2010) for a discussion on human cognitive development and the extended mind thesis.

That is, someone from Indonesia will react to crawling with the expression of disgust (they think crawling makes human look like animals and they do not appreciate it so much), whereas people from other places may express joy. So, in the beginning, these emotional expressions from both sides are undifferentiated and they slowly acquire complex functions. Processes in the caregivers' brains, and the sociocultural background they come from, influence processes in the neonates so they can develop from simpler emotions (disgust, fright, pain, etc.) into complex emotions (empathy, shame, guilt, etc.) (Greenwood 2013). She writes:

The entire human emotional ontogenetic process involves the progressive fine-tuning of these mechanisms, and this occurs through progressively fine-grained mutual modulation of neonates' assistance-soliciting and caregivers' assistance-providing interpreter mechanisms. It is through these mutual modulation processes that neonate and caregiver become intimately acquainted and attached, that is, develop the very special social relationship upon which human emotional ontogenesis partly but crucially depends. (Greenwood 2013, 431)

Notice how this process develops slowly over time. It depends on the learning and adjusting of processes in both cognitive systems (of the caregivers and neonates) to evolve into a process that augments the neonatal systems.<sup>53</sup> Human ontogenesis provides us with an example of MCI, the evolution of a new cognitive process in the neonates, that takes place over time.

All of our faculties do not mature this way. Of course, we can imagine that a basic faculty like perception metaphysically integrates immediately in a

---

<sup>53</sup>Greenwood's broader conclusion is that deep functional integration, especially the kind observed in human ontogenesis, points towards cognitive transcranialism. That is, that our cognition can extend outside our crania. Also, her case of human emotional development exhibits that cognition does not always extend from the brain to the world; it can also extend the other way. Human ontogenesis is an example of cognitive extension initiated by caregivers in the world that bring about an augmented process in the brains of the infants (Greenwood 2013).

neonate.<sup>54</sup> This is why we can tell that a newborn begins perceiving things right away. The important question here is if epistemic integration of the perception faculty also takes place this immediately? While the newborn begins to perceive, does he also begin to *know* right away? Let's consider these questions in the next section.

## Epistemic integration at birth and later

Clark's quote (mentioned above) mainly focuses on the future technology that will be able to immediately *epistemically* integrate such that it can provide us with knowledge right away. In this section, I will examine how and if this claim is possible in terms of the epistemic integration account that Pritchard and Palermos support– i.e. weak ECI. The reason behind this investigation is that weak ECI, as discussed in previous chapters, has been claimed to snugly fit with the hypothesis of cognitive extension. So, if the extension of cognition in future technology can happen immediately then it is worth scrutinising if epistemic integration, according to weak ECI, can occur the same way.

Very briefly, recall the weak ECI account. It requires that a target process should form cooperative interactions, of a high degree, with other processes in the cognitive system, such that these cooperative bonds can allow the agent's cognitive system to monitor the reliability of this target process (Pritchard 2010, 2018c; Greco 2010; Palermos 2011). The monitoring will be possible when there are strong interactions between the target process and the agent's cognitive system. A higher degree of interconnectedness is only likely to be a result of more connections the target process makes with the agent's cognitive agency.<sup>55</sup> And, for this to be possible, the target process should be employed by the agent's cognitive mechanism frequently for a long time, in a way that makes

---

<sup>54</sup>Deep functional integration is basically the neuroscientific sketch of continuous reciprocal exchange of information between two systems, and how it brings about a temporally evolved system. This account also does not properly establish how metaphysical integration occurs immediately. The account that does explain immediate MCI is one that will be unpacked in detail in the next chapter; it is the predictive processing model of the brain.

<sup>55</sup>I discussed in the previous chapter how both ECI and MCI have cooperation, interaction and complementarity as their basic fundamental unit.

more connections between the target process and the agent's cognitive system.<sup>56</sup> Generally, it seems like immediate weak epistemic integration is unlikely to transpire immediately, at least on this view.

Consider an example of weak ECI. Let's say, I am a fitness instructor, and I excessively rely on my smartwatch to get me through my workday. If I want to be able to use my watch in a way that allows me to acquire knowledge from my watch, it should be epistemically integrated into my cognitive character. I will have to use the watch consistently for many days. And, I will also have to develop trust in the information that it provides me. And, when my process of employing the watch forms an appropriately high degree of interconnections with other processes in my cognitive system, I will develop counterfactual sensitivity to the reliability of my watch. This is when my process will have achieved weak epistemic integration; it is also a slow epistemic integration.

In contrast to this slow process, I can immediately integrate my reliance on this watch by gaining a perspective on its reliability. This process is fast and requires a strong cognitive agency (i.e. access to the reliability of the source). This is one way to understand why Clark would think that future technology can instantly integrate in a way that we can acquire knowledge from employing it. But, this kind of strong integration shows how we employ tools and instruments; it does not explain a cognitive process extending into the environment. Hence, this is not the concept of epistemic integration that Clark is referring to when he mentioned immediately integrating technologies.

There is a scenario in which weak ECI can occur relatively faster than described above but not as immediate as Clark would want it to be. To unpack this scenario, consider Pritchard's comparison between a resource epistemically integrating into an agent's cognitive system at birth and one that epistemically integrates later in life (2010). Tempo is fitted with a temperature-detecting device at birth, and he lives in an environment where relying on such implants for temperature beliefs is not unusual. In the other case, Tempo\*'s surgeons implant his device in his head at a later stage in life. When, after an accident, Tempo\* is brought into the hospital in an unconscious state, his doctors perform

---

<sup>56</sup>Wheeler (2018) also comments on how reliance and trust in a technology increases over time.

a surgery on him during which they implant the temperature-detecting device in his head, and he has no knowledge of the implant. Pritchard's intuitions regarding these two cases are that the former agent, Tempo, will begin to produce reliable, knowledge-conducive beliefs, soon after his birth when he starts to form beliefs. On the other hand, Tempo\*'s new process will epistemically integrate into his cognitive character after he has used his device frequently for a significant amount of time and developed counterfactual sensitivity to his device.

Tempo's case demonstrates that when a device is implanted at birth and is operating in an environment where reliance on it is a norm, it can undergo weak ECI relatively quickly. Pritchard states that the agency has a lower demand when the process integrates at birth, like our other innate processes. He writes:

What the Tempo case seems to illustrate, and this is in keeping with Clark and Chalmers's Parity Principle, is that what is important in this respect is only that the cognitive abilities are present from the off– and are thus in this sense 'innate' (henceforth we will use this term in this loose way)– and not whether they are natural rather than artificially added to the agent. (Pritchard 2010, 146–47) In one sense, integration of Tempo's implant can be called immediate, because as soon as Tempo is old enough to form beliefs he will be able to form knowledge. This is because Tempo's implant will have integrated during the period his cognitive character is forming. However, even this immediate integration is not immediate enough to explain Clark's claim. Tempo's process epistemically integrates during the period that his cognitive character is developing, so by the time it is employed it is in the knowledge-generating business. Whereas, Clark has introduced a concept of integration that occurs at a later stage in life, in a cognitive character that is already mature. This concept is best understood in terms of the ECI-switch detector that I discussed above. Hence, Clark's epistemic integration cannot be understood in terms of the Tempo case that Pritchard illustrates.

Pritchard's characterisation of weak ECI (of non-neural processes) soon

after birth is in line with the functional parity principle. On his view, artificial implants at birth will epistemically integrate the same way as our innate belief-forming processes. When a child begins to form beliefs conscientiously, she will be able to employ her innate faculties as well as her artificial implants, fitted at birth, responsibly. Her implant will become a part of her cognitive character the same way, around the same time, that her cognitive faculties are forming her cognitive character.<sup>57</sup>

Contrastingly, processes that are not fitted at birth take longer to achieve weak epistemic cognitive integration.<sup>58</sup> Tempo\*'s device is not fit in him at birth but is surgically implanted at a later stage in life. His implant integrates by forming cooperative interactions with the rest of the cognitive system which comes from Tempo\*'s frequent employment of his implant. As a result of these cooperations between Tempo\*'s implant and other processes of his cognitive character, his agency becomes responsive to evidence against the reliability of his implant. This reminds us that the integration of a process without access to its reliability means that you ought to be responsive to evidence against its reliability.

Tempo and Tempo\*'s devices have integrated via weak ECI. This means that both agents have developed a sensitivity to the reliability of their reliable belief-forming processes, which is why they can detect any evidence against the reliability of their processes. The only difference is that Tempo reached the required sensitivity to his reliability faster than Tempo\*. The former, basically, attained the required level and degree of cooperation and interaction faster than Tempo\*'s device. Tempo's implant integrates much faster because it is fitted at birth, whereas Tempo\*'s device has to be employed for much longer to become a stable disposition of his cognitive character. I will explain the 'why' more thoroughly in the next section, in terms of the neuroscientific understanding of these cooperations and interactions that I mentioned.

Weak ECI is swift at birth and much slower at a later stage in life, but

---

<sup>57</sup>While epistemic integration occurs quite seamlessly and swiftly when faculties come together at birth, it still follows metaphysical integration. MCI occurs before and then ECI follows.

<sup>58</sup>This is how we acquire most of our cognitive character, that includes skills and methods of inquiry that we learn along the way.

there is a kind of epistemic integration that is immediate in the sense that it allows the agent to acquire knowledge the very first time that the agent engages with the process (unlike weak ECI at birth, which requires some time). This is the strong ECI account that I mentioned before in this chapter (and discussed in the first chapter). For an agent to integrate a process in this way, she will need to gain access to the reliability of her process. Tempo's device may be confused as one that has integrated via strong ECI, but that is not the case.<sup>59</sup> It would be strong ECI if Tempo could acquire access to the reliability of his process at birth, but we cannot reflect on the reliabilities of our processes at birth. Perhaps, we ought to rely on the access that Tempo's surgeons and doctors have to the reliability of the process that they are implanting in Tempo's brain. But they are not the epistemic agents in this case, and so their access to the reliability does not integrate the implant in Tempo's agency. Recall that in order to successfully meet the ECI requirements the agent is supposed to employ her process in an epistemically responsible manner. Tempo cannot employ his process epistemically responsibly if his surgeon has access or responsiveness to his implant's reliability. Tempo needs to have access to the reliability of his process or sensitivity to it himself, to employ the said process responsibly. His implant does not immediately integrate via strong ECI; it integrates via the weak route to epistemic integration. But it happens more swiftly than usual in the Tempo case, because his device was fit at birth.

---

<sup>59</sup>On a side note, consider Pritchard (2009a)'s Jenny case, inspired by Lackey (2009), that looks very similar to that of Tempo's, i.e. exhibits some features of immediate integration. Jenny looks at a stranger for directions in a new city and asks where the Holyrood Palace is. He tells her where it is and she, most likely, acquires knowledge from his testimony. Before Jenny inquired about the Holyrood Palace, she evaluated her informant and figured he was a decent local who would know about the city. At first glance, it may look like Jenny has immediately integrated her informant's process. Although, firstly, the Jenny case is not a case of extended cognition or even any other case of reliance on an external process. There is in fact no new integration taking place in this case. Jenny is relying on a faculty that she has already integrated, i.e. her ability to trust testimony. It does not matter if she lacks responsiveness to the reliability of her informant's process because it is not his process that she is recruiting to form a belief; she is recruiting a faculty from her own cognitive character, which integrated over time. Again, what may look like immediate epistemic integration is actually the agent's integrated faculty of trusting testimony.

## The differences in integrations

This section aims to explain why Tempo's device integrates differently from Tempo\*'s implant when they both undergo the same weak ECI process. In the course of my comparison, I will also discuss how deep functional integration is a milestone on the way to weak epistemic cognitive integration.

I unpacked in an earlier section that deep functional integration requires a certain degree of complementarity and complexity with other processes of the agent's cognitive system, such that the two integrated processes can function as a whole. Weak ECI also requires something similar. Cooperation and interaction are the fundamental units of both these integrations. The difference between deep functional integration and weak ECI is that the latter demands cooperation and interaction of the target mechanism with the agent's cognitive system, such that the agent becomes indirectly sensitive to the reliability of the target mechanism. Whereas, deep functional integration requires a degree of integration that makes the target process one with the agent's cognitive process. In this section, these cooperations and interactions will be discussed in the order they are achieved; deep functional integration of a process occurs first and then weak ECI of the said process follows.<sup>60</sup>

The Tempo and Tempo\* cases are not just cases of epistemic integration; they are cases of metaphysical integration as well. Not only are those implants integrating to form cognitive abilities for our two agents, but they are trusting, relying on, and carrying around with them a new cognitive process that is an extension of their original, neural cognitive make-up. Simply put, Tempo and Tempo\* are cases of extended cognition as well as cases of extended cognitive abilities. Their implants integrate and develop into knowledge-conducive reliable belief-forming processes, but first, they undergo deep functional integration and become metaphysically integrated into Tempo and Tempo\*'s cognitive systems.

Recall some important features of deep functional integration. It can be observed between processes within a cognitive system and also between processes in a cognitive system and extraorganismic vehicles that they couple

---

<sup>60</sup>It is important to note that not all processes that undergo deep functional integration also undergo weak ECI. That is, not all processes that qualify as cognitive are also epistemic. This idea has been discussed at length in the previous chapter.

with. Deep functional integration requires a certain 'degree and complexity of functional complementarity and integration' (Greenwood 2013, 425). On this degree of interconnectedness, the two processes come together to affect other processes and produce output as one synchronous unit. Furthermore, functional integration "has to be triggered initially and then developed through processes in which functional complementarity, complexity, and integration are progressively increased" (Greenwood 2013, 425). For instance, a toddler touching something really hot might trigger her respective processes to interact with other relevant processes, which may give her the ability to make sense of hot and cold.

Consider how a process fitted at birth undergoes MCI first and then how weak ECI follows. Say, Tempo touched the hot pot as a toddler and the external stimulus he received from the hot pot triggered his non-neural device to integrate with other innate processes in his system. Meanwhile, just like Tempo's non-neural device, his neural cognitive faculties will also be receiving initial triggers from either extraorganismic or intraorganismic factors and functionally integrating with each other. These early non-neural and neural processes will yield output and based on these output integrational complexity and complementarity will progress. These interconnections will help form Tempo's initial cognitive system, and therefore a developmental cognitive agency. The degree of interconnectedness of Tempo's implant will be proportional to the frequency and reciprocity of its use in the developmental stage. What is of import here is that his implant will be interconnecting and, therefore, integrating at the same time as his other neural processes and therefore, the degree of connections that the implant makes with his cognitive system will be similar to the amount that his innate processes make with his cognitive system. So, if there is a certain degree of interconnectedness that needs to be achieved for deep functional integration, Tempo's implant and his innate processes are both working towards that degree together. In other words, his implant and his neural faculties are all integrating at the same time to form his initial cognitive system; they have all been promoted to the status of *cognitive processes* together. So, now, Tempo's cognitive architecture has cognitive processes. Tempo's implant develops into an epistemic process, from a cognitive process, when Tempo becomes sensitive to its reliability.

Things will occur differently in Tempo\*, whose implant was added in his brain at a later stage in life. Tempo\* has an already integrated system and a mature cognitive character. His new implant, that was inserted in his brain unbeknownst to him, will have to integrate into a system that has already formed complex integration patterns. Unlike Tempo, Tempo\* device will have to integrate with an already integrated cognitive system. His device will first have to develop deep functional integration with his cognitive system. He will have to use his device such that it can acquire a certain degree of functional complementarity and complexity of integration. Notice how this is in line with Pritchard's claim about how Tempo\*'s device *epistemically* integrates into his cognitive system, over time. For weak ECI, Tempo\*'s implant will be employed habitually for a longer period, so it can become a stable disposition of his cognitive character. This will require cooperative interactions between the new resource, the process of employing the implant, and the other processes in the agent's cognitive system.

I will further support my argument in terms of this puzzle analogy that compares integration in a developing cognitive character with the integration that develops in an already mature cognitive character. Tempo's initial cognitive system, soon after birth, is like one of the first pieces of 5000-piece puzzle you pick. Let's say it is the 4th piece to be exact. The way you start the puzzle is by joining the first few pieces together, along with the 4th one and that sets the foundation for the rest of the pieces. The 4th piece, which is analogous to Tempo's implant, integrates with the other few pieces that make the foundation. The implant reaches a higher degree of integration very quickly and becomes a part of the agent's cognitive system. It becomes a cognitive process of the agent's cognitive system. Because the implant is like the 4th piece of the puzzle, it reaches a high degree of integration instantly. In other words, the foundation of the puzzle is not very demanding for the 4th puzzle piece like it is for the 1000th piece. The 4th piece only needs to make a few connections with the other foundational pieces to integrate into the foundation.

On the other hand, Tempo\*s device is like the 3500th piece of the puzzle. This piece has to fit into the spot where it has to make connections with twelve other pieces that are already positioned properly, so it has to fit right with all

ten of them to integrate into the puzzle. Clues from the semi-formed picture help the fitting of the 3500th piece, and in a way, the role of the puzzle is a little more substantial than it was when we were feeding in the 4th piece of the puzzle. The 3500th piece will need to achieve a degree of integration similar to the ones before it. In this case, the 3499 pieces of the puzzle that are already in place are analogous to our cognitive agency. This is why integrating and becoming Tempo\*'s cognitive process is more time consuming and requires more interconnections. His cognitive agency is already well-connected. A new process will need to match the degree of integration (or cooperative interaction) that processes already present in his cognitive system have amongst each other.

This analogy illustrates how weak ECI after birth can be faster than weak ECI at a later stage in life. But, does it accommodate how Clark's new technology will immediately integrate into an already mature cognitive character? It does not.

### Where do we stand on the univocal view?

As discussed in the previous chapters, the univocal view is the idea that ECI and MCI conditions partially or fully overlap. The stronger symmetric version is the idea that there is symmetric entailment between the two integrations, and the weaker asymmetric entailment suggests that epistemic integration is a sum of metaphysical integration and some other conditions. Clark's initial claim about technology that can immediately integrate epistemically strongly supports a univocal view, especially a symmetric entailment between the two integrations. The integration he has in mind is immediate and in one fell swoop, it makes an external process a cognitive process of the agents (MCI) and a knowledge-conducive process (ECI) as well. This means that Clark supports an understanding of metaphysical integration that is the same as (or very slightly different to) epistemic integration.

Pritchard's weak ECI can explain a somewhat fast epistemic integration but only in cases where the implantation has occurred at birth. The Tempo case precisely shows that a process can swiftly integrate into one's cognitive character when the character is still in its developing stages. But, when the resource is

encountered later in life, it takes more time to epistemically integrate into a developed cognitive character. This is best understood in terms of how a lower degree of interconnectedness is easily achieved when one's cognitive character is still developing, but for a new process to integrate into a mature cognitive character requires a higher degree of complementarity and interconnectedness.

The puzzle analogy, I discussed above, and the discussion leading up to it indicate that cooperation between processes is the basis of both metaphysical and epistemic integration, but the requirement for the latter may be more demanding than the former. That is, epistemic integration may require a higher degree of cooperation than metaphysical cognitive integration. MCI requires cooperation (between the target process and other faculties in one's cognitive system) sufficient for the target process to become one unit with the agent's cognitive process. Whereas, ECI requires interconnectedness sufficient for the target process to develop into the agent's cognitive process and also enough for the agent to be in a place where she can sense the failing reliability of her target process. Even in the Tempo case (almost immediate weak ECI) and Tempo\* case (slow weak ECI), epistemic integration follows metaphysical integration. First, it seems, that deep functional integration develops and then the interconnectedness of the new process with other faculties is such that it can be employed as an epistemic process.

This picture points towards an asymmetric entailment between the two integrations. It shows that instead of the symmetry between the two integrations, that Clark's immediate integration hints at, the two integrations have an asymmetric entailment. MCI is required to achieve epistemic integration, but it is not sufficient to meet epistemic integration. Conditions for epistemic integration are more demanding because the agent has to gain sensitivity to the reliability of her process to be able to responsibly (which is necessary for a knowledge-conducive process) employ her process.

## Conclusion

The discussion in this chapter revolves around Clark's claim that advanced technology will soon be able to immediately integrate, metaphysically and

epistemically, in our cognitive systems. I analyse this claim as someone who supports weak ECI as a fairly good explanation of how extended cognitive processes can also be epistemic processes that generate knowledge. Weak ECI cannot provide an understanding of immediate epistemic integration. The most immediate form of weak ECI can occur when a device is implanted in the agent's brain at birth. But, even this formulation is not the same as Clark's concept of immediate epistemic integration. Hence, weak ECI does not explain Clark's immediately integrating advanced technologies.

The other important finding of this chapter is that weak ECI, even in its most immediate form (i.e. right after birth), still does not occur as instantly as metaphysical integration, which makes the symmetric univocal view hard to defend. Clark's immediate integration suggests a strong symmetric entailment between the two integrations. I elaborate how deep functional integration develops and weak ECI follows in the Tempo and Tempo\* cases to show that a symmetric entailment between the two cases does not exist, but an asymmetric view may be possible.

I began the chapter with the argument that metaphysical integration cannot be understood only in terms of the glue and trust conditions and the CRC requirement (3+1 model). This model cannot appropriately describe Clark's immediately integrating processes. It is apparent from this disparity, that Clark's concept of the weakest form of epistemic integration is different from what Pritchard and Palermos have in mind. And, the weakest possible understanding of epistemic integration is what we need to make sense of extended epistemic processes. We have already unpacked and discussed the weak ECI account; it is time to explore the subpersonal ECI account. The debate from this chapter will carry on into the next chapter, where I aim to evaluate Clark's subpersonal MCI and ECI account based on the predictive brain approach. Although Clark's claim about immediate integration makes sense when understood in light of his subpersonal approach to epistemic integration, such an understanding falls prey to a range of problems that will be discussed in the upcoming chapters. The problem mainly lies in how a subpersonal epistemology fails to meet the responsibility requirement on knowledge.

## Chapter 4 Subpersonal epistemic integration

This chapter introduces and elaborates the foundation, i.e. the predictive brain model, of Clark's subpersonal epistemic integration account. The predictive processing account establishes Clark's understanding of perception and cognition, and how they come together. It gives insight into what metaphysical integration means on this account; how an external mechanism can integrate into one's cognitive system and extend it. This, we will see, is a wholly subpersonal affair. Building on this metaphysical cognitive integration (MCI) account, Clark's subpersonal epistemic integration will be presented as the mechanism of precision estimation. According to the process of precision estimation, already integrated processes come together to assign reliability weightings to new processes. On the basis of these weightings, new processes can become cognitive abilities of the agent and are able to provide knowledge.

This chapter also includes a comparison between Pritchard's weak epistemic integration and Clark's subpersonal integration. I point out that these two accounts share an important principle, like the cooperation and interaction of the cognitive system. However, there are more dissimilarities between the two. A crucial difference is their treatment of the connection between what goes on in the agent's cognitive processes and how it is attributed to the agent.

There are two main arguments in this chapter that cast doubt on an entirely subpersonal epistemology. The first is the ubiquitousness of Clark's subpersonal epistemic integration as presented by the predictive processing literature. According to the view, the precision estimation function occurs at all levels of the brain's functional hierarchy. Some of these levels are only concerned with states that are not consciously accessible to us and other levels deal with the mental states that we can be aware of. I argue that states that are beyond conscious accessibility should not fall in the epistemic domain.

The second problem I present for Clark's epistemic integration account is that it is wedded to only one possible personal level scenario. Since on Clark's view, epistemic integration can be fully subpersonal, it should not matter what happens at the personal level; but, it does matter. There is only one personal picture to his subpersonal epistemic integration which makes his account the

very same as Pritchard and Palermos' weak epistemic integration. He will need to separate his view from that one personal level scenario to make the argument that there is no personal level agency involved in his most minimal interpretation of epistemic integration.

## Predictive processing

Subpersonal metaphysical cognitive integration is supported by the Predictive Brain (Clark 2013, 2015, 2016b, 2017; Hohwy 2013, 2014; Wiese and Metzinger 2017), which is best understood in terms of how our perceptive faculties work.<sup>61</sup> The traditional feedforward model of perception claims that we perceive things from the external world and then make sense of it in our brains. Predictive Processing (PP) argues that the feedforward model is actually incorrect and our brains make predictions about the sensory data even before it is received. Clark's explains this in the following way:

... the idea here is to perform a kind of 'Bayesian flip' upon the standard (passive, feedforward) image of sensory processing. Instead of trying to build a model of what's out there on the basis of a panoply of low-level sensory cues, these models aim, in effect, to predict the current suite of low-level sensory cues from their best models of what's likely to be out there. (2015, 3766)

This means that perception actually involves perceiving from within the brain as well as from the external world. To make more sense of this quote, we need a general layout of what the structure of cognitive processing is like.

Cognitive processes have different hierarchical levels. Take our perceptual faculty for example. At the lowest level, there is sensorimotor exchange or the transmission of sensory data from sense organs. On intermediary levels, the data that is brought in is assembled into information about object recognition, shape

---

<sup>61</sup>It is interesting to note that work on predictive brain has also been used to argue against the extended cognition theory, and therefore extended knowledge. See, Hohwy (2014) for a discussion on predictive brain that entails that there is a strict boundary between the mind and the environment.

detection, colour classification, etc. And, at higher levels, the intermediary data come together so we can process mental states, form beliefs, feel emotions, etc.

Take an example of an agent looking at a car. At the very low levels of brain function, receptor neurons receive and carry signals to the relevant processes in our brains. For instance, electrical signals from our retinal receptors receive and transmit information on blue wavelengths and some other neurons transmit signals on red wavelengths. Both kinds of electrical impulses, then, come together in neurons that carry signals for both these wavelengths. Accordingly, the integratedness of the processes at the lowest level have formed information that will be required for higher levels. This kind of integratedness between processes occurs at different levels of functions. Take, for example, colour classification, shape detection, etc., that are somewhat intermediary level functions of the brain. The information about wavelengths becomes available to the intermediate level which assimilates it to form new information about the size, shape and colour of the object. At even higher levels, this information will be processed by the integrated processes to form mental states. For example, the object in front of me is a car.

PP claims that the levels on top make predictions about the output that is received from levels below them, like the higher levels will predict the expected output of the intermediary levels, with the help of 'prior expectations' or prior information present to them. These expectations come from 'a unified body of acquired knowledge' (Clark 2015, 3766). And these predictions are supposed to match the actual information of perception at these different levels. But when the match is not completely accurate it generates the difference between the expected information and what is actually perceived. This difference is called 'prediction error.' The prediction error value is communicated by the lower levels to the higher levels so future predictions can be improved accordingly. What you have in the end is an interplay between these top-down (predictions making) and bottom-up (prediction error relaying) chains, which is what helps us make sense of what we perceive (Clark 2015).<sup>62</sup>

---

<sup>62</sup>The predictive brain model is an eminent advancement in philosophy of mind and cognitive science. It provides a very good account of perception and cognition, and how they come together. The issue that I want to address does not have to do directly with the predictive

Recall the dress that divided the internet between those who called it blue and black and others who called it white and gold. Neuroscientists suggest that the split points towards how we are different in the way we predict illumination cues. What we expect is based on our prior information, which is different for different people. What happened with the dress was that the people who expected to see the dress in daylight saw it as white and gold, and those who expected to see it in artificial light thought it was blue and black (Wallisch 2017; Witzel, Racey, and O'Regan 2017). This explanation suggests that perception is a result of a complex interplay between what happens inside the brain and what happens outside the brain.

Of course, if perception was only a result of prediction then we would not get a very accurate picture of the world. Clark states, 'This is a potentially dangerous process. It would not do to always see what we expect, even if the use of top-down expectations is essential for dealing with noise and uncertainty. This is where the second core feature comes into play' (Clark 2015, 3767). An important second-order feature of PP, called 'precision estimation,' assigns reliability weightings to these predictions made at different levels and the sensory input present (Clark 2015, 3766). The operation involves assigning reliability to prediction errors and to see if some need to be ignored sometimes in accordance with the sensory environmental cues present, or if sensory input needs to be assigned reliability according to prior expectations. Think of it like this. A part (subpersonal) of our cognitive agency, in accordance with the context, either assigns reliability to data from outside (environmental cues) or data from prior knowledge. Clark gives us an example of an agent driving on a foggy day. Her precision estimation units will assign less reliability to visual cues (like signboards), and instead, her brain will rely more on previous knowledge of the area. On a sunny day, her precision estimation units will assign more reliability to visual cues, instead of prior expectations (Clark 2015; Carter, Clark, and Palermos 2018).

It is important to note that the precision estimation mechanism works at all the different levels of the brain's functional hierarchy, i.e. lower to 

---

brain account. My concern is Clark's claim that the predictive brain provides the weakest epistemic integration account.

higher levels. Its job is to mediate the interaction between the top-down and bottom-up chains. Recall this interaction: the higher levels constantly predict the output of the lower levels, error signals are generated that are a difference between what was predicted and the sensory data that was actually received, and these error signals are communicated from the lower levels to the higher levels. Meanwhile, at all these levels, these error signals are assigned reliability weightings by the precision estimation units, that determine if the agent will rely on a specific process or not. So, precision estimation is the agency's operation that determines if a certain process should be employed or not.

The PP model, that I have just discussed, tells us how cognitive agencies employ and integrate cognitive processes. It is the interplay between the predicting and error-communicating chains, and so a multilevel operation, that engages a cognitive process. Now that we have an idea of what cognition looks like we can use this model to identify cognition outside our skin and skull boundaries.

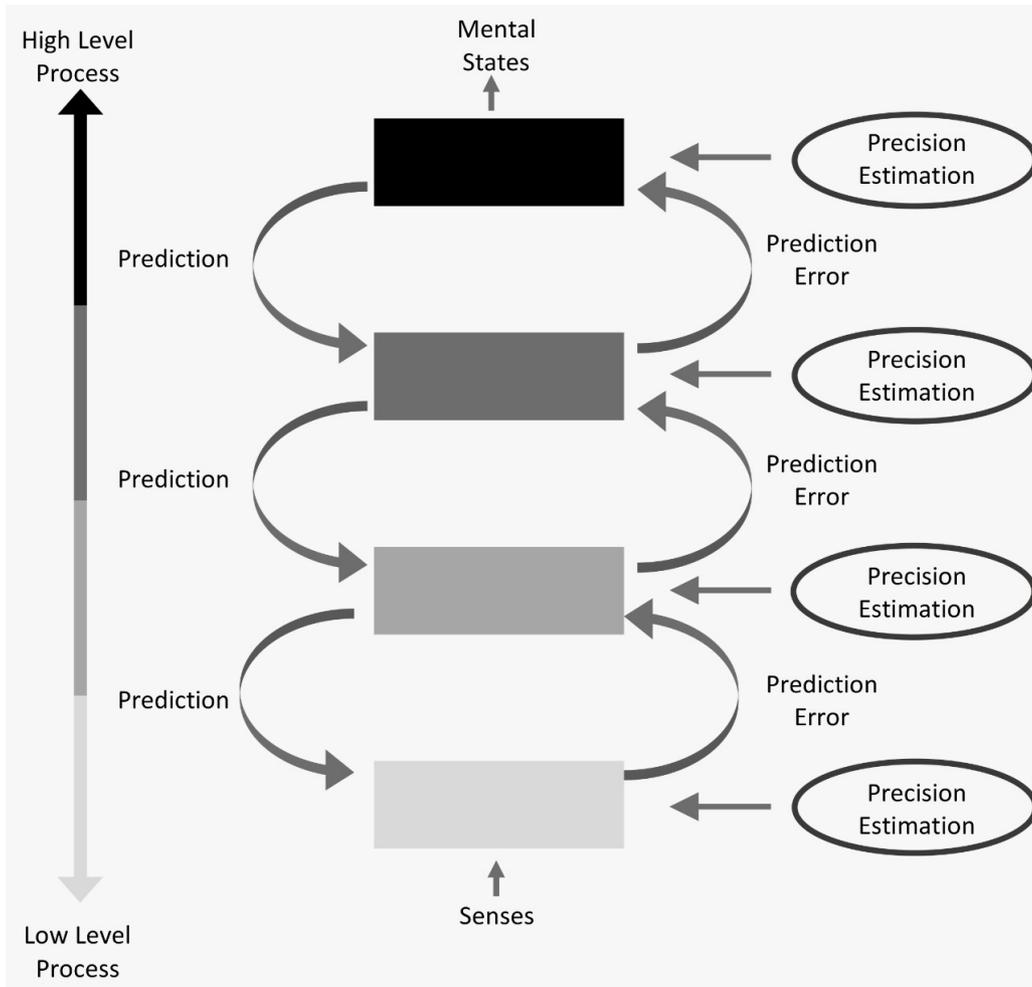


Figure 1: Predictive Processing.png

## Weak and subpersonal epistemic integration

The predictive brain model is not simply an account of metaphysical integration for Clark, it is also the most minimal understanding of epistemic integration. In other words, with the predictive processing account, Clark claims to have explained how a new process becomes a part of our cognitive system, and, also, how it can be employed as a cognitive ability (i.e. the integrated process that is able to provide knowledge) of our cognitive system.

The reason we are looking for a minimal epistemic integration account is that it will successfully respond to the epistemic hygiene dilemma (discussed in Chapter 2). According to this problem, the version of epistemic integration (weak epistemic integration that Pritchard and Palermos' support) that is considered a snug fit with the extended cognition hypothesis is still too active. For it to be compatible with the extended cognition thesis, epistemic integration needs to be weaker. Only an even weaker interpretation of epistemic integration can help us make sense of resources that can extend our cognition and also become knowledge-conducive belief-forming processes of our cognitive system.

Clark's proposal is of an epistemic integration account that is even weaker than weak epistemic integration. And, according to Clark, it is the epistemology that is in line with extended cognition. In this section, the discourse will help identify features of predictive processing that motivate Clark's subpersonal epistemic integration. Also, I will compare and contrast the subpersonal integration approach with the weak epistemic integration account.<sup>63</sup>

Of his predictive processing account, it is the second-order mechanism of precision estimation (described above) that describes the most minimal interpretation of epistemic integration. Precision estimation units assign reliability to a process, one that may exist outside our skin and skull boundary, on the basis of which the cognitive system employs them.<sup>64</sup> In Clark's words, the

---

<sup>63</sup>Clark's epistemic integration proposal, according to him, is not necessarily a subpersonal virtue epistemology. He writes that it may even be a sophisticated form of process reliabilism (2015, 3770). Since Clark does not go into detail about the process reliabilism account and has an entire section titled 'subpersonal virtue epistemology,' I will mostly treat his proposal as a kind of virtue epistemology.

<sup>64</sup>Take into account that Clark's argument is not that this is the only understanding of epis-

process of precision estimation ought to be understood in this way:

... the use of that knowledge is subject to a constant kind of second-order assessment (known as 'precision estimation') that determines the weighting assigned to specific predictions at all levels of processing, and to different aspects of the incoming sensory signal. These weightings reflect the varying reliability, in context, of differing aspects of the generative model and of the sensory inputs currently available. It is this second feature that (I hope to argue) is suggestive of an important species of sub-personal epistemic virtue (Clark 2015, 3769).

This quote describes precision estimation as the *assignment of reliability* to a new process, by a group of already integrated cognitive processes working and interacting together. What is this assignment of reliability based on? In order to answer this question, we will have to recall how Pritchard and Palermos' weak epistemic integration works.

Weak epistemic cognitive integration (or ECI), that Pritchard and Palermos favour as compatible with extended cognition, is not very different from Clark's proposal. Weak ECI states that the agent ought to be (indirectly) sensitive to the reliability of her process in order to take responsibility for it, only then can the agent be in the position to acquire knowledge from her process. It does not require that the agent ought to be aware of her process's reliability or should have any sort of reflective access to it. However, even if in a very weak sense, the agent has to monitor (so to speak) her process's reliability. How is the reliability of one's process monitored if one does not need reflective access to it?

Proponents of the weak epistemic integration account build on John Greco's work to describe how an agent can monitor her process's reliability in the background. On his view, a target process ought to form cooperative

---

temic integration but that it is the weakest interpretation of epistemic integration. In fact, he supports Pritchard and Palermos' weak epistemic integration; he just does not consider it the weakest account of epistemic integration. More importantly, Clark does not take the latter to be consistent or compatible with cases of extended cognition.

interactions with other processes in the agent's cognitive system, like the interactions that are already present amongst the agent's innate processes. These interactions will help the agent monitor the reliability of her process in the background. If something goes wrong with the reliability of the target process, the problematic output from the said process will be detected by another process in the agent's interconnected system. The interconnectedness of the cognitive system, basically, monitors the reliability of its components.

Clark's subpersonal epistemic integration is based on similar principles of cooperative interaction and integratedness of the cognitive system. The difference is that the interconnectedness of Clark's predictive processing cognitive system assigns a reliability weighting to the target process and, on its basis, includes the said process into the cognitive system. On the weak epistemic integration account, the target process has reliability that the agent's cognitive system gradually begins to monitor. Consequently, if there is evidence available against the reliability of the process, the agent becomes aware of the misfit. On the subpersonal epistemic integration account, the agent is not in a position to become aware of anything; the entire process occurs subpersonally. The agent's cognitive system, as per a subpersonal operation, assigns reliability to the new process, based on which the process is either employed or ignored.

Recall the immediate integration debate that I explored in the previous chapter. From the description of the precision estimation operation, it is evident why immediate integration is possible. The subpersonal operation of one's integrated cognitive system only has to assign reliability to the new process on the basis of information present and sensory cues available; this can be done immediately. Contrastingly, weak epistemic integration is a gradual process. The agent has to develop indirect sensitivity to the reliability of her process, which comes from employing and assimilating the products of the target process over a period of time.

Another difference between these two integrations is how they connect cognitive processes in the agent's cognitive system to the personal level (i.e. belonging to the whole person) of the agent. The information exchanges that go on in parts of our cognitive system, like our cognitive processes, are discussed subpersonally (i.e. belonging to only parts of the person). How are

these processes connected to our person and why is it important to make such a connection? Weak epistemic integration overtly describes this link and makes it an important part of the integration account. Accordingly, if a process is integrated into one's cognitive character via weak ECI, the agent will be in the position to become aware of any problem with the reliability of the process. Hence, the agent is always in the position to become aware of the reliability of her processes, and this connects the mechanisms that operate on a subpersonal level to her person. She can take responsibility for the processes because she can monitor them, albeit in an indirect way.

On the other hand, subpersonal ECI presupposes the link between what goes on at a subpersonal level and the agent's personal level. It presumes that something produced at a personal level belongs to the whole person, without the need for an explicit description of how the subpersonal and the personal are linked. What they take for granted is that something being produced by a part of my mind obviously belongs to me because it is my mind that we are dealing with. Here is where the problem lies. For proponents of extended cognition, it is not very simple to demarcate where the mind lies. Think of an agent who has undergone 20 surgeries and now all her important innate faculties have been replaced by artificial high-tech gadgets that work the same way. Or, consider an agent whose innate faculties do not work properly so she has replaced them all with resources in her environment that she employs similar to how she employed her innate processes. Would you now think that the output from the subpersonal operations of these agents' are obviously their own? No, you are likely to demand more information about how the said agents employ these artificial and external processes, and what connects their output to the agent. You will want to know what makes the product of these external resources their product?

Building onto this critique, let's discuss how the main purpose of epistemic integration is to describe how the agent responsibly employs her process. And, how the subpersonal epistemic integration compares to weak ECI on this front. As I discussed earlier in the thesis, virtue epistemologists support the idea that reliability alone is not sufficient for a belief-forming process to be knowledge-conducive. It is also necessary that we employ our reliable

processes responsibly. This responsible employment of a reliable process is what Greco has described in terms of weak ECI. The epistemic agent responsibly monitors (in the background) the reliability of her process, on the basis of the cooperation and interaction in her cognitive character. In case of any difficulty, or counterevidence to the reliability of the agent's process, she becomes aware of the issue and deals with it reflectively. This is the complete story of a responsible epistemic agent, on the weak ECI account.

Does Clark's epistemic integration view meet the responsibility requirement? It has to do that in order to qualify as a form of virtue epistemology. Let's say X is a reliable-belief forming process. Our precision estimation units will have to assign sufficient reliability weighting to X for it to be employed by our cognitive system. They will assign reliability weightings to X in accordance with the context or other subpersonal data present to them. If precision estimation units assign sufficient reliability to X, such that it is employed on the basis of that reliability, then it means that they have established a link to X's reliability.

However, the said link to the reliability of the agent's process is not personal. It does not connect the agent to process X's reliability in any way. It only connects the agent's other processes, subpersonally, to process X's reliability. And, what I said earlier about connections that stay on the subpersonal level, and are disconnected from the personal level, applies here as well. That is, it is crucial to elucidate how these subpersonal phenomena play out on the personal level. Without an explicit description of this sort, it is difficult to attribute the products of these subpersonal phenomena (that may be between parts of the agent's cognitive system and other resources in the environment) to the particular agent.

Another dissimilarity between the two integrations is concerned with the debate on the univocal view. The examination of the concept of immediate integration, in the previous chapter, led to the conclusion that Clark favours a symmetrical univocal view. In other words, for Clark, the conditions that a process needs to meet metaphysical integration can also be sufficient for epistemic integration and *vice versa*. Here, let's unpack his predictive processing model to substantiate this conclusion.

Generally, Clark's PP-model explains metaphysical cognitive integration as a multi-level interaction in a cognitive process. More finely, the interaction between these processes should be understood in terms of an interplay between the two hierarchical chains of prediction-making and prediction error-relaying. A new process can integrate if it can cooperate with other processes of one's cognitive system (that is targeted to a specific function), in the same way, that they cooperate and interact with each other. If this new process is external to our skin and skull boundary or if it is non-neural, it will extend our cognition. This is Clark's metaphysical cognitive integration account. His subpersonal epistemic integration is, essentially, the mechanism of precision estimation units which occurs at all the different levels of interaction between the two hierarchical chains. This means that the subpersonal epistemic integration is satisfied almost everywhere there is metaphysical integration. This clearly indicates a symmetrical entailment. If there is MCI, there is also subpersonal ECI.

In contrast, the weak ECI account points towards an asymmetric entailment. That is, metaphysical integration conditions are a subset of epistemic integration. This, as discussed in previous chapters, makes more sense because all our cognitive processes are not epistemic.

### Subpersonal level: doxastic and subdoxastic states

I will evaluate Clark's wholly subpersonal understanding of epistemic integration and provide reasons to reject such an account. But before that, this auxiliary section will give insight into some important concepts that will be useful later. I have been using the terms subpersonal and personal to mean belonging to a part of the person and belonging to the entire person. Hereon, I will give a more in-depth understanding of where these terms are from and how they are used.

Daniel Dennett first coined the term subpersonal to discern the two different levels of explanation: the personal and the subpersonal. One of the foremost things to note is that the personal and subpersonal are only levels of explanation or theory. Even though they are predicated with processes, states, mechanisms, etc., Dennett only differentiated the two in terms of explanatory levels (Dennett 1969, 90–96). Since Dennett, these terms have come to be used

in many ways and some of these ways have been confusing.

Zoe Drayson (2012) writes that the personal explanation is a way to explain something in terms of the whole person, whereas the subpersonal is a way to explain something in terms of part of the person. She explains that the personal and the subpersonal level of explanations are derived from the general horizontal and vertical explanations in psychology. A horizontal explanation is one that has as its explanandum (which is to be explained) a particular event, and the explanans (which does the explaining) are other events that causally lead up to the said event (Drayson 2012). For example, the breaking of a plant pot can be explained in terms of how my leg hit the table and made the plant pot fall. In vertical explanations, the explanandum are understood as *capacities* or *dispositions* of things and the explanans are the parts or components (Drayson 2012). For instance, the breaking of the pot is explained vertically when it is discussed in terms of the pot's chemical composition and the force by which it hit the floor. When the horizontal and vertical explanations are applied to everyday folk psychology, concerned with mental state ascriptions in humans, we get personal and subpersonal levels of explanation. Take, for instance, a young child who looks at a friend enjoying an ice-cream cone and desires one as well. What you have here is a personal explanation of why the child has this desire. It is because she saw her friend enjoy one. A personal level explanation is one that relates to the mental state of the entire person. Whereas, a subpersonal explanation is about the state of part (or component) of the person.

Subpersonal theories try to explain a phenomenon in terms of the *state* that a *part of the person* is in. Part of the person can mean anything from an atom or neuron inside the person's body to a complex cognitive system, with several integrated processes performing specific functions. Within a cognitive system, the states that parts of a person are in can be discussed in terms of psychological states, such as liking, predicting, rejecting, connecting, receiving, collecting, etc.<sup>65</sup> Subpersonal explanations do not only describe what goes on in

---

<sup>65</sup>In *Philosophical Foundations of Neuroscience*, Bennett and Hacker argue that attributing psychological states to a part of the animal is to commit the mereological fallacy (2003, 72). But, Drayson notes that many philosophers have allowed this comparison because of the

the lower levels of the brain's functional hierarchy, they can also describe higher levels. For instance, the state that our cognitive system is in when we form a mental state, like belief, can also be explained subpersonally. The explanation would be in terms of the states that a part of the person is in.

Subpersonal explanations are often confused with subdoxastic states. Subdoxastic states are ones that our cognitive systems, or a part of them, are in when they bring about the information that is not consciously accessible for us. Subpersonal explanations can actually be about both subdoxastic and doxastic states. When our cognitive system, or a part of it, is in a doxastic state it brings the about information that is or can be present to our consciousness. For example, when parts of our person (or cognitive system) are in a mental state, we are said to believe *p*. Mental states like these occur at higher levels of the brain's functional hierarchy. Subdoxastic states, on the other hand, occur at lower levels. For example, the neural state in which neurons assimilate information about wavelengths and illumination cues to discover (subpersonally) the shape of the object. Stich was the first to make this distinction between subdoxastic and doxastic states and in his own words, '[s]ubdoxastic states occur in a variety of separate, special purpose cognitive subsystems' and doxastic states 'form a consciously accessible, inferentially integrated cognitive subsystem' (Stich 1978, 508).

## Problematic ubiquitousness

Clark's subpersonal epistemic integration is explained in terms of the operation carried out by the precision estimation units, as I have said before. These units work at all levels of the brain's hierarchy, as demonstrated in Figure 1. At all levels, these units assign reliabilities to processes on the basis of which their output is assimilated further. At lower levels, precision estimation function results in the use of processes that produce the kind of information

---

similarity, our parts often do function or process information like we do as a whole (2012). Dennett writes, 'It is an empirical fact, and a surprising one, that our brains — more particularly, parts of our brains — engage in processes that are strikingly like guessing, deciding, believing, jumping to conclusions, etc. And it is enough like these personal level behaviours to warrant stretching ordinary usage to cover it' (Dennett 2007, 86).

that subdoxastic states produce. Let's call this subdoxastic information, that is consciously inaccessible to us, and includes information on wavelengths, illumination, hues etc. Similarly, precision estimation at intermediary levels precipitates in more complex subdoxastic states, that give rise to shape, size, colour detection, etc. At higher levels, precision estimation units assign reliability to processes that provide us with doxastic information. And so, their function culminates in the production of emotions, beliefs, etc. Oddly, a singular function, that of precision estimation, describes the formation of subdoxastic as well as doxastic information.

Of course, there is a reasonable explanation for why subdoxastic states are being produced at some levels, and doxastic states are being produced at others with the help of the same function, it has to do with the type of processes that precision estimation units are assigning reliabilities to. At lower levels, precision estimation units enable the recruitment of processes that result in information about, say, wavelengths, whereas at higher levels the processes that they employ are concerned with mental state formation, for example, belief-forming processes.

However, the problem lies in the method of explanation that Clark has used. If the function of precision estimation units is supposed to describe subpersonal epistemic integration, the most minimal understanding of epistemic integration, then at all these levels we should see minimal epistemic integration. His theory should, then, also explain why at some of the levels epistemic integration results in subdoxastic states, that are not present to our consciousness, and other levels it results in doxastic states, that are or can be present to our consciousness.

Bring in the notion of epistemic responsibility and see how this problem aggravates. Recall that any theory of epistemic integration is supposed to explain how the agent meets the epistemic responsibility requirement in employing a reliable belief-forming process. If Clark's subpersonal epistemic integration is a legitimate theory of epistemic integration then it should describe how the agent meets the responsibility requirement, for which it has to explain how our agency is connected to the reliability of our process. Subpersonal epistemic integration provides an understanding of such a link by describing

the mechanism of precision estimation units, i.e. assigning reliabilities to processes and then employing on the basis of these reliabilities. However, if this happens at all levels of the brain's functional hierarchy because precision estimation functions occur at all these levels, then the epistemic responsibility requirement is being met at all levels. This probably means that part of my body is responsible for how it is collecting and assimilating information about different wavelengths in UV light.

This issue can be understood in another way. Epistemic integration should only be relevant to belief-forming processes, which form mental states and are at the highest level in the brain's functional hierarchy. Epistemic integration elucidates how we can employ a reliable belief-forming process responsibly because only reliable processes that agents engage with responsibly can give them knowledge. Maybe Clark's subpersonal epistemic integration does give an account for how agents can invoke reliable belief-forming processes in a responsible manner. But, the problem is, Clark's subpersonal epistemic integration also connects the concept of epistemic integration to the other cognitive processes, especially the ones at lower levels, that are not concerned with the production of beliefs. It seems that the information that we do not have conscious access to, that emanates from subdoxastic states, can be held responsibly by an agent or a part of him. This is an odd way to talk about epistemic integration. His explanation makes epistemic integration more ubiquitous than it is.

Consider two neural chains: one that carries information about red wavelengths and the other that carries information about green wavelengths. They connect at a junction where information about wavelengths is assimilated. This is the kind of information that comes from subdoxastic states and so we cannot have conscious access to it. Clark's subpersonal epistemic integration, which takes place at all hierarchical levels, puts this kind of subdoxastic information in the epistemic responsibility domain. It becomes possible to claim that the agent responsibly carries or acquires subdoxastic information about wavelengths. It is simply odd to talk about consciously inaccessible information as something that we (or a part of us) can acquire responsibly.

## Are these two integrations different?

In order to discuss the second problem with an entirely subpersonal concept of epistemic integration, let's compare the two rival theories of integration we have in front of us. First, I will quickly reiterate what these theories claim and then move on to discuss them in the context of the Temp-style cases, adapted from Keith Lehrer's *Mr TrueTemp* (1990).

Pritchard and Palermos' hold that weak epistemic integration is maximally and sufficiently weak in agency-involvement and is congruous with the demands of cognitive extension. It requires that one's agency gain indirect responsiveness to the reliability of one's process, which allows the reliable process to be knowledge-conducive. Clark claims that such an account is still somewhat personal and therefore not in line with cognitive extension. Cognition is a subpersonal affair, and only a subpersonal epistemic integration can be compatible with it. This subpersonal epistemic integration should be understood in terms of precision estimation.

Pritchard's Temp case (that has been discussed at length in the previous chapter but from a different angle) is about an agent who forms temperature beliefs about the room he is in by looking at a thermometer in the room. The thermometer is actually broken and gives random readings. However, unbeknownst to Temp, a genie in his room adjusts the thermostat so the reading on the thermometer matches the temperature of the room. Hence, Temp forms true beliefs and the process that he employs is a reliable process (Pritchard 2010). Of course, Temp does not have knowledge of the temperature of the room because he is not responsive to the reliability of the process that he is employing, in fact he does not even know which process he is really employing. The truth of the beliefs have nothing to do with Temp's cognitive agency.

Temp can integrate this external process in his cognitive system, and as a result acquire knowledge from this process. One way to acquire this is to gain awareness of the reliability of the actual process. In other words, Temp should become aware of the actual process, i.e. the genie adjusting the thermostat, and also become aware that the process is reliable and how. This way, his process of using the thermometer-genie mechanism will integrate into his cognitive system.

But this is not the only way that integration can take place, and of course, this active kind of integration is antithetical to cognitive extension.

Now, let's suppose that the process this genie performs, of adjusting the thermostat to fit Temp's temperature reading, is coded on a microchip, and the chip is fitted in Temp's brain at birth. He will be able to use the chip the same way that he uses his other faculties, provided that Temp is operating in an environment where this is not an oddity. After all, we rely on our innate faculties that are designed to function in this environment, even though we have no reflective access to their reliability (Pritchard 2010). So, Pritchard's proposal here is that if the process was functioning in its environment and was fitted in Temp's brain from the off then it would be integrated into his agency. Active awareness of the reliability of a resource is not a necessary condition for integration. Integration can occur via a passive route as well, i.e. weak epistemic integration.

Furthermore, Pritchard suggests that there are other passive integration routes. Even if Temp's device was planted in his brain at a later stage in life, it could integrate with his agency without any access to the reliability of the process. Suppose, Temp had an accident which resulted in him being comatose, and during his surgery, his doctors implanted this device in him, unbeknownst to him. At first, Temp will not be able to acquire knowledge by using this device, because the process of his agency employing his device will not have immediately integrated at implantation. However, after using the device for a few years, Temp will develop indirect responsiveness to the reliability of his process. Which is to say, if the process malfunctioned Temp would become aware of the unreliability and stop employing it (Pritchard 2010). This indirect responsiveness that the agent (his whole person) can become aware of is obviously a personal level phenomenon.

Here is where things get interesting. Clark puts forward his own version of the Temp case, and calls in Temp2, to demonstrate how an agent can epistemically integrate this microchip in an even less active way. His interpretation gets rid of any agential involvement whatsoever. Temp2 is a predictive processing agent (PP agent). She, too, forms her temperature beliefs by reading a thermometer in the room, and she has been doing that for quite

some time now. Her thermometer is also broken, but a genie regulates the thermostat to fit her reading, and for this reason, her beliefs are always true. Evidently, her process is reliable. So far, things seem to be going quite the same way as Pritchard would describe them. The only difference is the personal-level element in Pritchard's account.

Clark adds a twist to the Temp2 case; he wants us to imagine that at some point Temp2's genie begins to take Wednesdays off. So from that point onwards, Temp2's output on Wednesdays is unreliable, but on all other days, it is perfectly reliable. And according to Clark, as a PP agent, Temp2 will subpersonally and immediately integrate the new process in which the genie takes Wednesdays off. Let's see what Temp2 goes through in Clark's own words:

Over time, sensory cues indicative of the Wednesday context (e.g. seeing the word Wednesday on the computer date-screen that she views every morning before starting work) come to be associated with unreliable information from the thermometer. Sub-personal precision-weighting mechanisms then cause Temp2, even when she looks at the thermometer, to fail to attend to what it says. In effect, she simply ignores the thermometer readings on Wednesdays (Clark 2015, 3769).

Notice, that one of the foremost things that separate the Temp2 account from Pritchard's weak epistemic integration is that Temp2 can immediately integrate a different process in her cognitive system. When the genie takes a break, Temp2 immediately integrates a different process, in which genie takes Wednesdays off. Weak epistemic integration usually takes time that allows the agent to develop responsiveness to the process's reliability. I will not focus on the immediate integration in this example; there are some other aspects of subpersonal epistemic integration that I need to bring to light. I am going to argue that this example complicates subpersonal epistemic integration instead of helping us grasp the concept better. But most importantly, this case fails to separate the personal and the subpersonal, indicating that a very thorough separation might not even be possible.

First of all, the description of what goes on in Temp2's mind is very

rushed. And secondly, some of the description is giving us personal explanation of what is going on and the rest is subpersonal explanation. Both these explanations are jumbled together, which makes it very hard to demarcate Clark's subpersonal epistemic integration. For instance, the quote starts with how integration will take place, then begins to describe what occurs at the subpersonal level, i.e. the precision estimation weighting assignment, and then it claims that at the personal-level Temp2 will learn to ignore Wednesday readings.

Furthermore, we are given no understanding of how things transpire on the first Wednesday that the genie takes off which is what should have been the most important day to focus on. Pritchard's weak epistemic integration would find this account significant for the explanation of integration. After all, not only has a new process begun to integrate, but Clark is downplaying the fact that an original process has just ceased to be reliable. The first process in which the genie takes no breaks has stopped being reliable, and another process in which the genie takes Wednesdays off has come forward to be integrated. What explanation can subpersonal epistemic integration give us for the first Wednesday the genie takes off. What exactly happens?

I am going to attempt to fill in the gaps of the Temp2 case, and see what happens on the subpersonal and personal level on the first Wednesday the genie takes off. Before the genie decided to take days off, Temp2's precision estimation units were assigning sufficient reliability to the agent's use of the genie-thermometer mechanism. On the basis of this reliability assignment, Temp2 was employing this process just like she would employ one of her own reliable and knowledge-conducive belief-forming mechanisms. But on one random Wednesday, the genie-thermometer process stopped working properly because the genie began to take Wednesdays off. So, on this day, another process came to be, one in which the genie started taking Wednesdays off. On the first Wednesday, Temp2's precision estimation units did not assign sufficient reliability to the process that had just stopped working, and therefore Temp2 did not employ the process. I do not have reservations against this subpersonal picture but I am concerned about its personal counterpart.

Clark's epistemic integration account is entirely subpersonal, which means that it only describes the phenomenon in terms of part of the person. It does

not describe what happens at the personal level, which means any number of plausible scenarios can occur at the personal level. Clark has only described the subpersonal explanation for us. Which personal scenario it corresponds to, we do not know. I am going to begin by considering two plausible personal level scenario for this event, but I will conclude that only one of them makes sense. Scenario 1 is that Temp2 will become aware that she stopped relying on her thermometer reading, and Scenario 2 is that she will not become aware of ignoring her usual process.

In terms of the precision estimation mechanism, if a process is reliable, it will be assigned sufficient weighting and will be employed (and integrated) on its basis. But, if the said process is not reliable, then the precision units will not assign it sufficient weighting and it will be ignored. On the first Wednesday, the genie process, that used to take no days off, has become unreliable. There are two ways that scenario 2 can materialise. Firstly, the genie will produce flawed output, Temp2's precision estimation units will assign it insufficient reliability. Her cognitive system will ignore the output from the process that it has been relying for a long time and all of this will not manifest in her occurring thought. On the personal level, Temp2 will simply not become aware that she ignored the process that she usually relies on. This consequence does not just sound unlikely, it seems impossible. If we have been relying on a process for a long time, we do not suddenly ignore it. We, at least, become aware of the fact that we are ignoring it, which will most likely lead us to *why* we are ignoring it. Eventually, we will deduce that we have ignored our formerly reliable process because of a sudden problem with its reliability. This interpretation of scenario 2, it seems, will necessarily culminate into scenario 1. Temp2 will become aware that she has stopped relying on the process that she was relying on for very long.

The second version of scenario 2 may be that Clark's PP agent will keep employing the unreliable process for a while before integrating a different process. After all, Clark states that the sensory cues will come to be associated with unreliable information *over time*. According to this account, Temp2's cognitive system will take some time before it figures out that it should not assign sufficient reliability to the genie process anymore. By the time it has figured out, Temp2 will go on to use an unreliable process for a while, and, with

it, she will produce many false beliefs.<sup>66</sup> If she does become aware that she is forming false beliefs, then this approach also culminates into scenario 1. And, if the agent does not become aware of forming false beliefs then there are other problems to deal with, like how this is not a very good epistemic integration account if it cannot maintain epistemic responsibility.

If Clark's epistemic integration account was genuinely and wholly subpersonal then it should not matter what happens at the personal level of the integration; a number of scenarios should be possible. Nonetheless, I have shown that the agent will become aware of her flawed process, one way or another. Hence, the only personal level scenario that makes sense is scenario 1. This is the same scenario that proponents of weak epistemic integration use to describe their account. That is, the consequence of successful integration means that if the reliability of the target process suffers for any reason, the agent would become aware of it.

If Pritchard's Temp case had the same problem as Clark's, and the genie had started to take Wednesdays off, Temp would become aware of the sudden unreliability. Weak epistemic integration states that the agent develops indirect responsiveness to the reliability of her process in virtue of cooperation and interaction amongst the processes of her cognitive system (Palermos 2014b). It is indirect in the sense that any counterevidence to the reliability allows the agent to become aware of the problem. And, maybe Clark's Temp2 will not become aware of the unreliability of her process, she will at least become aware

---

<sup>66</sup>Some might argue at this point that this is not a serious worry. After all, the processes that we have not yet integrated into our cognitive characters often give us beliefs that are not true. However, note that there is a significant difference between stumbling upon unreliable belief-forming processes when you are in the process of integrating them and keep employing an integrated faculty even when it is malfunctioning (of course, it might not remain integrated then). The former is not yet a part of our cognitive character and, therefore, not pernicious to epistemic responsibility. The latter indicates a serious threat to our concept of epistemic responsibility. Also, I want to clarify here that I do not take issue with the fact that Temp2 takes some time before it begins to assign the correct reliability weighting to her altered process. It may be that Temp2 employed her process but her mind was occupied elsewhere, and she was not conscious of the belief produced. Although, as soon as she becomes aware of the belief she produced, she should be able to detect the discrepancy. Otherwise, epistemic responsibility will be compromised.

of ignoring a process that she has been relying on for a while. In this way, then, their fate will be the same. They will become aware of the employment of their processes.

So, the upshot is that Clark's subpersonal epistemic integration is wedded to a personal scenario, which is the same scenario that weak epistemic integration defines as its personal scenario. That is, the agent becomes aware when the evidence against the reliability of the process comes forward, even if this evidence is only presented to the agent's subpersonal agency. In order to demonstrate that epistemic integration can be a wholly subpersonal phenomenon, Clark should demonstrate that any number of scenarios can be possible at the personal level of subpersonal epistemic integration. As long as we have only one personal level understanding, in which the agent gains awareness of the situation, we will find no differences between Clark's epistemic integration and Pritchard and Palermos' weak epistemic integration, except the fact that Clark's account simply does not explicitly link the subpersonal to the personal.

## Conclusion

I have argued in preceding chapters that a symmetric univocal view, i.e. one that understands epistemic integration and metaphysical cognitive integration in terms of the same conditions, can be problematic. Such a view gives the impression that all cognitive processes are epistemic processes. Clark, with his epistemic hygiene dilemma, hints at a similar symmetric univocal view that I have (finally) fully unpacked in this chapter. Here, I elucidate the cognitive science and philosophy of mind aspect of Clark's account. Before, his account was only discussed in terms of its epistemology, but this chapter fully develops the cognitive science grounds that his epistemology is based in. Clark's subpersonal epistemology is based in his predictive brain model. I have tried to give a full account of how the predictive processing mind works, and what metaphysical and epistemic integration looks like in such a mind. Furthermore, with the exposition of the predictive processing account, I have been able to describe why Clark's univocal view is a symmetric entailment. The reason lies in how Clark formulates his subpersonal epistemic integration (not a term he uses)

in terms of the precision estimation mechanism. The latter is a subpersonal operation that assigns reliability weightings to target processes, on the basis of which they are either employed and integrated or ignored.

After providing the framework for Clark's subpersonal integration, I argue that there are two serious problems that his account faces. The first problem is the ubiquitousness of his subpersonal epistemic integration. The way Clark explains the latter is in terms of precision estimation mechanisms that occur at every level of the brain's functional hierarchy. Some of these levels are concerned with the production of mental states, whereas others only deal with the processing of subdoxastic information that is not consciously accessible to us, like wavelength data, hues, etc. It would appear that the precision estimation units operate at all levels and assign reliability weightings to processes to bring them on-board. I argue that it is odd for an epistemic integration theory to explain the formation of doxastic as well as subdoxastic information. Subdoxastic information and the states that produce them are not in the epistemic dimension. If they are not consciously accessible to us, they are not something we can ever know. An epistemic integration account that applies to these states the same way it applies to doxastic states cannot be proper.

The second worry I bring forward is the similarity between the weak and subpersonal epistemic integration. Clark's account is supposed to be entirely subpersonal, and what happens at its personal level should not be significant for his theory. With discussion and examination of Clark's version of the Temp case, I propose that Clark's subpersonal is linked to only one possible scenario on the personal level, which is the same one that the weak ECI supporters describe. If Clark wants to make the point that epistemic integration can take place entirely subpersonally, he will need to separate his account from that one possible personal level. Because, for an account that is wholly subpersonal, any number of personal level scenarios should be possible. Eventually, I ask if his account is sufficiently different from the weak epistemic integration that Pritchard and Palermos favour as a snug fit for extended cognition. While I believe that it is sufficiently different, I emphasise that Clark's explanation has not made that specific point clear.

The one major difference between subpersonal ECI and weak ECI that I

highlight in this chapter is how they understand the connection between the personal and the subpersonal level. Weak epistemic integration overtly explicitly describes how our cognitive processes (what goes on at a subpersonal level) are linked to us. Subpersonal ECI takes this connection for granted. I suggest that proponents of extended cognition thesis should not take this connection for granted. They need to find a way to link what happens at the subpersonal level to our personal level so that beliefs and knowledge can be attributed and credited to the right sources. This is a theme that I will explore further in Chapter 6. The idea is to point out that if the subpersonal is not sufficiently linked (in terms of explanation) to the personal level, we will find it hard to understand which processes are producing knowledge that belongs to us. Adding to this problem, I will also establish, in Chapter 6, that a weaker understanding of the said link will make it hard to delineate the epistemic agent. Before that debate, however, I will examine (in Chapter 5) the notion of epistemic responsibility further, in terms of the defeasibility theory this time.

## Chapter 5 Understanding epistemic responsibility via defeaters

Epistemic integration accounts describe how reliable belief-forming processes can become a part of our cognitive characters so that we can recruit these processes as responsible epistemic agents. An important question is, what about epistemic integration allows the agent to responsibly recruit the said process? One way to answer this question is to look towards *defeasibility*. Defeasibility theory elucidates how an epistemic justification theory deals with defeaters, and in turn gives us an understanding of epistemic responsibility. Generally, defeaters are propositions that the agent may or may not be aware of that attack the truth of an agent's belief or the reliability of its source. Agents' responsible employment of their processes can be conceived in terms of what kind of defeaters are present against their beliefs and how they deal with them.

In this chapter, I first give an overview of the literature on defeaters. I bring to the fore different types of defeaters, some that are concerned with externalist justification, others that relate to internalist justification, etc. Once I have introduced several types of defeaters, I will explain Pritchard and Palermos' weak epistemic integration account in terms of a defeasibility theory. I will elaborate how two types of defeaters, understood together, describe how an agent can responsibly employ her process via weak epistemic integration. These two types of defeaters include (1) normative defeaters, i.e. propositions that the agent ought to be aware of because they seriously attack the truth of the agent's belief or its source, and (2) conscious mental state defeaters that the agent is aware of. I argue that epistemic integration should be understood in terms of no-defeater conditions for both these defeaters. This means, that if there are no normative defeaters and there are no conscious mental state defeaters against the reliability of a candidate process, the process can be considered epistemically integrated. And, the agent can employ such a process responsibly. The idea is that an epistemic agent should (first) use the target process frequently over a period of time so she can become aware of all the important things about the reliability of the said process. Once that has been

achieved, she can monitor the reliability of her process further. So, she should not become aware of any problems with the reliability of her process then it has successfully integrated through weak epistemic integration. Overall, the agent will be aware of all that she needs to be aware of, and if some new problem comes up, she will be in the position to become aware of it. Weak epistemic integration, hence, provides a dependable conception of epistemic responsibility.

Further, I contrast weak epistemic integration to Clark's subpersonal epistemic integration in terms of defeasibility. For this, I have to formulate a subpersonal defeasibility account (Clark does not provide one, but it can be extrapolated from his work). I propose two different subpersonal defeasibility accounts that rely on different kinds of defeaters, and I analyse both of them respectively. I find both of them unsatisfactory in capturing epistemic responsibility.

For the first account, I propose we understand subpersonal defeat in terms of reflective mental state defeaters. These are defeaters that are our mental states but we are not aware of them. My suggestion is that these defeaters may defeat at a subpersonal level. By looking at a few examples, however, I find that reflective defeaters defeat the same way as normative defeaters (ones we ought to be aware of). But, on their own, they cannot provide a complete understanding of epistemic responsibility. I conclude that if subpersonal defeat is understood in this way, it will not inform us about epistemic responsibility in any way.

My second account of subpersonal defeasibility introduces a new kind of defeater, a subpersonal defeater. These defeaters may be our mental states (not conscious to us) or they may be states that work on a level below mental states (subdoxastic states, like the ones that carry information about wavelengths of light, etc.). What is important about this idea is that these defeaters are present to our cognitive system, even if they are not present to our conscious selves. Nevertheless, examining un-integration in Clark's Temp2 case (also discussed in previous chapter) shows that this account can also not uphold epistemic responsibility. Clark's predictive processing agent is found irresponsibly recruiting her malfunctioning process.

In short, this chapter compares two plausible subpersonal defeasibility

theories with the one that supports weak epistemic integration and finds both of them wanting. It concludes that Clark's wholly subpersonal integration does not provide an appropriate understanding of epistemic responsibility.

## What are defeaters?

Epistemic defeaters may be true propositions that the agent is not aware of, or propositional attitudes and experiences that the agent is aware of, that reduce the justification status of one's belief. For example, my belief that I had breakfast this morning can be defeated if I discover that my bowl of cereal is still sitting, as it is, on my kitchen table. It casts doubt on my belief that I had breakfast this morning. Similarly, my belief that I saw a black cat passing by can be defeated by the true proposition that the cat was brown, even if I am not aware of the said proposition.

A theory of epistemic defeat can be useful because it provides insight into the concept of epistemic justification (Bergmann 2006, 2005). For, if there is a defeater present for your belief then your justification for this belief has been called into question. What is required for epistemic justification is that there are no defeaters attacking your belief. Defeaters point towards a problem in your epistemic justification for B, and therefore they warrant withholding B, looking for more support for B, or outright dismissing B.

Epistemic justification can be internalist, the kind that agents have access to, or it can be externalist, the kind that agents do not have access to. And both of these can be explained in terms of defeasibility theories.

An example of externalist defeat<sup>67</sup> is that there are true propositions that are not present to our consciousness, but can defeat our belief. This sort of defeat indicates that our belief or its source lacks an externalist property like proper function or reliability. For example, I go to my ophthalmologist for an eye exam and she asks me to read a whiteboard with small letters on it. I read the last sentence from the left and believe I am looking at the letter F. The true proposition (I am not aware of) that the letter I am looking at is a P defeats my

---

<sup>67</sup>The reason I use the term externalism with defeat (following Bergmann) here is to refer to defeaters that are related to externalist justification. There is nothing else I mean by externalist defeat.

belief. This true proposition is a *Propositional Defeater* to my belief that I am looking at the letter F (Bergmann 2006). It also indicates that my eyesight does not work properly or is unreliable when it comes to reading small writing that is far away from me. This kind of defeat, therefore, characterises an externalist property like reliability. The lack of such defeaters is an indication that the source of the belief is reliable.<sup>68</sup>

There is also an internalist understanding of defeat<sup>69</sup>, according to which defeaters are the agent's conscious mental states. Let's call these *Conscious Mental State Defeaters*. So, for instance, after reading the last line wrong, I get up from my seat and walk closer to the board and discover that the letter that I believed was an F is actually P. I am now aware of a mental state that I was looking at the letter P all along, and this new mental state defeats my original belief. This internalist concept of defeat says something about my epistemic responsibility as an agent. Now that I am aware that one of my mental states defeats another belief I have, it will be irresponsible of me to hold the defeated mental state on the basis of its original justification status. Epistemic responsibility demands that I renounce or double-check my defeated belief.

Defeaters are especially useful for externalist theories of justifications, such as proper function theories and reliabilism. Externalism about justification, as I briefly mentioned above, is the view that the quality that justifies a belief may not be accessible to the agent. So, you may not have reasons in support of your belief, but there may be an externalist quality about your belief that makes it justified for you to form the said belief. Reliabilist externalist theories of justification propose that this externalist quality is the reliability of a belief-forming process, i.e. its frequency to produce true beliefs rather than false. A reliable belief-forming process, therefore, justifies belief even if the agent does not have access to the reliability of her process.<sup>70</sup>

---

<sup>68</sup>However, note that this does not ascertain that the belief cannot be false. The belief may still be sometimes false even if the source of the belief is reliable.

<sup>69</sup>This only means that these defeaters are related to internalist justification in some way.

<sup>70</sup>Recall the Norman case from Chapter 1. Norman has a reliable process, clairvoyance, that helps him form true beliefs about the whereabouts of the president. He has no connection to the reliability of this process, however. Epistemologists mostly agree that Norman's true

The problem with these externalist theories is that they lack a kind of subjective justification or a link to the reliability of the process.<sup>71</sup> Virtue reliabilism states that an agent should responsibly employ her reliable process in pursuit of knowledge. Before this chapter, I have tried to capture the agent's link to the reliability of her process in terms of epistemic cognitive integration, which also elucidates how an agent can responsibly employ a reliable process. In this chapter, I will expand on this understanding in terms of the defeasibility theory. It will become clear how an internalist sort of defeat as well as an externalist defeat can help inform epistemic responsibility. Simply put, the agent ought to be aware of the necessary defeaters so she can make adjustments, and her new belief is not defeated by any other mental state she holds. In order to unpack this notion and connect it to epistemic responsibility as well as the epistemic integration accounts, I need to introduce a few other kinds of defeaters.

## Types of defeaters

In the way they defeat, defeaters can be *overriding* and *undercutting* (Pritchard 2018b). An overriding defeater provides counterevidence to the truth of our belief, which means if an overriding defeater is present we cannot justifiably hold our belief on the same basis as we did before. For instance, Saima solved a math problem and concluded that the answer was 28. Her teacher then tells her that the correct answer is 24. The correct answer is an overriding defeater for Saima's belief that the answer is 28. Contrastingly, an undercutting defeater attacks the basis for our belief. It may bring into question the reliability of our belief-forming process. If the process that Saima uses to solve the math problem is brought into question then she has an undercutting defeater that

---

beliefs are not knowledge. Some epistemologists add that Norman can have knowledge if he meets a no-conscious-defeater condition, i.e. if he has no defeaters (propositional and conscious) against the reliability of his source or the truth of his belief. This, in general, is how defeasibility can help externalist theories of justification like reliabilism.

<sup>71</sup>By subjective justification, I mean a connection or link to the reliability of the agent's cognitive process. I do not mean the internalist notion of subjective justification that has to do with having reflective access to something. I have discussed this point in chapter 1 as well.

points out a problem with the source of her belief. For instance, suppose Saima stumbles upon a proposition that defeats the algorithm she used to derive her answer. This defeater is an undercutting defeater because it casts doubt on the algorithm or the source of Saima's belief.

Most overriding defeaters can collapse into undercutting defeaters. For instance, if one has a defeater that calls the truth of their belief into question, it will most likely also bring the basis of their belief into question. However, it is sometimes possible to get a false belief using a reliable belief-forming process, provided that the process is not 100 per cent reliable. On the other hand, if one has reason to doubt the basis of one's belief, one may not necessarily, also have grounds to doubt the truth of one's belief. That is, a non-reliable belief-producing mechanism can sometimes produce true beliefs. This is because non-reliable belief-forming processes can accidentally produce true beliefs.<sup>72</sup>

In the previous section, I mentioned propositional defeaters as the true propositions that the agent is not aware of but ones that can still defeat her belief or the reliability of its source. Here, I will discuss an important type of these defeaters. Some propositional defeaters are *normative*; these are true propositions that we are unaware of but ought to be aware of (Lackey 1999, 2010; Pritchard 2018b). They often directly indicate a problem with the reliability of our belief-forming process and may be undercutting defeaters or overriding defeaters that collapse into undercutting defeaters. The main idea is that these defeaters point out something worrying about the agent's belief that the agent ought to be aware of.

Consider an example of a normative defeater. Amina has a pocket watch that she uses to form beliefs about time. Unbeknownst to her, the watch broke at 5 AM this morning and has not moved since. At 5 pm, she checked her watch to form the belief that the time is 5 pm. Her belief is true because it is 5 pm. Her watch shows 5 o' clock so she does, indeed, have justification for her

---

<sup>72</sup>Sometimes in the course of this chapter I have used phrases like: defeater against a process, or defeater against a process's reliability, or defeater against the process's integratedness. I want to clarify that of course defeaters do not present themselves directly against reliability of a process or its integratedness in a system. When I write such phrases I am thinking of undercutting defeaters that question or attack the source of the belief, and this source may be the reliability of the process or its integratedness.

belief, and this is how she would usually get justification for this kind of belief. But, in this case, there is a normative defeater for Amina's belief, that her watch is broken. If the watch had not been broken, Amina would have been justified in believing the time by consulting her watch. But, because there is a normative defeater, Amina's justification runs into a problem and her true belief is not knowledge.

This example is a classic Gettier case, named after the philosopher Edmund Gettier. Normative defeaters are especially helpful in understanding the relationship between truth and justification in Gettier cases. Gettier presented three cases in his paper to the conclusion that a justified, true belief is not always knowledge (1963). These cases are mostly examples of agents who come to acquire true beliefs via justification that is in some way not connected to the truth of their beliefs. Normative defeaters can help trace this problem. The fact that the agent's justification does not link to the truth of their belief is because they are unaware of defeaters that they ought to be aware of.

More specifically, the true proposition that the watch is broken, in Amina's case, is an undercutting, normative defeater for her belief that it is 5 pm. It is undercutting because it can give our agent reason to doubt the basis, or the source, of her belief. Her belief might still have been true but she would have reason to consider her belief-producing mechanism unreliable and might not have employed the source at all. It is a propositional defeater because Amina is not aware of it. And it is normative because it is the kind of defeater that she ought to be aware of. It is the main reason why her source cannot help her achieve knowledge. If she had been aware of the defeater, she would not have formed the belief.

Like conscious mental state defeaters<sup>73</sup>, normative defeaters can also help describe how an agent can be epistemically responsible.<sup>74</sup> In fact, there is an interplay between these two kinds of defeaters that can best explain how an agent can be epistemically responsible. To give a general idea, if an

---

<sup>73</sup>I mentioned these defeaters in the preceding section. These are the defeaters that the agent is aware of. I will discuss and unpack them even further in the next section.

<sup>74</sup>Pritchard (2018b) points out that the ought in the definition of normative defeaters should be understood as an ought that describes epistemic responsibility.

agent is aware of all the defeaters that she ought to be aware of (normative defeaters), she will be in the position to monitor any conscious mental state defeaters and ensure epistemic responsibility. In other words, if there are any normative defeaters, the agent will be aware of them such that they will become her conscious defeaters that she can address. This is what makes an agent epistemically responsible in employing a target process.<sup>75</sup>

This is why an account of epistemic integration ought to be an explanation of how an agent's normative defeaters can become her mental states. For a process to integrate in the agent's cognitive character means that the agent should become familiar with this process's reliability in a way that if it were to malfunction she would not have a normative defeater, but she would be *aware* of a mental state defeater against it.

To understand this, consider the following example. One day, Amina's watch stops working reliably. There is a normative propositional defeater (that she ought to know of) against any belief she forms using her watch, i.e. that her watch is broken. If her process of checking her watch was properly integrated (via weak epistemic integration) in her cognitive character,<sup>76</sup> she would have a

---

<sup>75</sup>Goldberg (2018) disagrees that normative defeaters have a direct effect on making an agent epistemically responsible or not. He argues that a lack of normative defeaters may disqualify one from being assessed as a responsible or irresponsible epistemic agent. Palermos, in a one-on-one discussion, has also elaborated that he agrees with this conclusion. On their understanding, Amina is not being epistemically irresponsible in any way just because she is not aware that her watch is broken. Here, I would like to comment that normative defeaters are a difficult topic to grasp. Some epistemologists would argue that as a good knower or inquirer Amina should be aware of her surroundings and how they affect her beliefs. She should not be free of blame for never checking her watch. Although, if we grant that Amina otherwise takes very good care of her watch and also checks several times if it is working properly, and it is only that one time she missed, then the scenario changes. Now, Amina is free from blame, and it does make sense to state that she is not epistemically irresponsible for not checking her watch that one time. I can (and will in my discussion in the following sections) bypass taking a side on this debate because it is not necessary for the arguments I want to make in this chapter.

<sup>76</sup>In this particular case, I am assuming that if Amina's watch is properly integrated in her cognitive system, via weak epistemic cognitive integration, it is somewhat like an implant. I do acknowledge that checking time on a watch can also be like using an instrument to get to a result.

conscious mental state defeater against any belief she formed using her watch. The integration of the process would explain why Amina has the relevant mental state defeater against the unreliability of her watch. At least, in case of weak epistemic integration, we know that this is because weak epistemic integration requires that Amina's watch-employing process cooperates with other processes in her cognitive system, and her former process has produced output that has fit well with output from her other processes. And so, when the watch is broken the output fails to assimilate in other processes, giving her a mental state defeater against the reliability of her watch-process. This may be that the time displayed on the watch does not correspond with the daylight outside. When I discuss epistemic integration accounts more specifically, I will come back to a more in-depth development of this concept.

One last note on propositional defeaters is that they are not all normative. Some may be *non-normative* defeaters, i.e. the agent ought not be aware of them. For instance, Samiya looks at the water on the road and forms the belief that it has rained today. But, there is a propositional defeater for Samiya's belief: she lives in Lahore where mirages are in fact very common on hot, summer days. As it turns out, it did rain earlier that day, and Samiya is looking at rainwater and not a mirage. The defeater in this case is non-normative and also *misleading* (Pritchard 2018b). If Samiya had become aware of this defeater, she would not have acquired cognitive success (i.e. true belief). And so, since Samiya did not become aware of the misleading defeater, she did acquire cognitive success. Furthermore, she manifested significant agency in forming her true belief, which makes her case a case of knowledge. With this, our discussion on types of propositional defeaters comes to an end.

Let's now discuss mental state defeaters. There are two types of mental state defeaters: conscious and reflective mental state defeaters. A conscious mental state defeater, as I discussed above, is one that the agent is aware of. To be exact, the agent is aware of the content of the mental state as well as its defeating nature towards her belief. For instance, Aneeqa believes that she has run out of coffee and she needs to buy some more. But, while cleaning the pantry, she finds more coffee and has a conscious mental state defeater (that she has not run out of coffee) against her belief that she has run out of coffee.

The other mental state defeaters are called reflective mental state defeaters. These are the ones that the agent is not aware of, but can become aware of upon reflection (Bergmann 1997). In the following example, Halima has a reflective mental state defeater. In March, Halima read on the World Health Organisation's (WHO) website that the novel coronavirus will not be affected by heat. But when, a few months later, the number of new cases began to drop in Pakistan, Halima formed the belief that the stifling summer was slowing down the spread of the virus. Distracted by the shred of hope she was feeling, she could not recall what she had read on the WHO website. The mental state about what she read on the WHO website is her reflective mental state defeater against her belief that the heat has affected the spread of the virus, and if she manages to recall her reflective mental state defeater then it becomes her conscious mental state defeater.

An interesting question to ask is if reflective mental state defeaters can defeat? We know that they can defeat when they become consciously available to the agent as conscious mental state defeaters. But, do they defeat when they are not consciously available to the agent? If Halima cannot recall from memory what she read on the WHO website, can it defeat the belief she forms later?

Since she is not aware of the defeater, it cannot defeat the same way a conscious mental state defeater defeats. However, it does defeat in the way that propositional defeaters defeat. Propositional defeaters are true propositions that defeat one's beliefs, even when one is not aware of them. This is the kind of externalist defeat I discussed above. They may be normative defeaters. That is, they defeat the agent's belief and the agent ought to be aware of them. What Halima read on the WHO website is something she ought to remember, but she does not. In that sense, in the way that normative defeaters defeat, her reflective mental state defeater defeats as well. I will come back to reflective defeaters in the section after the next.

## The no-defeater condition

Now that I have discussed the types of defeaters, I will bring attention to some important, and the possibility of other, no-defeater conditions on epistemic

responsibility. Simply put, a no-defeater condition means that a certain kind of defeater should not be present (either out in the world or should not be present to the agent's awareness), in order for the agent to responsibly employ her belief-forming process. Epistemic responsibility can be described in terms of a no-defeater condition. I suggested above how a no-normative-defeater and a no-conscious-mental-state-defeater condition, together, can help us make sense of epistemic responsibility. This simply means that if there are no normative defeaters for the agent's belief (produced by a certain process) and the agent is not aware of any (conscious) mental state defeaters, then she is most likely being epistemically responsible in recruiting the said process.

Just like defeaters are related to externalist or internalist justification, the no-defeater conditions can also be categorised in this way. A no-defeater condition for propositional defeaters (externalist kind of defeat) would demand that there are no true propositions that attack the truth of the agent's belief or the source of her belief. Similarly an internalist no-defeater condition would be about the internalist kind of defeat, like conscious mental state defeaters. That is, the agent should not possess any experience, or propositional attitude, or a combination of these two, that defeats the truth of her belief or attacks its source.<sup>77</sup>

Let's discuss defeasibility or the no-defeater conditions in terms of the overarching debate between weak epistemic integration and the subpersonal epistemic integration. The weak epistemic integration account describes epistemic responsibility in terms of the agent's responsiveness to evidence against her process's reliability. The subpersonal epistemic integration account describes epistemic responsibility in terms of the subpersonal operation of precision estimation, whereby a target process is assigned a reliability weighting by the agent's cognitive system. These two accounts of epistemic responsibility

---

<sup>77</sup>A no-defeater condition of this sort is also supported by Alvin Plantinga's account. He states that 'warrant' is the quality that makes true belief knowledge, and that proper 'proper functioning' of an agent's belief-forming mechanism is necessary for warrant (1993). If an agent is aware of a defeater against the proper functioning of her cognitive faculty, she cannot employ the said faculty responsibly. Hence, there is a no-conscious-defeater condition for warrant.

can be described in terms of no-defeater conditions.<sup>78</sup> Let me describe what these conditions will be and how they fit into these theories of integration. First, let's discuss weak epistemic integration.

Virtue reliabilism has taken help from epistemic integration accounts to explain how a reliable process can be recruited responsibly by an agent. Pritchard's weak epistemic integration is one such account. It describes how one can responsibly employ a belief-forming process even when one cannot directly access its reliability. In this section, I will discuss how integration of a process and its employment can be understood in terms of two no-defeater conditions: the no-normative-defeater condition and the no-conscious-defeaters condition.<sup>79</sup>

Let's first discuss the no-normative defeater condition. According to weak epistemic integration, a process initially or *diachronically* integrates in the agent's cognitive agency if it is reliable and has been used as a stable disposition (i.e. frequently for a considerable period) of the cognitive agency (Greco 2010; Palermos 2011).<sup>80</sup> In this initial period, one's agency becomes familiar with the new process by employing it several times (the process has not yet integrated so the agent is not employing it responsibly). The mental states or beliefs produced by the new process in this period will cohere with the rest of the agent's system (beliefs as well as other processes). And, if there are any undercutting normative defeaters, ones that would make the agent doubt the reliability of her process if she was aware of them, they would become apparent. Their presence will keep the new process from properly cohering with the rest of the system, and will also keep the agency from becoming familiar with the new process. In other words, if there is a normative defeater against the reliability of the said process, the agent

---

<sup>78</sup>Note that Clark does not present a defeasibility account for his predictive processing account of subpersonal epistemic integration. I will attempt to extrapolate such an account in order to evaluate how his integration fares in providing us with a concept of epistemic responsibility.

<sup>79</sup>Some of the discussion has been influenced by Orestis Palermos' unpublished article, "Virtue Reliabilism and Epistemic Defeat" (n.d.). In this paper, Palermos systematically defends the thesis that weak epistemic integration automatically satisfies the no-defeater condition that informs epistemic responsibility. He argues that this no-defeater condition is a no-conscious-defeater condition. He does not argue for a no-normative-defeater condition.

<sup>80</sup>See Carter, Clark, and Palermos (2018) and Palermos (n.d.) for diachronic and synchronic agentive control on one's target process and its relation to epistemic responsibility.

will not be able to integrate it. So, the initial integration will take place in terms of a no-normative-defeater condition.

When the agency has made substantial connections with the new process, all these (possible) normative defeaters will have become mental states of the agent. In other words, the agent will become aware of whatever she ought to be aware of about this new process. And, therefore, the no-normative-defeater condition will be met, if initial integration is successful. From here on, if the agent monitors the reliability of this process with another no-defeater condition, the agent can responsibly employ this process. This next step is the *synchronic* monitoring of the reliable process. It is governed by the no-conscious-defeaters condition. Accordingly, now that the agent is aware of what she ought to be aware of, and if there are also no conscious mental state defeaters against the reliability of her target process, she can employ it responsibly.

In case her process malfunctions, its belief ought to be met with a conscious mental state defeater, because the agent is aware of all that she ought to be aware of. Such monitoring of reliability is therefore automatic. It is the integratedness of the system that helps keep a check on the reliability of other processes of this system. If a process produces an obviously false belief, other beliefs or mental states will come forward to defeat it in shape of a conscious mental state defeater, and will alert the agent of the unreliability of her process.<sup>81</sup>

The initial stage of integration of a process and its monitoring can be understood in terms of an analogy. Let's say you purchase a new bike and then ride it frequently for a substantial amount of time. In this time you get very familiar with all its mechanisms: how the brakes work, how to check the wheels and put air in them, how the gears operate, etc. In a way, you now have a vast number of mental states about your bicycle's operations. So, if your bike was to malfunction one day, the problem will not go unnoticed. It is likely that you are aware of all the things that you ought to be aware of about your bike. But, if you did not use your new bike for very long after purchasing it, then there are likely to be some things that you ought to be aware of but you are not aware of

---

<sup>81</sup>Palermos (n.d.) suggests that these two stages of weak epistemic integration should be understood as diachronic and synchronic contributions to epistemic responsibility.

(i.e. some normative defeaters about the reliability of the bike). As a result, you will not be able to monitor the smooth working of your bike because you are not familiar with the concept.

Hence, the no-normative-defeater condition along with the no-conscious-defeater condition, together, account for the agent's epistemic responsibility. On their own, they do not play the role that weak epistemic integration has assigned them. For instance, the no-normative-defeater condition alone does not fulfil requirements for epistemic responsibility. For, that means the agent is aware of all there is to be aware of about the process, and has no defeaters that she ought to be aware of, but she might be conscious of a defeater and still go on to use the said process. The no-conscious-defeater alone does not work either. It may be that the agent is not aware of a normative defeater that defeats her process's reliability.

On the same note, it is true that the agent is not responsibly recruiting her reliable process when the process is still in the course of initial integration. But the agent is working towards her responsible employment of her process, which makes the initial integration a part of the epistemic responsibility account. It would be wrong to assume that an account of epistemic responsibility can be fully grasped if we consider only the agent's satisfaction of no-conscious-defeaters condition, or her monitoring of the reliability of an already integrated process. Of course, the latter is extremely important for epistemic responsibility, but it is grounded in the initial integration.

In terms of the bike analogy, the agent will manifest responsibility when she has already made herself familiar with the operations of her bike. But, this does not mean that we ignore the initial process in which the agent gets to know her bike. She may not be manifesting responsibility at that time, but the responsibility she later manifests in monitoring her bike's mechanisms is based on her first getting to know her bike. Hence, we should understand weak epistemic integration in terms of both these no-defeater conditions together (neither one without the other).

## Subpersonal defeasibility via reflective defeaters

In the previous section, I unpacked how no-defeater conditions inform responsibility in the weak epistemic integration account. Here, I am going to take a similar course of action for subpersonal epistemic integration. An important thing to note is that Clark does not provide a defeasibility theory for his subpersonal epistemic integration account, but that does not mean that one cannot be extrapolated from his account. Any fallible epistemology can be understood in terms of defeasibility. The epistemology that Clark puts forward is only different to Pritchard's account in terms of the level of explanation that can meet the responsibility requirement. Hence, it would be fair to understand his account as a fallibilist subpersonal virtue epistemology. This is one reason I am interested in an account of defeasibility for Clark's subpersonal epistemology account. Furthermore, integration can be accounted for in terms of no-defeater conditions like we have seen above. This is another reason why it would be a good idea to sketch a defeasibility theory for subpersonal epistemic integration.

Before I begin to devise a subpersonal epistemology, I want to bring attention to a comment from Clark (2015). Even though Clark mostly defends his epistemology as a subpersonal virtue epistemology, he also mentions that it may pass as a sophisticated process reliabilism as well. Either way, I think, a defeasibility account will work just the same. All fallibilist epistemologies, that accept that agents adjust the justification for their beliefs when presented with new evidence, can be interpreted in terms of a defeasibility theory. And, we can analyse such a theory to evaluate how it addresses epistemic responsibility.

Like weak epistemic integration, Clark's subpersonal integration also describes how a target process integrates in the agent's cognitive character such that the agent takes responsibility for the said process. The main difference between the two integrations is that subpersonal epistemic integration allows the agent to meet epistemic responsibility requirements without the agent (at a personal-level) becoming involved in the process. So, the suggestion is that sometimes epistemic responsibility can be explained only in terms of what goes on at a subpersonal level. This means that a defeasibility account for Clark's integration will have to do with the subpersonal level that is not conscious to

the agent. It is not temporarily unconscious to the agent; in fact, it is not even potentially conscious to the agent.<sup>82</sup>

Which defeater will be involved in a subpersonal defeasibility account? We are looking for defeaters that are in our minds but not consciously accessible to us. Reflective mental state defeaters are a good candidate. They are present inside our minds but they are not conscious to us. Some might argue that they may not be conscious to us but if we reflected and were successful in recalling them, they would become conscious to us. And, this is not in line with the things that occur at the subpersonal level. What happens at the subpersonal level is not consciously accessible to us.

However, as soon as reflective mental state defeaters become conscious to us, they become conscious mental state defeaters and do not remain what they were. They only remain reflective defeaters until they are unconscious to us, in this sense they, in fact, are consciously inaccessible for us. It seems that they serve the purpose here because they are not consciously accessible to us unless they become some other defeaters. Let's see what happens when we understand subpersonal defeat in terms of these reflective defeaters and how, if at all, they defeat.

Maryam has a reliable device implanted and integrated in her brain. This device gives her an account of the items she has in her fridge, even when she is away from home. She has used it for most of her life and it has worked really well for her. This also means that she has been responsibly employing her device all this while. On her way to the grocery store today, she uses her device to form the belief that she has 12 eggs in the fridge. However, Maryam has a reflective mental state defeater against her implant, which is that 'she has no eggs left in the fridge.' Let's suppose that she picked up the last egg from the fridge today but has no recollection of this fact because she was busy thinking about something that happened at work. If she was asked to reflect on how many eggs there are in the fridge she will be able to recall that there are none left. In this case, when Maryam uses her implant to form the belief that there are

---

<sup>82</sup>Here, I am discussing states that are not consciously accessible to the agent ever, compared to states that the agent may not be conscious of but can become conscious of upon reflection.

12 eggs, she is distracted again about something that happened at work which is why her reflective mental state defeater does not become a conscious mental state defeater. She goes on believing that there are 12 eggs in the fridge and that she does not need to buy more eggs. Since the content of the reflective defeater does not present to her consciousness, we can assume that Maryam will go on believing that there are 12 eggs in her fridge. Of course, that is not very responsible of Maryam.

As I discussed earlier, reflective defeaters defeat the same way that normative defeaters do (externalist sense of defeat) and not the way conscious defeaters do (internalist sense of defeat). So, Maryam ought to be aware of the fact that she took out the last egg from her fridge today. But, the crucial question is, what does this case tell us about epistemic responsibility? Do we think it is epistemically irresponsible of Maryam to not recall her defeater against her belief?

Consider an example of reflective defeat from Bergmann (1997). Chuck has a cognitive malfunction and thinks that the reports he hears between 4.15 pm and 4.30 pm are highly unreliable. This prevents him from learning some things, but in most other ways he is just like other people. One afternoon, at 4.20 pm he hears a noise outside his house. When he goes out to inquire he finds heavy equipment and city trucks parked outside his main door. When he asks a person what is happening, he is told that there is work to be done on the main water supply. He is also told that his wife was informed of it a day before. Chuck believes this man, and it turns out that this man was telling the truth. But, if Chuck had reflected on his mental state, that the reports he hears within a certain time frame are unreliable, he would not have believed the man outside his house (Bergmann 1997).

Bergmann's intuition on this case is that Chuck has knowledge, even though he has a reflective mental state defeater in this case. This indicates that it is not necessary for epistemic responsibility that an agent should not have any reflective defeaters (a no-reflective-defeater condition).<sup>83</sup> In other words, a

---

<sup>83</sup>Bergmann explains that such cases are cases of *Evidential Luck*, a kind of epistemic luck that is not incompatible with knowledge (1997). Evidential luck simply means that the agent is lucky to have the kind of evidence she has. Chuck is simply lucky to have not recalled his

no-reflective-defeater condition is not necessary for epistemic responsibility. One can responsibly form a belief even though there is a reflective defeater against it.

Some might object to this conclusion by focussing on how Chuck's reflective defeater is only a misleading defeater (and not an actual defeater). If Chuck becomes aware of it, he would be misled. In this way, Chuck's reflective defeater is not like Maryam's actual defeater. An argument can be made that the agent ought to be aware of her reflective defeaters, unless it is a misleading defeater. This would suggest the conclusion that there is a no-actual-reflective-defeater condition on responsibility.

To respond to this objection, consider Chucky. Chucky is in the exact same scenario as Chuck. Let's suppose that the man Chucky speaks to outside his house is actually lying about working on the main water supply. Chucky also has a reflective mental state defeater about the false reports he hears between 4.15-4.30 pm. Since the man is lying about working on the main water supply, it means that Chucky's reflective mental state defeater is an actual defeater (not a misleading defeater). Is it epistemically irresponsible on Chucky's part that he failed to recall his reflective mental state defeater?

I think it would be wrong to commit that Chucky is irresponsible and Chuck is not. As agents, we have no access to our reflective mental state defeaters and also no access to whether these defeaters are actual or misleading. So, we cannot commit to the claim that one of our agents is irresponsible and the other is responsible based on factors that are outwith their control. What we can conclude from this discussion is that reflective mental state defeaters do not help us understand epistemic responsibility. So, if subpersonal defeat is supposed to be understood in the way reflective defeaters work, it does not tell us anything about how an agent has been responsible or irresponsible.

Recall my discussion from the previous section. According to weak epistemic integration, both normative defeaters (externalist defeat condition) and conscious mental state defeaters (internalist defeat condition) are important for epistemic responsibility. The agent first gets acquainted with her process by employing it several times, while in the background her no-normative-defeater condition is being met (if the process is reliable). She will be aware of  

---

reflective mental state defeater and that good fortune has led him to knowledge.

most propositions that she ought to be aware of about the reliability of her process (thanks to the no-normative-defeater condition). Then, her process's reliability can be monitored by the no-conscious-defeater condition. As long as there are now no conscious defeaters against the reliability of her process (now that she is aware of most of what she ought to be aware of about her process), she can employ the said process responsibly.

So, a no-normative defeater condition alone does not capture epistemic responsibility. It is together with the no-conscious-defeater condition that it provides the weak epistemic integration account. It is important that the agent first gets acquainted with her process to be able to monitor it for conscious defeaters.

Let me bring my discussion together now to make my point. In this section, I posed a question about whether subpersonal defeasibility can be understood in terms of reflective mental state defeaters. I came to the conclusion that reflective mental state defeaters only defeat in the way normative defeaters do. That is, the agent ought to be aware of the defeater against their belief or the reliability of their belief forming process. However, according to some of the cases I considered, agents are not epistemically culpable or irresponsible if they fail to recall their reflective defeater. That means that the presence of a normative defeater alone does not make you irresponsible. It is when the no-normative-defeater condition is understood together with the no-conscious-mental state defeater that we are able to devise a complete account of epistemic responsibility. This is what occurs in a weak epistemic integration. The agent first gets sensitive to the reliability of her process and is then able to monitor it because she has no normative defeaters.

We see this phenomenon in play in Clark's Temp2 case (2015). Temp2 is a predictive processing agent whose integrated (via subpersonal epistemic integration) process gives her true beliefs about the temperature of her room. Her process consists of a genie, a thermostat, and a thermometer. The genie adjusts the thermostat in the the room such that Temp2's reading corresponds to the actual temperature of the room and her beliefs are true. This process is reliable. Although, one Wednesday, Temp2's genie decides to take the day off (and all other coming Wednesdays off as well, but that's not relevant for what I

have to say). Temp2's beliefs are likely to be all over the place on this day. Let's assume that Temp2 has something like reflective defeaters against these beliefs (or their source).

Reflective defeaters defeat the way that normative defeaters do. So, Temp2 has a defeater that she ought to be aware of, which means that it should not just be understood at the subpersonal level. It is an undercutting defeater and provides direct evidence against the reliability of the source of her belief. Some epistemologists, at this point, will argue that Temp2 is epistemically irresponsible because she is not aware of this normative defeater, and she cannot have knowledge in this case. Others might assert that she cannot acquire knowledge, but she is not epistemically irresponsible either. I am not going to pick one of these sides. I have not argued for either side here. I have only pointed out that this understanding of integration in terms of reflective defeaters does not give us an account of epistemic responsibility. We cannot decide if Temp2 is responsible or irresponsible because the subpersonal epistemic integration account does not provide an appropriate understanding of epistemic integration, at least not in terms of reflective mental state defeaters. For a complete understanding, an externalist condition like no-normative/reflective-defeater condition needs to be paired with a conscious mental state defeater condition.

## Subpersonal defeasibility via subpersonal defeaters

I will now discuss the possibility that the subpersonal defeasibility account that I presented in the previous section is false. And, perhaps reflective mental state defeaters are not the right way to conceive subpersonal defeat. Here is a different account.

Reflective mental state defeaters may not be helpful to grasp subpersonal defeat because they only refer to our mental states. These defeaters are our mental states that defeat the reliability of a target process. It is possible that subpersonal defeat will include information that does not come from our mental states but also from our subdoxastic states (pertaining to what lies under the formation of beliefs), i.e. the information that our cognitive systems

(e.g. perceptual faculty and memory) interchange at a subpersonal level like data on wavelengths of light, size and shape of objects, etc. Let's suppose that the defeaters that can help capture subpersonal defeat are ones that include mental states and subdoxastic states that can be made sense of at a subpersonal level. That is, our cognitive system can engage with this sort of defeat. I will call these *subpersonal defeaters* and examine if they display a better picture of subpersonal defeasibility. Let me unravel how these defeaters will perform.<sup>84</sup>

Subpersonal epistemic integration is the function of precision estimation units that assign reliability weightings to candidate processes so they can integrate in our cognitive character. These units assign reliability weightings in virtue of the information available to them (subpersonally). This information that is present to the precision units (parts of our cognitive system) may be in form of subpersonal defeaters, i.e. subdoxastic or doxastic (relating to beliefs). Also, notice that our cognitive systems will be able to detect these subpersonal defeaters and assign reliability to the target process accordingly. If there are no subpersonal defeaters, the precision units will assign sufficient reliability weightings and integrate the process. If there is a subpersonal defeater against the reliability of the said process, the precision estimation units will assign insufficient reliability weighting to the process.

Subpersonal defeaters are not like reflective mental state defeaters, and so they do not defeat like normative defeaters do. These defeaters are accessible to the agent's cognitive system, even if they are not consciously accessible to the agent. What happens when a predictive processing agent's integrated process begins to malfunction? A subpersonal mental state defeater becomes available to the agent's cognitive system (precision estimation units specifically) and so the system assigns insufficient reliability to the said process.

At this point, the reader will be forced to think that this seems to be a good account of epistemic responsibility. The task that the agent performs in weak epistemic integration account has been delegated to her cognitive system.

---

<sup>84</sup>These defeaters are not entirely internal like conscious mental state defeaters, but they are not entirely external either like normative defeaters or reflective defeaters that one cannot recall. They are internal in the sense that they are accessible to a part of our cognitive system, even if not us on a personal level.

Everything occurs the same way, albeit at a subpersonal level. The agent's cognitive system assigns insufficient reliability to the candidate process if there is a subpersonal defeater against it. And, accordingly, the process will un-integrate. Overall, it seems that if we introduce a no-subpersonal-defeater condition, the agent can meet the epistemic responsibility requirement.

There is a catch. Consider how Clark describes the un-integration of Temp2's process when her genie begins to take Wednesdays off.<sup>85</sup>

Over time, sensory cues indicative of the Wednesday context (e.g. seeing the word Wednesday on the computer date screen that she views every morning before starting work) come to be associated with unreliable information from the thermometer. Sub-personal precision-weighting mechanisms then cause Temp2, even when she looks at the thermometer, to fail to attend to what it says. (Clark 2015, 3769)

According to this quote, Temp2 will use her process for several Wednesdays, before her subpersonal agency or the precision estimation mechanism will realise that her process is not reliable on Wednesdays. On these Wednesdays, when Temp2 is forming false beliefs, she does not become aware of any problems with her process.

In terms of the defeasibility theory, this means (and this is the catch) that there are subpersonal defeaters present against Temp2's false beliefs (on Wednesdays), but her cognitive system does nothing about them. On the weak epistemic integration account, Temp2 will have a conscious defeater against her source's reliability on Wednesday, and this means that she will be aware that the reliability of her source has been defeated. Whereas, on Clark's subpersonal integration account, a subpersonal defeater may be present, while Temp2's cognitive system may still keep employing the broken resource. This shows epistemic irresponsibility on the part of the agent.

---

<sup>85</sup>Clark would not call it un-integration. According to his view, Temp2 simply adjusts to the genie taking a break on Wednesdays. I think it is fairly derivative to understand this as Temp2 un-integrating one process (in which the genie takes no breaks) and integrating another (in which the genie takes Wednesdays off).

It will simply be irresponsible of one's agency to continue employing a process that it has a subpersonal defeater against. The purpose of a defeater in a responsibility account is to allow the agent or, in case of subpersonal mechanisms, the cognitive system of the agent, to successfully and timely un-integrate a process that is no longer reliable (or manifest responsibility by dealing with defeaters in some other way). This response determines if the agent (or her cognitive system) is sensitive to the reliability of her process. It seems that on Clark's view, the agent's cognitive system is slow at responding to the sudden unreliability of her target process.

Some might object that even weak epistemic cognitive integration demands that the agent should employ a process frequently over a substantial period, and since the process is not yet integrated, this employment is not responsible. The agent, when she comes across a new process that she begins to employ frequently, has not yet developed any link to the reliability of this process and is therefore recruiting it irresponsibly. But, notice, that this is not the same as deploying a process one has a subpersonal or personal defeater against. When a process is integrating, the agent is not aware of any mental states against the said process. So, even if the agent is not yet responsibly recruiting the process, she is not outright irresponsible either. But, when Temp2, the predictive processing agent, continues to employ a process that her agency can discern a subpersonal defeater against, she is doing so irresponsibly because it does not seem like her agency has a link to the reliability of its process anymore. If there was such a link, the agency would not employ a defeated process.

This issue with subpersonal defeat demonstrates yet another weakness of the subpersonal epistemic cognitive integration account. It proves that subpersonal epistemic cognitive integration is an inadequate understanding of (minimal) epistemic responsibility or perhaps that it just cannot be understood in terms of defeasibility. The issue with the latter conclusion is that a fallible epistemology can always be interpreted in terms of defeaters. And it is necessary that an account of epistemic hygiene or responsibility explain fallibility of a reliable process and the sensitivity one's agency has to such a fallibility.

## Conclusion

Weak epistemic integration provides a thorough understanding of epistemic responsibility. It depends on two no-defeater conditions: no-normative-defeater and no-conscious-defeater condition. Accordingly, an epistemic agent should first get to know the target process by recruiting it frequently over a period of time. In this time, the agent will most likely become familiar with the reliability of the process. Then, the agent can monitor (in the background) any evidence against this reliability. This is a summary of how processes integrated via weak epistemic integration can be responsibly invoked by epistemic agents.

Subpersonal epistemic integration also needs to describe how an epistemic agent can responsibly employ her process. I devise two plausible subpersonal defeasibility theories for this purpose but find them both wanting. The first one is based on reflective mental state defeaters, i.e. defeaters that we have in our minds but we are not aware of. In their defeat, they are the same as normative defeaters, that we ought to be aware of. I argue that a theory based on these defeaters does not help us deduce if the agent is responsible or irresponsible in employing her process. Further, I suggest a new type of defeaters, subpersonal defeaters, that may be able to describe subpersonal defeasibility better. These are not entirely external like normative defeaters, neither are they completely internal like conscious defeaters. They are accessible to our cognitive system, instead of our consciousness. However, even a theory of defeat based on subpersonal defeaters does not inform epistemic responsibility of a predictive processing agent.

## Chapter 6 Knowledge bloat and demarcating the epistemic agent

This chapter introduces an elaborate case of extended cognition based on a future model of Amazon Alexa. This AI, Alexa, is integrated into my cognitive system and has the ability to abet other advanced technologies to integrate into my cognitive system as well. With the help of this case, I bring forward two more challenges for a purely subpersonal understanding of epistemic integration. Firstly, it leads to a knowledge bloat situation where the agent can know far more than it is possible to know, and may have no conscious accessibility to where all of this knowledge is coming from. Secondly, the fully subpersonal concept of epistemic integration makes it difficult for us to demarcate the epistemic agent properly. Let me give an overview of my discussion in this chapter.

As an extended part of my cognition, Alexa is in a constant, two-way subpersonal exchange of information with my cognitive system. I suggest that this subpersonal exchange can result in the realisation of a belief. This belief will partially be realised by my neural states and partially by physical states present in Alexa. Drawing comparisons with the famous Otto-Inga case, I suggest that this belief is also similar to Otto's belief that he has but is not aware of. Then I ask why and how should this belief be creditable to me in light of why Otto's belief is ascribed to him. From hereon, I take my discussion from metaphysical cognitive integration into epistemic cognitive integration. I put aside the belief attributability problem and wonder if the belief being realised by my cognitive system and Alexa, working together, can be knowledge for me. And, if so, will the knowledge be creditable to me?

Further, I discuss the possibility of multiple layers of cognitive extension. For instance, suppose Alexa, an extension of my cognitive system, can help the Amazon Pantry to also extend my cognition, and the Pantry further collaborates with other advanced technologies to extend my cognitive system, and so on. The idea is that if epistemic integration, that makes belief-forming processes knowledge-conducive, is an entirely subpersonal notion then all these layers of

cognitive extension may be producing true beliefs that are knowledge, and they ought to be creditable to me. However, I am not aware and not even in the position to become aware of most of this extension. This is how I describe the knowledge bloat issue.

The other problem that Clark's minimal concept of epistemology runs into is the problem of delineating the epistemic agent. It is unclear why a belief that is realised by Alexa and I should be ascribed to me, or why knowledge of such a belief should be creditable to me. If the belief should be attributed to a new agent that composes of Alexa and I, then things get complicated when you add multiple layers of extension. Some beliefs will be attributed to evolved systems that include many advanced technologies along with my neural apparatus. Subpersonal epistemology, because of how it understands epistemic integration, does not have a solution to this problem. I conclude by arguing that Pritchard and Palermos' weak epistemic cognitive integration has a theory that demarcates the epistemic agent, the one who ought to be creditable with knowledge in these cases.

## Alexa

The Amazon Alexa (call it Alexa 2020) is a virtual assistant. Alexa 2020 and I exchange information throughout the day in different ways. It monitors the temperature of my surroundings and tells it to me. It is connected to my smartwatch so the temperature of my room shows on my watch as well. I can dictate my grocery list to Alexa, and it can repeat it for me when I need it. It keeps a record of what I buy on Amazon and uses AI to infer what other things I would be interested in. I can tell Alexa to make lists of plants I want to buy. I can tell it to look up seeds of the plants I want to grow. Alexa can keep a record of the money I am spending and keep updating me with that kind of information. It can send me updates and notifications about weather, news, wallet information on my phone and my watch.

It is now the year 2030, and there is an Alexa 2030 in my home. My virtual assistant has, in a manner of speaking, grown with me. It has all our information exchange from the last (approximately) ten years. There are more

sophisticated algorithms by which Alexa now processes information. There is more AI inculcated in Alexa. But, most importantly, Alexa has a lot of data on me and about me, and I have a track record of Alexa's output as well. The kind of information exchange between Alexa and I is still reciprocal and on-going. When Alexa 2030 recommends me a book online, it is usually the kind of book I want to read. It can order the right fertilisers for my plants because it has all the information about the different fertilisers I have tried, when I need them, which of them works best for my plants, where to get this fertiliser, etc. Alexa 2030 is more useful to me because our information exchange is more tuned to my character— it learns from collecting data about me.<sup>86</sup>

It may look like the information exchange, between Alexa and I, has only improved in terms of how Alexa now has access to more information. That is, Alexa 2020 was privy to less information about me than Alexa 2030, which is why it can perform better at recommending me things and purchasing things for me. But, my relationship with Alexa has improved in the sense that there is two-way flow of information now. Not only is Alexa better at information about me because of the reservoir of data that it has on me, but I have also received a lot of data from Alexa in all these years. This means that I have also come to trust Alexa more and more, and rely on it for more things, and I have adopted more of its features in my life because I have found them accessible and reliable. And, I might not have reflected on this detail at all, but Alexa 2030 is more incorporated in my life because I find our information exchange more accessible, swift, and reliable now.

Now, imagine a day in my life with Alexa 2030. I have a presentation to make at work today and my mind is occupied thinking about it. I take an egg from the fridge to make myself a quick breakfast but I don't realise that I

---

<sup>86</sup>When devising this case, I had supervised machine learning in mind. When an AI is given more data, it improves its algorithms based on the scores of data provided to it and the goals that it is directed towards. This, in general, is supervised machine learning. Such learning is a result of continuous and reciprocal exchange of information. The AI receives data on something related to the agent, upgrades its function according to the data, receives feedback on the upgrade, improves its upgrade according to the feedback, and so on. I have tried to show continuous reciprocal causation between Alexa and my cognitive system in terms of this exchange.

have picked up the last egg from the fridge. There are no eggs left in the fridge. Since Alexa oversees my activities, it has registered that I have no more eggs left in the fridge and that I will need to get more eggs before breakfast tomorrow. I am very occupied with my presentation work so I fail to register the information about the number of eggs I have left.

Furthermore, Amazon Alexa is connected to the Amazon Pantry and, therefore, sends the information about my egg situation to the pantry. It then receives a response from the pantry that 12 new eggs will be delivered by noon today. Around noon, someone from the Pantry comes home to deliver 12 new eggs, and Alexa lets the person in to refill the egg-shelf in the fridge. Now, I have 12 new eggs in the fridge, which is another piece of information that I do not have access to, but Alexa does.

### Does Alexa extend my cognition?

Let's start with a very basic picture of what seems to be going on in this case. There is an exchange of information between Alexa (I will use Alexa for Alexa 2030 from here on) and I, and it results in my acquiring a new batch of eggs the next day. I am going to consider, in this section, if I am employing Alexa as an extended cognitive process. Some of the conditions that I can look for to see if Alexa meets the extended cognition requirement are the functional parity principle, the glue and trust conditions, and the continuous reciprocal causation (CRC) requirement.

I have Alexa on my phone and on my smartwatch. I automatically trust the information that it shares with me. I easily access and reliably invoke my readily available process of employing Alexa to fulfil my cognitive tasks. This relationship with the process of manipulating Alexa can indicate that the glue and trust conditions are being satisfied. But, we have seen in previous chapters, that the glue and trust conditions do not sufficiently or necessarily confirm extended cognition.

Is there functional parity in how I employ Alexa and how I employ my innate cognitive processes? It can be argued that there is no functional parity in this case, since cognitive processes that lead to me buying eggs are so different

that issues of parity simply do not arise. Employing Alexa is not equivalent to recruiting an artefact that functions similar to my biological memory, for instance. If there was no Alexa in the picture, several cognitive states would have led me to buy more eggs and restock my fridge. With Alexa in the scene, all these cognitive processes can be performed by Alexa. In a broader sense, it seems Alexa can be employed to meet the cognitive goals that I usually use my innate faculties for. For instance, different innate faculties in my cognitive system interact with each other to bring about cognitive states the same way some of my innate cognitive faculties work with Alexa's operations to bring about the cognitive states that I have discussed above, namely cognition tasks that result in my purchasing of more eggs for my next breakfast, etc. However, this is not the argument I make here. I agree with Menary (2010) that the functional parity principle is only an intuition pump for extended cognition. It is not an argument in support of the thesis.

I will argue that we should regard Alexa as an extended cognitive process of mine on the basis of the CRC requirement which is both necessary and sufficient for extended cognition (Palermos 2014b). The CRC condition, you may recall, is based on the Dynamic Systems Theory. Accordingly, when two systems reciprocally and continuously exchange information, they can temporally evolve into a new hybrid system. The two systems can, in this way, produce output and affect other processes together as a whole unit. Their output can drive the cognitive loop forward. Alexa and my cognitive system have a similar relationship. Alexa's mechanisms and some innate faculties in my cognitive system can continuously and reciprocally interact to process information. This allows my employment of Alexa to drive my cognition forward. Some of the cognitive output produced is realised by Alexa and parts of my cognitive system together. All in all, if the interchange and processing of information between my automatic employment of Alexa and my innate processes meets the continuous reciprocal causation requirement, we can say that Alexa extends my cognition.

## Does Alexa extend my mind?

Another argument that can be made is that Alexa extends my mind. The *extended mind thesis* is more radical than the extended cognition thesis (Clark and Chalmers 1998, 12; Carter et al. 2018; Gertler 2007; Menary 2010). It proposes that not only are our cognitive states brought about by physical states that may partially reside outside our head, but that our mental states may be realised by such physical states as well. So, mental states, like our beliefs or desires or other affective states (Colombetti and Roberts 2015), may be realised by AI algorithms, like Alexa, or external mechanisms of other devices.

Note that, so far, I have only discussed my exchange with Alexa in terms of information. As I mentioned in the previous section, we only need to look at how information exchange between a part of my cognitive system and Alexa's operations are bringing about my cognitive states. So, the extended cognition thesis can be furthered in terms of this subpersonal exchange of information between my cognitive system and Alexa.<sup>87</sup> In terms of cognitive integration, it means that Alexa's operations can metaphysically cognitively integrate into my cognitive system entirely subpersonally because of the high degree of interconnectedness and complementarity of exchange between our systems.

What I am proposing in this section is that Alexa and my exchange of information may bring about more than just a cognitive state; it may bring about a mental state. Mental states, like beliefs and desires, supervene on physical states.<sup>88</sup> This means that there cannot be a change in the mental state without there being a change in the physical states that realise the mental state. Some of these physical states that realise my mental states, like beliefs, may be present in Alexa.

To further this theory, it is important to find out what mental states are at play in this case. Let's say a mental state is brought about because of my

---

<sup>87</sup>Some might understand subpersonal exchanges only in terms of what goes on in the neural apparatus of the biological agent, but I take Dave Ward's (2012) understanding of the notion that subpersonal exchanges do not stop at the neural region. They extend into the environment.

<sup>88</sup>For an understanding of the concept of supervenience, see McLaughlin and Bennett (2018).

information exchange with Alexa. This mental state may be the belief that 'there are no eggs left in the fridge.' This may have been brought about by the cognitive information exchange between Alexa and I, or (another way to put it is) physical states in Alexa and I may realise a mental state that 'there are no eggs left in the fridge.'

The thing about mental states is that they are assigned to a subject. It is a characteristic of beliefs and desires, for instance, that they are attributed to someone (or something). This is why, if our discussions revolve around a belief that 'there are no eggs left in the fridge,' then it is also important to address whose belief it is. In a scenario, where all the physical states that realise the said mental state are present inside my cognitive system, it makes sense to attribute me with the said mental state. But, if some of the physical states that realise this mental state are present in Alexa, then who (me, Alexa, or Alexa and I as a unit) do we attribute the belief to and why? The latter is the more important question here; that is, why do we attribute a belief to someone?

Clark and Chalmers' Otto-Inga case (1998) is an extended mind example. Otto is an Alzheimer's patient who keeps a notebook with him everywhere he goes. All important details that Otto wishes to remember, like dates and events, he writes them down in his notebook so he can retrieve the information when he needs it. At an event that Otto and Inga are both attending, the speaker mentions an exhibition at the Museum of Modern Art (MoMA). Otto looks into his notebook to locate the information that 'MoMA is on the 53rd street.' Contrastingly, Inga recalls from her memory the proposition that 'MoMA is on the 53rd street.' This proposition is a belief (mental state) of Inga's, and it is realised by (or supervenes on) her internal neural physical states. Clark and Chalmers emphasise that Otto has the same mental state as Inga; he also has the (dispositional) belief that 'MoMA is on the 53rd street.' The only difference is that his belief is realised by physical states spread out in his cognitive system and his notebook.

The ideas that motivate the extended cognition thesis are the same that motivate the extended mind proposal. *Metaphysical cognitive integration* (MCI) allows external physical vehicles to realise mental states, just as well as it allows external vehicles to realise cognitive states. After all, mental states supervene

on cognitive states. And, MCI can be confirmed by the way an external process is integrated, i.e. via the Glue and Trust criteria, the *functional parity* (Clark and Chalmers 1998) between the external process and our innate processes, and the *continuous and reciprocal exchange* (CRC) (Clark 2008; Palermos 2014b) of information between the new process and the agent's cognitive system. The integration of Otto's notebook is also facilitated because there is functional parity between how Otto uses his notebook, and Inga operates her innate process. The notebook meets the three Glue and Trust conditions as well. He carries it around everywhere, trusts the information in the notebook, and can access this information readily. And if this is not enough to guarantee that Otto's notebook is integrated into his cognitive system, note that he is also continuously and reciprocally exchanging information with his notebook that allows his process of manipulating his notebook to couple with his cognitive system.

The Alexa case is also a case of extended mind like the Otto-Inga case. Like Otto's dispositional belief is realised by his cognitive system employing the notebook as an external vehicle, Alexa partially realises my belief. I can form, store, retrieve my dispositional belief by employing Alexa a certain way. My process of employing Alexa for day-to-day tasks is coupled with the rest of my cognitive system because it is trustworthy, readily available and easily accessible for me. And, there is a constant exchange of information between my employment of Alexa and the rest of my cognitive character. My employment of Alexa is also functionally equivalent to how I employ my neural processes. Otto's mental state is at least partially realised by a physical state that is situated outside his body, i.e. his notebook. And, my mental state that 'there are no eggs left in the fridge' is also partially realised in Alexa.

So, what is it that allows us to attribute this belief to Otto? Before we can discuss this question, let's consider different forms of believing. The literature on forms of believing will then be used, in the section after the next one, to understand why we attribute beliefs the way we do, and to relate these concepts to the Alexa case.

## Forms of believing

The property of believing is dispositional (Audi 2015), or, in other words, to believe something is to have a disposition. Most of our beliefs are present in us in a kind of dormant form, and they become (sort of) activated in appropriate contextual circumstances. For example, I have the belief that 'hedgehogs are very cute.' This belief is not present to my awareness at all times. But, when a friend mentions hedgehogs or when I see one, my belief manifests in my occurring thought and I am reminded of how I believe hedgehogs are very cute.

You can approach this concept by simply asking what a disposition is. It is a property of an object that manifests itself in specific conditions. For example, sharpness is a disposition of a knife. When a knife is used on a cake, it cuts it. This is because its disposition, i.e. sharpness, manifests itself. But, a disposition may not always be manifested (or is at least manifested in a different way than occurrently); as in, the knife's sharpness may not be able to cut a stone in half. So, to have a disposition means to have a property, like sharpness, that manifests itself in specific conditions. Believing is like such a dispositional property.<sup>89</sup>

Sometimes believing manifests itself in occurring thought, and those beliefs are called *occurrent beliefs*. Example of an occurrent belief is when my friend mentions hedgehogs and I think of my belief that 'hedgehogs are very cute.' But, most of the beliefs I have are not presently occurring in my thought. For example, I believe that the sun is in the centre of the solar system, the moon is very far away from the earth, the tap water in Pakistan is undrinkable, there are no approved vaccines for the coronavirus, etc. When these beliefs are not being manifested in my occurring thought, they are my *dispositional beliefs*. I am not conscious of them, but I still have these beliefs. In short, to believe something does not require awareness of the belief.

Audi explains the difference between these two forms of beliefs in terms of a computer analogy that I find very helpful. An occurrent belief is like a representation that is on the computer screen, whereas dispositional beliefs are

---

<sup>89</sup>For an in-depth understanding of beliefs and different forms of believing, look at Schwitzgebel (2019).

like representations stored in the computer's memory system. With a simple click of the button, the things stored in the memory can be brought to the computer screen. That is, by a simple retrieval process, dispositional beliefs can occurrently manifest in our thoughts.

We also have *dispositions to believe* (Audi 2015, 1994). These are propositions that we are disposed to believe but do not yet believe because they have not undergone the process of belief-formation. For example, I am disposed to believe that I have never been to the Republic of Nauru, even if I have never heard of this country before. Because, as soon as I think of this proposition, I will realise this belief. Similarly, I am also disposed to believe that there are more than two words in this sentence (Audi 1994). I can very easily not form this belief (or this belief may not be realised by my physical states) but I am disposed to believe it.

Audi (Audi 2015) explains that we are disposed to believe many things that do not become our beliefs (neither occurrent nor dispositional). For example, I have more than three hair on my head, there are more than four molecules in the air, many people are using the Internet right now, etc. We do not form or realise all these beliefs that we are disposed to form. But these dispositions are almost beliefs, the grounds (or basis) for these beliefs are already there, we only need to think of them to realise them. Our dispositions to believe can inform the process of belief-formation. I am going to describe this process in terms of grounds and realisation.

Consider what these grounds may be like. Grounds may be other beliefs that the dispositions to believe can be inferred from. For instance, my disposition to believe that I have never been to the country Nauru is grounded in a belief and can be inferred from it, i.e. I have never heard of Nauru. Other important grounds are sensory, introspective, and memorial experiences. 'Consider seeing. We need not form beliefs about everything we see. But seeing something disposes one to form various beliefs about it' (Audi 2015, 16). The grounds for the belief that there are more than 3 hairs on my head is already present in me, i.e. there are a lot of hairs on my head. When I appropriately consider this proposition, for instance, when I am asked if I have 3 hairs on my head, I will form this belief.

Further, let us discuss what it means for a belief to be realised. The grounds of our dispositions to believe are already present in our cognitive states; what is absent is the realiser. Once a disposition to believe is realised, it becomes a belief. In most cases, Audi writes that the realiser is simply *thinking of p*, 'even in a less focal way than in entertaining it' (Audi 2015, 42). So, when a disposition to believe is thought of appropriately, it is realised into a belief. The appropriate way, in this case, is to think of *p* in a way that somewhat links the grounds of *p* with *p*. This is sufficient to bring *p* before one's mind and realise the belief.<sup>90</sup>

In term of Audi's computer analogy, our dispositions to believe are neither on the computer screen nor its memory. These are not fully-formed representations yet. However, they can easily be generated from (i.e. they easily follow or are derived from) the representations on the computer screen or saved in the computer's memory. This is similar to saying that the grounds for these beliefs are already present, so they can easily be derived from beliefs already present in one's cognitive system (Audi 2015).

Most of the belief-formation that occurs in us results in occurrent beliefs. However, the important thing to note is that some belief-formation may directly result in dispositional beliefs. Audi writes about this in these words: "occurrence of belief formation apparently does not entail that of an occurrent belief" (Audi 2015, 13). Our dispositions to believe often become our occurrent beliefs when we think of these propositions, but sometimes they may directly become dispositional beliefs. In what follows, I will describe both these accounts.

Our dispositions to believe are mostly realised as occurrent beliefs because when we are *thinking of* a disposition to believe, we can realise it and manifest it in occurrent thought. Often, later this occurrent belief passes into our 'inactive memory' and is realised as a dispositional belief (Audi 2015, 49). For example, when I was six, I read about the sun in my geography book and formed the occurrent mental state that 'the sun is in the centre of the solar system.'

---

<sup>90</sup>Audi's classification helps us understand that we have fewer beliefs than we actually think we possess because we have been conflating dispositional beliefs with dispositions to believe. The way to separate the two is to reflect on the statements and consider how the latter has not undergone belief-formation, whereas the former has.

Similarly, I talked to my friend on the phone yesterday and had the occurrent belief that 'she is getting married next year.' But, I have forgotten about this belief from yesterday and the one I formed when I was younger; these beliefs have passed into my inactive memory. In the right circumstance, I might be able to recall them back to my awareness as occurrent beliefs, but right now, they are realised as dispositional beliefs in me.

Some of our beliefs, on the other hand, are realised directly as dispositional beliefs.<sup>91</sup> Audi gives the following example of the realisation of a dispositional belief:

Such belief formation seems possible where one acquires beliefs in certain perceptual ways without thinking of the propositions thereby believed. While absorbed in conversation, one might come to believe, through hearing a distinctive siren, that an ambulance went by, but without thinking of this proposition or considering the matter. This is the formation of a dispositional belief. (Audi 2015, 14)

So, in some cases, like cases of subliminal perception, what we are disposed to believe is realised (as a dispositional belief) in a way that is different from thinking of  $p$ . Our dispositions to believe directly become our dispositional belief. Consider another subliminal perception example before I move onto unpacking how this transpires.

On some mornings, I have short chats with my neighbour when she is watering her plants. These chats are never about her plants, but I can see her plants when she is watering them. It is possible that even though thoughts about the types of her plants have not occurred in me in a way that is present to my awareness, I still have dispositional beliefs about them. Hence, I might be able to answer what types of plants my neighbour owns if I was asked to reflect on it. This is because while my conscious thought is focussed on our conversations, physical or cognitive states in my cognitive system realise dispositional beliefs about the types of plants she has (I am very keen on plants).

---

<sup>91</sup>Most dispositional beliefs are formed when occurrent beliefs pass onto inactive memory.

Hence, realising, and, therefore, forming a belief, does not necessarily require thinking of  $p$  appropriately so it comes before one's mind. In the case of the ambulance passing by, Audi writes, that the proposition of one's disposition to believe does not come before the mind, but it is 'implanted by the perceptual experience'— i.e. 'the siren (of the ambulance) leaves a doxastic trace' (Audi 2015, 15). Similarly, the mental state about the type of plants my neighbour owns is also implanted in me by the perceptual experience.

Still, however, our dispositions to believe are often conflated with our dispositional beliefs. The reason for that is that they are, usually, both recalled to awareness by thinking of the said proposition. So, when I think of my (already formed) dispositional belief that 'the sun is in the centre of the universe,' it manifests in my occurring thought and becomes an occurrent belief. This phenomenon can be confused with how we realise propositions (or form beliefs) that we are disposed to believe by thinking of them. For example, when I think of the proposition 'there are more than two words in this sentence,' I form the belief that 'there are more than two words in this sentence.' How, then, are dispositional beliefs different from what we are disposed to believe if we become aware of both when we think of them?

The difference lies in understanding that dispositions to believe are not yet fully-formed beliefs. This becomes clear if we consider the earlier example of the Republic of Nauru. If I have never heard of this place, I cannot have a dispositional belief about this place. So, only when I reflect on this proposition can I form the belief that 'I have never been to the Republic of Nauru.'

## Attributing beliefs

The literature on forms of believing and our dispositions to believe can inform our understanding of the Otto-Inga case. Otto and Inga's belief that 'MoMA is on 53rd street' is their dispositional belief. This belief is realised in Inga by the physical states in her neural cognitive system. Otto, on the other hand, uses a resource present in the environment as his inactive memory. And, his dispositional belief that 'MoMA is on 53rd street' is realised by physical states present in his cognitive system as well as the notebook he writes his beliefs in.

For both of our agents, the belief that ‘MoMA is on 53rd street’ is not just something that they are disposed to believe, it is their dispositional belief. This is because Inga and Otto have come across this proposition before and have endorsed it. Otto consciously formed an occurrent belief once that MoMA is on 53rd street and wrote it in his notebook like he does most of his beliefs. Inga has also formed this belief which is why it is already present in her biological memory, and all she has to do is recall it to her consciousness when she needs it.

We already know that Inga’s belief ought to be attributed to her because it is being realised by her neural states. Why should Otto’s belief be attributed to him?<sup>92</sup> One answer is that Otto, like Inga, previously endorsed this belief. He first formed an occurrent belief when he first heard about where MoMA is, wrote it down in his notebook, and, now, the physical state in his cognitive system and the notebook together realise his dispositional belief. When he needs his dispositional belief, he can recall it. Hence, the previous endorsement clause allows us to ascribe the said belief to Otto and also successfully demarcate the epistemic agent (as Otto).<sup>93</sup>

In light of the literature of belief forms and how it applies to the Otto-Inga case, consider the Alexa case once again. Suppose, it is the year 1985 and there is no Alexa in my house— such an AI has not been invented. I have a presentation at work (like in the Alexa case) and I take out the last egg from

---

<sup>92</sup>The discussion in this section is concerned with how we attribute belief. Breyer and Greco (2008) write about what makes an agent own a belief. They propose different accounts of what it means for an agent to own a belief, but their approach to this debate conflates the two integration accounts I have tried to keep separate. The account of belief ownership that they prefer is based in taking responsibility of the belief-forming process. Which means that their concept of belief ownership is associated with epistemic integration instead of metaphysical cognitive integration. More on the concept of belief ownership can be found in Breyer (2010) and (2013).

<sup>93</sup>Wikforss (2014) argues that Otto only has a disposition to believe whatever is there in his notebook and does not have standing dispositional beliefs. Her argument is based on the premise that the beliefs in the notebook are not well connected to the rest of her belief system. My worry is that the dispositional beliefs that are realised entirely by our neural states are also not as well-connected as our occurrent beliefs. And, as much as dispositional beliefs are interconnected with the rest of one’s belief system, Otto’s are too. The idea is that they should keep the cognitive loop moving forward when they manifest in thought.

my fridge for breakfast, unreflectively. According to the literature discussed above, it is still possible that the dispositional belief that 'there are no eggs left in the fridge' has been realised in me. I may not have picked up the last egg in a way that this proposition comes before my mind, but it is still possible that it may have been implanted in my mind by my perceptual experience, like the ambulance case that Audi described.

Now, let's go back (or forward) to the year 2030. I have Alexa to rely on for my day-to-day tasks, and the kind of information exchange there is between Alexa and I fulfils the requirements for extended cognition. The part of my mind that is concerned with awareness is occupied by the presentation I have at work. But, a process that is metaphysically integrated (via the CRC condition) in my cognitive system, i.e. Alexa, processes the information with other faculties in my cognitive system in such a way that it realises a dispositional belief. The dispositional belief is that 'there are no eggs left in the fridge,' and it can be realised by my cognitive system and Alexa working together. The part of my cognitive system involved may not be the part that is conscious to me at present. Even so, since the interaction between my cognitive system and Alexa results in this cognitive state, the dispositional belief can be attributed to Alexa and I (Alexa+I henceforth). Later in this chapter I am going to refer to this phenomenon as the extending agent. My reason for using this concept is to merely show that it is not clear if the cognitive agent in these cases is only the biological agent or if it is possible for the biological agent to extend when new artificial processes are added to it. I will continue this discussion later.

Some might object that 'there are no eggs left in the fridge' is not yet my dispositional belief, it is only something I am disposed to believe. And, only once I receive the notification from Alexa, that I have no eggs left in the fridge, will this belief present before my mind and be realised.<sup>94</sup> Another objection

---

<sup>94</sup>On this related note, consider John Preston (2010) argument that we have a sort of first-person authority on our beliefs. We do not have to find out what we believe. Preston objects that Otto does not have the same kind of first-person authority on his beliefs. He has to remember what he remembers by first forming the belief that the address is in the notebook and then locating the address and forming the belief about the address. Preston assumes that Clark's reply to this worry will be that Otto should be considered a cognitive system together with his notebook. This will show that he is not remembering something, but simply employ-

may be that the Alexa case is not sufficiently like the Otto-Inga case. Otto's dispositional belief was once his occurrent belief; it was previously endorsed by his cognitive agency. In the Alexa case, I never endorsed this proposition before; how can it be a dispositional belief of mine?

In response to the first objection, note that if Alexa is a part of my cognitive system (like my neural processes) and I can directly form dispositional beliefs (i.e. I can form beliefs that I am not aware of), then as an extended cognitive process of mine Alexa can partially contribute to realise the dispositional belief that 'there are no eggs left in the fridge.' So, subpersonal mechanisms in my cognitive system and my extended cognitive process have realised my dispositional belief; my awareness of it is not necessary.

Secondly, the Alexa case is indeed not like the Otto case in that it represents a directly formed dispositional belief. But, we have seen above in Audi's explanation that this is entirely possible. An ambulance passing by can leave a doxastic impression such that I directly realise a dispositional belief that an ambulance passed by. I do not have to be aware of this belief. If we grant this, then we can also see how an extended cognitive process of our cognitive system can also process a doxastic impression that 'there are no eggs left in the fridge.' Consider an Otto in 2030, who has two instead of one extended cognitive process. One extended cognitive process is his notebook that stores his beliefs like the original Otto case. A second automated device, integrated in his cognitive system via CRC,<sup>95</sup> writes some important details in his notebook automatically, even if he is not conscious of it. Otto comes across a board that says MoMA is on the 53rd street. His automated resource writes this down in his notebook, without Otto ever consciously entertaining this information. It can

---

ing his own cognitive process that happens to be outside his biology. Preston responds to this anticipated reply from Clark by pressing that it is the person (Otto) who does the thinking and believing, not the Otto and notebook system. Clark discusses Preston's worry as the Otto Two-Step problem and responds that it adds unnecessary complexity to the case. See more on this debate in Clark (2010b).

<sup>95</sup>I imagine, such an AI-based device will monitor the kind of information I jot down in my diary, and then pick similar information from around my surroundings and automatically write it down in my diary. Further, it can monitor how often I use the information it notes down for me, to adjust its algorithm to find more useful information to record from my surroundings.

be said that Otto still has the dispositional belief about MoMA's address.

To sum up the discussion here, the Alexa case is a case of extended mind just like the Otto case. Both these cases describe how our dispositional beliefs may be realised by structures partially outside in the environment. The only difference is that Otto's dispositional belief was previously endorsed (or was an occurrent belief once) and my dispositional belief was directly formed without any reflection. The belief that 'there are no new eggs left in the fridge' is realised in Alexa and my cognitive system so it should be attributed to me or at least Alexa+I.<sup>96</sup>

## Double and triple extension

In this section, I am going to discuss the rest of the Alexa case, but before I do that I want to present a clarificatory note. The argument that I am making in this chapter, rests on the understanding that the CRC condition provides a metaphysical cognitive integration account that is both necessary and sufficient for cognitive extension. That is, if there is CRC between two systems, they can come together to drive the cognitive loop forward, and they will form a case of cognitive extension.<sup>97</sup>

So far, I have only discussed the possibility of Alexa integrating in my cognitive system via satisfying the CRC condition. Here, I will discuss the possibility of another layer of extension, i.e. if there is CRC between another external vehicle (that operates with AI) and Alexa+I. Suppose, Alexa+I have a continuous and reciprocal exchange relationship with the Amazon Pantry. The information that Alexa has on the plant seeds I want to buy, it shares it with the Amazon Pantry. It also sends the information about the books I want to read to the Amazon Pantry. Similarly, it shares most of the output, that my cognitive system's manipulation of Alexa creates, with the Amazon Pantry. Reciprocally, the Amazon Pantry also keeps Alexa up to date on its new products. It sends information like recommendations for new products that are honed to target the

---

<sup>96</sup>For a related discussion on belief attributability, in cases of extended mind, consult Preston (2010).

<sup>97</sup>Those who do not find CRC necessary and sufficient for cognitive extension will need to be persuaded with a different case. But, similar conclusions can be achieved.

agent's interest. It improves its recommendations based on the information it receives from Alexa. Overall, the Amazon Pantry and Alexa+I share information similar to how Alexa and I share information, i.e. it satisfies the CRC condition. Hence, the Amazon Pantry is probably a cognitive extension of Alexa+I.

On the day I have my presentation at work, Alexa sends the information to the Pantry that there are no eggs left in my fridge and that I require more eggs. The Pantry informs Alexa that 12 new eggs will be delivered to my place by noon. A delivery comes at noon and Alexa unlocks the automatic door to let in the delivery person. I now have 12 new eggs in my fridge. Alexa, an extended part of my cognitive system, is privy to this information. Amazon Pantry, that also has CRC with Alexa and my cognitive system, has this information. Whereas, I am at work and do not have access to this information. However, if my processes of manipulating Alexa and the Pantry are metaphysically integrated in my cognitive system, it is possible that these processes can help realise the belief that '12 new eggs will be delivered by noon.'

I acknowledge that I may be making a leap here. Just because my extended cognitive processes have some information does not mean that I have necessarily formed the belief. Belief formation does not only require that our cognitive processes (external or neural) have certain information, and I understand that. My aim, however, is to only point out that this may be a possibility. We can suppose that whatever is missing in this case of belief formation is present. My extended cognitive processes (Alexa and the Pantry) have the information that new eggs will be delivered by noon and the mental state that '12 new eggs will be delivered by noon' is realised by physical states present in my cognitive system and those present in Alexa and the Pantry.

The belief that '12 new eggs will be delivered by noon' is similar to the belief that 'there are no eggs left in the fridge' in that I am not aware of it. My suggestion again is that this belief can also be a directly formed dispositional belief. Once again, we look towards Audi's ambulance case. He says that an agent can form the dispositional belief that an ambulance passed by, because the ambulance passing by leaves a doxastic impression. In this case, there is some subpersonal information that is present directly to the perceptual faculty; the agent on a personal level does not become aware of her perception. In the

Alexa and Pantry case, I am suggesting that we suppose that such doxastic impressions can be present to Alexa and the Pantry subpersonally, and I do not become aware of them on a personal level. Given this scenario, the Pantry, Alexa, and parts of my cognitive system will realise the dispositional belief that '12 new eggs will be delivered by noon.'

Let's discuss who this belief will be attributed to. The non-occurrent belief that '12 new eggs will be delivered by noon' may be attributed to me (the biological organism) or it may be attributed to the evolved system that now has three components: parts of my cognitive system, Alexa, and the Pantry (i.e. Pantry+Alexa+I system). This question presents a debate that I have reserved a complete section for, and I will delve into it later. For now, I just want to point out that our discussion on belief attributability has narrowed down (in the previous section) as to where the cognitive agent lies, but we have not demarcated the cognitive agent exactly. It is either I, the biological organism, or the extending agent that includes my cognitive system, Alexa, and the Pantry.

Consider one more layer of extension. There is an advertising agency, called Fishbowl, that exchanges information with the Amazon Pantry, the same way I exchange information with Alexa, and Alexa exchanges the data it has with the Amazon Pantry. Fishbowl is a modern advertising AI that works with the Pantry to improve its advertising and marketing for consumers. The Amazon Pantry, since it is also a part of my cognitive system like Alexa, has access to my information and can share it with Fishbowl. In general, it can share details about my life, like where I live, what I am doing, how I feel, etc., and it can share all the information about my purchases and the things I look-up on the Pantry. From this data, Fishbowl determines if and when I will be willing to try a different brand for a particular product I use, or if I will entertain a detailed advertisement of a sample of a new product.<sup>98</sup> It advertises its products at different times and collects data on my mood and my readiness to try new things to improve its algorithms and give me better suggestions.

Let's say that my presentation at work, that I was worried about, went

---

<sup>98</sup>AI-infused advertising is already violating all sorts of privacy laws by monitoring our behaviour online, sharing our online data with other companies, and targeting us individually. Imagining such a link between Amazon and other companies is not difficult.

great. And I came back home relaxed and content with myself. Now, I am sitting on the sofa and watching TV. Meanwhile, new information is generated by the exchange between the Amazon Pantry and Fishbowl (or the Amazon Pantry's employment of Fishbowl). This new information is that I am likely to change my toothpaste brand if I tried a new one right now. Once again, because of the nature of information exchange between these advanced technologies and my cognitive system, one can argue that this proposition is also my dispositional belief. It is being realised by physical states distributed in my cognitive system, Alexa, the Pantry, and Fishbowl. Once again, we know that this dispositional belief should be either attributed to me, perhaps because I am the only living organism in the scene, or it ought to be attributed to my cognitive system that is spread out in my body, Alexa, the Pantry, and Fishbowl.

The information exchange in the first layer of extension, i.e. Alexa and I, seems slightly different than the information exchange in the third layer of extension, i.e. Fishbowl and I. One reason for this impression is that Fishbowl is connected to my neural cognitive system via Alexa and the Pantry. So, Fishbowl has other AI systems, which are (technically) parts of my cognitive system, that connect it to my neural cognitive system. And, my neural cognitive system is still the only part involved in bringing about conscious mental states. That is, consciousness, at least so far, supervenes only on my innate neural states that reside in my biological brain. This means that the third layer of extension (i.e. extension into Fishbowl) is farther away from the things that I can be conscious of, as compared to the first layer of extension (Alexa). Moreover, the way I employ Alexa is slightly different from how Fishbowl is being employed. Ten years ago, I made a conscious and reflective decision of installing Alexa in my home, office, car, and phone. This is not the same as how the Pantry exchanges my information with Fishbowl. It is possible that I am not aware, and have not been aware, of my cognitive extension into Fishbowl. It may have been in the *terms and conditions* agreement that came with Alexa but maybe I did not read that section very well.

However, notice that I am not arguing here that the first and the third layer of extension are different, only that they appear so. And, I am providing some reasons as to why they appear so. Practically, it should not matter if

advanced technology is integrating into my biological cognitive system or if it is integrating into my enhanced cognitive system that has a few acquired technological processes already integrated into it (i.e. Alexa and the Pantry, in this case). Some integration should not seem more distant than others. I am going to elaborate further as to why I think this may not be a problem for metaphysical cognitive integration but is indeed a problem for epistemic integration. The idea is that it matters for epistemic integration if the layers of integration seem like they are further away from the conscious level of the agent. I will unpack these ideas in the next sections.

## The knowledge bloat problem

So far, the discussion has revolved around metaphysical integration (MCI) of processes that exist outside our skin and skull boundary. And, we have seen that the nature of information exchange determines which processes in the environment may be embedded in our cognitive architecture, such that their products, i.e. beliefs, can be attributed to the agent (or the extending agent). So, why a belief ought to be attributable to a cognitive agent is understood in terms of MCI, i.e. what makes a process and its product a part of the agent's cognitive system and attributable to the said cognitive agent.

Now, if knowledge is true belief that results from a process that is appropriately integrated in one's cognitive system, then a case can be made that the true beliefs being formed by my employment of Alexa can be knowledge for me. In this section, I am going to discuss if and when such a claim can be made. I will discuss the Alexa case in terms of epistemic integration. This means that we ought to consider not only how and why beliefs are attributed to a cognitive agent, but, also, what makes knowledge creditable to an epistemic agent. Knowledge, as outlined by virtue epistemologists, is a cognitive success (or true belief) that is at least significantly creditable to one's cognitive ability or the manifestation of one's cognitive agency. For a belief to be attributed to you, it ought to be produced by a cognitive process that is metaphysically integrated into your cognitive system. And, just like that, for knowledge to be creditable to you, it ought to be produced by a process (your cognitive ability) that is

epistemically integrated into your cognitive system.

Let me reiterate the two rival theories of how a reliable belief-forming process develops (or sometimes immediately turns) into a cognitive ability: weak epistemic cognitive integration (ECI) and subpersonal ECI. We have discussed both of them in detail in the previous chapters, so I will outline them only briefly here. Weak ECI requires that the target process should have the kind of information exchange (cooperative interactions), with other processes in the agent's cognitive system, that allows the agent to be sensitive to the (evidence against the) reliability of the said process. On the other hand, Clark's subpersonal ECI suggests that there may be a more minimal understanding of epistemic integration wherein the agent does not have to be personally involved in the epistemic integration. So, the information exchange between the target process and other processes in one's cognitive system will be such that the reliability of the target process will be vouched for (or assigned to the process) subpersonally. The agent, on the personal level, will not gain sensitivity to the counterevidence to her process's reliability in this case. Only the subpersonal mechanism will have access to the reliability of the target process.

Is Alexa a cognitive ability of my cognitive system? On the weak ECI account, a reliable belief-forming process is integrated in our cognitive system when we have developed counterfactual sensitivity to its reliability. This case is slightly odd from the usual cases. Even if the agent is sensitive to the reliability of her target process, she will not be able to exhibit this sensitivity unless she becomes aware of a belief that she produces via the said process. Although, as long as she can easily come to be in the position to exhibit that she is counterfactually responsive to the reliability of her process, we should consider the process epistemically integrated in her cognitive system. Therefore, on the weak ECI account, we may be able to argue that Alexa is a cognitive ability of my cognitive system. I will discuss this premise further at the end of this section and consider why some might not want to grant this.

Contrastingly, subpersonal ECI states that Alexa (because it is an extended cognitive part of me and can epistemically integrate most minimally) is my cognitive ability if my subpersonal mechanisms have assigned it a high reliability weighting (discussed in detail in Chapter 4) (Clark 2013, 2015). Accordingly, I

know the proposition, that ‘there are no eggs left in the fridge,’ entirely based on subpersonal activities in my head and in Alexa. In this sense, subpersonal ECI is the same as MCI. That is, some subpersonal information exchange between parts of my cognitive system and Alexa can bring about metaphysical integration of Alexa into my cognitive system, such that the cognitive states being produced can be attributed to me. Similarly, subpersonal goings-on between Alexa and my cognitive system can bring about epistemic integration as well, and Alexa can be employed as a cognitive ability by my cognitive system. So, entirely subpersonal mechanisms determine the dispositional belief that ‘there are no eggs left in the fridge’ is knowledge for me.

Since knowledge, in such cases of extended cognition, can be determined entirely subpersonally, I can know the proposition that ‘12 new eggs will be delivered by noon’ (realised in Pantry, Alexa, and I) and I can also know that ‘if I tried a new brand right now, I would change my toothpaste.’ On this note, imagine if Alexa, the Pantry, or Fishbowl, had continuous on-going exchange of my information with another mechanism in the environment, and then another, and so on. If the subpersonal alone can confirm epistemic integration, then I may be in the position to know a very large number of propositions, depending on how many advanced technologies have integrated into my cognitive system. While there is nothing wrong in knowing a large number of propositions, it is odd to know things and employ faculties that are this far removed from your conscious self. Hence, I will conclude my argument, here, by suggesting that Clark’s subpersonal minimalist understanding of epistemic integration falls prey to a *knowledge bloat* problem (Palermos 2018; Wikforss 2014).<sup>99</sup>

To unpack knowledge bloat, let’s go back to the concept of why a belief

---

<sup>99</sup>Palermos uses the term knowledge bloat to indicate the worry that follows from a cognitive bloat. The idea is that if cognition can leak perniciously into the environment, then it is also possible that we can know far too many things than is humanly possible. I use ‘knowledge bloat’ the same way Palermos does, i.e. to denote the uncontrollable and problematic expansion of what we know. However, the overall issue that I outline here is different from Palermos’s discussion. This knowledge bloat is a result of the very weak interpretation of epistemic integration and is not as such linked to cognitive bloat. According to this, if epistemic integration is entirely subpersonally determined, we may know innumerable beliefs that are being realised in AI systems and advanced technologies all over the world.

is attributed to someone. Belief attributability is determined by MCI, which is an entirely subpersonal affair. If a certain process is metaphysically integrated into the biological agent (or the extending agent), then the beliefs produced by the said process ought to be attributed to this biological agent (or the extending agent). Clark's subpersonal ECI account proposes that we can also credit an epistemic agent with knowledge the same way we attribute a cognitive agent with belief, i.e. in terms of entirely subpersonal goings-on. So, Alexa, the Pantry, Fishbowl, other advanced technologies and AI systems, and powerful supercomputers far away from us, etc., can all interchange information with my cognitive system (or my extending cognitive system, for example, Alexa+I, Pantry+Alexa+I, etc.) in a way that realises several dispositional beliefs. Similarly, merely subpersonal exchanges (on Clark's subpersonal ECI account) between these processes and my cognitive system may also allow all these dispositional beliefs to be knowledge for me.

Clark's account of subpersonal epistemic integration leads us to a view where we can know a lot of things from employing sources that we have no idea we are employing. Intuitively, it makes sense to think that my employment of Alexa can result in knowledge because I consciously installed the AI in my surroundings for this specific purpose. But, how can Fishbowl be associated in my acquiring knowledge, when I may have no idea that the Amazon Pantry continuously shares my information with Fishbowl to improve its ability to sell new items to me. The Pantry may have an interchange of information with supercomputers in Alaska, and I may be generating knowledge in virtue of their subpersonal interaction. Should this knowledge be creditable to me? On Clark's view, the supercomputer can become my cognitive ability based on entirely subpersonal activity, and the resultant true belief will be creditable to my cognitive agency and will be my knowledge. There seems to be something wrong with this account of epistemic integration.

Note another clarificatory point, knowledge bloat is not the same as *cognitive bloat* (Clark 2001b; Rupert 2004; Rowlands 2009; Lynch 2014) that we discussed in Chapter 2.<sup>100</sup> According to the cognitive bloat worry, cognition

---

<sup>100</sup>I have distinguished cognitive bloating from knowledge bloating. The former I understand as the bloating of belief, and the latter is the bloating or harmful leaking of knowledge into

leaks uncontrollably into the environment. To solve this problem, we need a proper account of why a cognitive process should be considered ours, or why a cognitive product should be creditable to us.<sup>101</sup> This worry was dissolved by the dynamic systems theory (DST)(Thompson and Varela 2001; Palermos 2011, 2016, 2014b, 2017). DST added that the nature of information exchange between two systems can allow it to evolve into one system and produce output as a whole.<sup>102</sup> So, an external mechanism can interact with our cognitive system such that it becomes our cognitive process, and only the external mechanisms that evolve in this way can be considered our cognitive processes. But, in the knowledge bloat problem I mentioned earlier, I am not pointing out that we need a better understanding of what makes a process part of our cognitive system. I am discussing the problem that arises from our understanding of epistemic integration being so weak that it allows all or most cognition to qualify as capable of generating knowledge; which it is not.

In my understanding, Pritchard's weak epistemic integration does not run into this kind of knowledge bloat problem. Weak ECI states that for an agent to be creditable with knowledge, she ought to be responsive to her process's reliability. This immediately limits the external sources that the agent can employ to produce knowledge. Only the processes that are reliable and the agent has sensitivity to their reliability can be employed such that they produce knowledge for the agent. We have discussed before how Alexa is metaphysically integrated in my cognitive system. It is in a continuous reciprocal exchange of information with other processes of my cognitive system. The interconnectedness is such that it not only furthers my cognition but it also allows me to be sensitive to Alexa's reliability. If Alexa forms the belief that I have 12 eggs in my fridge, when I actually picked the last one today and have none left, Alexa is malfunctioning. Given the oddity of the case, I will not

---

the environment.

<sup>101</sup>The cognitive bloat problem was elaborated in terms of the Telo case in Chapter 2. The Telo case demonstrates that our understanding of MCI in terms of solely the glue and trust conditions was very weak. It would allow Telo to know everything in his phone, just because he trusts it and carries it around in the same way that Otto does.

<sup>102</sup>Van Gelder (1995) also discusses the coupling of two systems in terms of the dynamical systems theory.

become aware of this inconsistency in Alexa's output right away, but there are several ways that I will become aware of the issue very soon. I may receive a notification on my phone that I have 12 eggs and no more need to be ordered, and this information may trigger my memory that I took the last egg. Or, I may come home and find the fridge empty when Alexa notified me saying that I have 12 eggs.

Some might argue that this is not in line with Pritchard's weak epistemic cognitive integration, because the agent does not become aware of the problem as soon as the belief is produced. Indeed, this case is somewhat strange since the agent is not aware of when the belief is being produced. In which case, as long as she exhibits sensitivity to the reliability of her process when she becomes aware of the belief, she would manifest sufficient cognitive agency. Accordingly, as soon as she becomes aware of the belief she should be able to detect the inconsistency in her belief. We should allow this for all cases because there is always a chance that the agent will not become aware of the belief that she has just realised. What is important is that as soon as she becomes aware of it, she should be able to detect if there are any problems with it. In strange cases like these, an agent can manifest cognitive agency if she is sensitive to the reliability of her process even if it takes her a little while to exhibit this sensitivity. What matters is that the agent is in the position to counterfactually monitor the reliability of her process, and become aware of any inconsistencies that arise. It is far less important how soon she can do it. She demonstrates sensitivity to the reliability of her process, if she can detect the inconsistency of her belief as soon as she becomes aware of it.

We cannot have this strict condition on weak epistemic integration: that the agent ought to become aware of her malfunctioning process as soon as it forms the unreliable belief. This is because our target process may not be 100% reliable. There is always a chance that it may produce some inconsistent beliefs. Furthermore, our target process may malfunction in a way that it does not produce a very obviously problematic output. For instance, let's suppose that Otto's actual notebook gets mixed up with someone else's notebook. He retrieves MoMA's address from this new notebook. It says MoMA is on 52nd street. We cannot expect Otto to immediately become aware of this

malfunctioning process. But, perhaps he will use it a few more times and will realise that this is not his notebook.

However, if epistemologists still do not want to accept this part of my argument then they can opt for the conclusion that Alexa is not epistemically integrated in my cognitive system and I cannot employ it to acquire extended knowledge. My ultimate goal, in this chapter, was to point out that if Alexa's epistemic integration is understood in the same subpersonal terms as its metaphysical integration, then that can be very problematic. And, I have shown that above. My account, therefore, emphasises that epistemic integration ought to be understood in some way that is connected to the agent's awareness, even if minimally.<sup>103</sup>

I want to make one last comment on this point. Fishbowl is only integrated in my cognitive system if I can be responsive to its reliability. In case Fishbowl malfunctions and it is recruited to form a belief, I should be able to become aware of the inconsistency when I become aware of the belief. Although, if I have no idea that Alexa is sharing my information with Fishbowl in this way, then even when I become aware of the inconsistent output I will not be able to pinpoint which of my processes is unreliable. I need to at least have the concept that my extended cognitive process, Alexa, further connects to the Pantry or Fishbowl in a certain way. So, when I become aware of a belief that is problematic and inconsistent, I can also become aware of the process that is malfunctioning.

## Demarcating the epistemic agent

Another related problem that arises when we understand epistemic integration as an entirely subpersonal phenomenon is that we do not have a proper criterion on how to demarcate the epistemic agent. Perhaps it is not very important to distinguish the cognitive agent for all cognitive tasks, but cognition that is relevant to belief-production needs a clear understanding of where the cognitive agent lies. This is because belief is a personal-level phenomenon— a mental state

---

<sup>103</sup>I understand that the argument from Preston (2010) that discusses that agent does not find out what she believes can be applied here. After all, it does seem like that the agent finds out, when she becomes aware of her belief, what she believes.

that has to be ascribed to a person. It is, therefore, crucial that we know where the cognitive agent lies.

It is time to discuss the issue I put on hold earlier. Recall that I was unsure exactly who to attribute the beliefs to in the Alexa case. Let me give a quick reiteration of the problem. The belief that 'there are no eggs left in the fridge' is being realised by physical states that are present in both Alexa and I. However, Alexa processes information that partially realises the mental state. This means that Alexa is playing a stronger part in realising the belief than my neural states. Why, then, should we attribute the said belief to me when most of the subpersonal occurrence that realised this belief transpired in Alexa?

One reason to think that I am the cognitive agent in the Alexa case is that the dynamic systems theory (DST) and the CRC requirement give an understanding of how two systems evolve into one and produce output as a whole. Although, this phenomenon only explains that two systems merge into one whole unit; it does not give us reason to think that one system belongs to another. That is, we do not have a reason to think that Alexa belongs to my cognitive system, only that part of my cognitive system and Alexa have evolved into a new system, i.e. Alexa+I, to which the belief ought to be ascribed.<sup>104</sup>

I want to bring attention to how this problem with belief attribution carries on into Clark's subpersonal epistemology, mainly because subpersonal ECI, like MCI, is based entirely on subpersonal processes. If knowledge is creditable to an agent based on entirely subpersonal developments then it ought to be attributed to Alex+I as well. The subpersonal mechanism, after all, is occurring in my cognitive system and Alexa. So, if the true belief that 'there are no eggs left in the fridge' is knowledge, then it is creditable to the combined system that is Alex+I. Similarly, if '12 new eggs will be delivered by noon' is knowledge then it should be creditable to Pantry+Alexa+I. And, if 'I will change my toothpaste brand if I tried a new one right now' is also knowledge, then it belongs to Fishbowl+Pantry+Alexa+I. We can perhaps call this the problem

---

<sup>104</sup>This is in line with the argument that Miyazono (2015) presents in reply to Clark's functionalism. According to this article, what follows from Clark's argument 'is not that beliefs of Otto are physically realized in the notebook but rather that the beliefs of the hybrid system consisting of Otto and his notebook are physically realized in the notebook' (2017, 3523).

of the extending agent or the *agential bloat* worry. The worrying aspect of this problem is that we do not have a theory that helps us delineate the epistemic agent that you can accredit knowledge to, especially in cases where it is not apparent what and where are all the parts are that make up the agent.

Clark describes *cognitive agency* in a way that is supposed to demarcate the cognitive agent (i.e. who belief is ascribed to) and the epistemic agent (i.e. who knowledge is creditable to). He writes that cognition is organism-centred, even if it is not organism bound. The cognitive agency always resides and is exhibited by the biological organism.<sup>105</sup> This would be Clark's reason to think that the biological agent is the cognitive (or epistemic agent) agent and should be attributed with beliefs in cases of cognitive extension. He writes:

... in rejecting the vision of human cognitive processing as organism-bound, we should not feel forced to deny that it is (in most, perhaps all, real-world cases) organism-centered. It is indeed primarily (though not solely) the biological organism that, courtesy especially of its potent neural apparatus, spins and maintains (or more minimally selects and exploits) the webs of additional structure that then form parts of the machinery that accomplishes its own cognizing. (Clark 2007, 175–76)

Clark's concept of cognitive agency is the ability to spin, maintain, monitor, select, and exploit external vehicles, and he suggests that these abilities are always exhibited by the neural and innate apparatuses of the biological organism. The exploitation, selection, and monitoring that he ascribes to the biological organism can all be performed by Alexa. Clearly, Alexa (along with my cognitive system) is engaging in some of these pursuits when it engages the Pantry (and then the Pantry employs Fishbowl the same way).<sup>106</sup>

---

<sup>105</sup>Clark does not claim that agency extension is impossible, neither does he argue for this conclusion. He claims in parentheses and a footnote in Clark (2007) that in almost all real-world cases cognition is organism-centered.

<sup>106</sup>Robert Wilson also argues for the conclusion that one's cognitive agency lies in one's biology, i.e. where the 'locus of control' is present (2014, 26). Wilson discusses *Lethocerus*, a

Agency is a personal-level phenomenon. It is something we manifest when we exhibit intentional action or initiation of action. What it means for us to be agents is that we can perform actions, and the manifestation of this capacity is called *agency* (Michela 2010). The term agency is used to refer to many different kinds of intentional actions, like epistemic, cognitive, moral, etc. Moral agency, for instance, prompts a moral agent to make a moral judgment based on some idea of right and wrong. Similarly, when we act in a way to fulfil our epistemic goals or act on our beliefs, we express our epistemic agency. The cognitive agency, that we are concerned with here, is manifested when we, or our cognitive system, performs a cognitive task.

We manifest cognitive agency in ways that are conscious to us and in ways that are not conscious but are consciously accessible to us. And, the weakest formulation, that Clark argues is possible, is the exhibition of cognitive agency in a way that is not even consciously accessible to us. Consider these categories in cases of knowledge. Knowledge, as you may recall, is a cognitive success that one achieves when one manifests at least significant cognitive agency. Epistemic integration accounts elaborate on how an agent manifests agency, and why, then, should she be creditable with knowledge. Clark takes subpersonal ECI to be the most minimal expression of cognitive agency that can bring us knowledge. Further, this manifestation of agency is not consciously accessible to the cognitive agent. For instance, I may be expressing (significant) cognitive agency, that I have no reflective access to, when the subpersonal exchange in Fishbowl and other parts of my cognitive system form the belief that 'I will change my toothpaste brand if I try a new one right now.' And, this proposition may be knowledge for me. Similarly, I may be manifesting sufficient cognitive agency and know innumerable propositions (via several belief-forming processes spread out into the environment) grounded in entirely subpersonal exchanges. This is the same knowledge bloat problem that I discussed in the last section.

---

giant water beetle, that digests its prey outside its body. His idea is that even though the bug digests outside of his body, the process is still considered his because he is the locus of control. Furthermore, he reasons that many of our actions go beyond our skin and skull boundary, i.e. they manifest in the environment. But, they still remain our actions. The issue is that the locus of control can shift if we bring forward an artificially intelligent system like Alexa, who can help you with controlling your environment when you are not paying any attention to it.

However, another complication that goes hand-in-hand with this problem is that if agency manifestation is not linked to what the agent can be conscious of, then there is also no way of demarcating where the agency lies. If entirely subpersonal mechanisms can exhibit it, then it may be in my cognitive system, or Alexa and I, or Alexa and other gadgets and I, etc.

The conclusion I am drawing is that at least in cases of epistemic integration, cognitive agency has to be exhibited in a way that is consciously accessible to the agent, even if the agent does not become conscious of it. It is because Clark's epistemic integration approach does not include this condition that it makes it difficult for us to locate the epistemic agent that deserves credit for the knowledge produced. And, to counter this problem, Clark has to force the proposition that cognitive agency is always found in the biological organism. We need better reasons as to why this is the case.

Once again, the weak epistemic integration account does not run into trouble in delineating the epistemic agent. This epistemic integration explains the most minimal notion of cognitive agency (on which knowledge can be creditable to an agent) as one that may not always be conscious to the agent, but it is at least consciously accessible to the agent. Weak ECI demands that the agent should be in a position to become aware of any problem with the reliability of her process. The agent is in a place of potential awareness, which means that she may not be aware of her expression of agency but can become aware of it. Since, the manifestation of cognitive agency is linked with awareness, albeit weakly and indirectly, it is easy to demarcate the epistemic agent.<sup>107</sup>

The epistemic agent is the one who can become aware of her

---

<sup>107</sup>Someone might point out here that this method only helps us delineate epistemic agency and not cognitive agency. In other words, weak ECI pinpoints who the knowledge is creditable to; it does not identify who the belief should be attributable to. This is a legitimate worry. I understand epistemic agency as a subcategory of cognitive agency. And, my proposition solely suggests that the epistemic agent can be demarcated properly, in cases of extended cognition (extended knowledge, more specifically), if we understand epistemic integration in the way that weak ECI account recommends. I am not sure how beliefs ought to be attributed in cases where epistemic integration has not taken place, and only metaphysical integration has taken place. But, I also think that there is not as urgent a need to delineate that sort of agency.

malfunctioning cognitive process (and do something about it), and, therefore, knowledge should be creditable to her. If I am sensitive to Alexa's reliability, and the true belief that 'there are no eggs in the fridge' is knowledge, then it ought to be creditable to me. The same is true when the belief is being realised by Fishbowl, the Pantry, Alexa, and my cognitive system. The pivotal question is: am I sensitive to the reliability of Fishbowl? If not, then it has not epistemically integrated into my cognitive character, even though it is still possible that Fishbowl may have metaphysically integrated into my cognitive character. In other words, Fishbowl can extend my cognition, without also developing into a process that can extend my knowledge as well.

## Conclusion

I have shown in this chapter that a solely subpersonal epistemology falls prey to two serious problems. If epistemic integration ought to be understood in terms of nothing more than a subpersonal exchange of information between artefacts and parts of our cognitive system, then it will become difficult to link knowledge to the agent or be able to delineate where the agent lies. So, my conclusion is the same as in previous chapters. Clark's proposition of a wholly subpersonal epistemic integration is not helpful for our understanding of knowledge. The hypothesis of extended cognition can benefit from a proper account of epistemic integration. Furthermore, the right epistemic integration account will share and assist the concept that external artefacts can be employed by our cognitive agency in the same way that our innate processes are recruited. And, if there is an account of how external artefacts help us achieve knowledge, then it is obvious that cognition can extend via these external sources as well. The subpersonal epistemic integration approach has many issues to address before it can benefit the extended cognition theory in this way. The weak epistemic integration account, meanwhile, fits snugly and corroborates cognitive extension.

## Conclusion

This is a thesis on the subject of extended knowledge. I have tried to accomplish two main aims here. The first is to show that a need for a subpersonal epistemology to explain extended knowledge does not arise. The second aim is to argue that the subpersonal epistemology based on Clark's subpersonal epistemic integration account does not provide a thorough understanding of extended knowledge.

I begin with two integration accounts, one that explains what makes a process our cognitive process, i.e. metaphysical cognitive integration, and the other that characterises an epistemic process, i.e. epistemic integration. We are also told that there is a relationship between these two integration accounts, and this relationship is what explains how an extended cognitive process can also help us acquire knowledge. Of these two relationships, one is a symmetric entailment. It argues that the conditions required to make a reliable belief-forming process cognitive are the same conditions that make an external vehicle an extended cognitive process of one's cognitive system. The symmetric entailment suggests that all cases of extended cognition are also cases of extended knowledge. This claim is false. Hence, our understanding of metaphysical integration and epistemic integration ought not lead us to a symmetric entailment.

The other relationship between metaphysical and epistemic integration is called the asymmetric entailment. It proposes that there is some overlap between the conditions of metaphysical and epistemic integration, but the overlap is not complete. Accordingly, not all cases of extended cognition are necessarily cases of extended knowledge. This is because some conditions that allow metaphysical cognitive integration do not also lead to epistemic integration of a target process.

Clark's account of *subpersonal epistemic integration* suggests a symmetric entailment with metaphysical integration. That is, if the target process meets requirements for extension, it also meets the minimal subpersonal conditions required for the said process to be epistemic. I take issue with this understanding. Firstly, this view suggests that there is always an extended

epistemic process where there is an extended cognitive process. Secondly, Clark puts forward his subpersonal epistemic integration approach in lieu of Pritchard's *weak epistemic integration* account. He argues that the weak epistemic integration is not an appropriate account to make sense of extended cognitive abilities (i.e. knowledge conducive reliable belief-forming processes that can generate knowledge for the agent).

The reasoning behind Clark's argument is that the weak epistemic integration account involves an active agential role, whereas the role of agency in extension is passive and entirely subpersonal. Extension requires a sort of seamless dependence on an external vehicle. And, while weak epistemic integration is not a very active account of incorporating a reliable belief-forming mechanism in one's cognitive system, it does necessitate that the agent gain a counterfactual sensitivity to the reliability of her process (to epistemically integrate it). Clark demands an epistemic integration account that does not require such personal level (even counterfactual) involvement of one's agency. Therefore, he suggests an epistemic integration that is entirely subpersonal.

Clark's criticism makes sense if we are necessarily looking for a symmetric entailment between the two integrations. There is no need to look for such an entailment. The relationship between metaphysical integration and epistemic integration can be understood via the asymmetric entailment. Accordingly, *some* conditions (not all) that are required for metaphysical integration or extension are also required for epistemic integration. I elaborate in Chapter 1 and 2 that these conditions are interconnectedness, continuous cooperation, and non-linear interaction between the candidate process and other processes in one's cognitive system. These features are required for both types of integration. However, weak epistemic integration requires more than just continuous and reciprocal cooperation amongst processes, it requires a certain degree of interconnectedness that is sufficient for the agent to monitor her target process's reliability in the background. Of course this means that the conditions for metaphysical integration do not entirely overlap with the requirements for epistemic integration, but this poses no serious problem. Thus, Pritchard's weak epistemic integration account still snugly fits with extended cognition and gives insight into the concept of extended knowledge.

In chapter 1 and 2, I set the scene for this debate and I argue that if we accept the asymmetric entailment, then the need for a subpersonal epistemic integration does not arise. In chapter 3, I highlight that Clark's subpersonal epistemic integration is at odds with Pritchard's weak epistemic integration. Clark envisions that modern technology will be able to immediately (subpersonally) integrate in our cognitive systems. That is, we will soon have such advanced gadgets that the process of manipulating them will immediately (as in, the first time we encounter them) epistemically integrate in our cognitive characters. I have shown how even the most immediate of weak epistemic integration, i.e. the epistemic integration of innate processes after birth, is not consistent with Clark's conception of immediate integration.

In chapter 4, I ground Clark's subpersonal epistemic integration in his predictive brain model. According to the predictive processing account of how perception and cognition come together, subpersonal epistemic integration occurs at all levels of the brain's functional hierarchy. It is characterised by the function of precision estimation units. These units operate at all levels, i.e. on levels that are concerned with forming beliefs and also on levels that deal with processing only subdoxastic information. They assign reliability weightings to target processes on these different levels, and based on these reliabilities, the said processes are either recruited or ignored. After unpacking the elaborate picture of Clark's subpersonal epistemic integration, I question how it can lead to beliefs on some levels and subdoxastic information on other levels. If the precision estimation units are an indication of epistemic integration, why are they associated with levels that have nothing to do with belief-formation? I present another argument against subpersonal epistemic integration in this chapter. I use a case in which an integrated process suddenly becomes unreliable to show that the subpersonal integration account directs us towards only one possible personal level scenario, which is evidently the same one that the weak epistemic integration validates. This is the same scenario in which the agent becomes aware of any counterevidence to the reliability of her process. I suggest, on this point, that the subpersonal integration is too tightly wedded to the personal level scenario it shares with Pritchard's weak epistemic integration account. We need to be able to devise a case in which when a subpersonally

integrated process suddenly becomes unreliable, the agent does not become aware of the sudden inconsistency in her belief. Such an agent will not be able to use her process (epistemically) responsibly.

Further, in chapter 5, I discuss epistemic responsibility in terms of epistemic defeasibility (how an epistemic agent deals with defeaters). How an epistemic agent deals with defeaters against her beliefs provides insight into how responsibly she has formed those beliefs. In this chapter, I outline the defeasibility theory that describes weak epistemic integration. I also try to devise a subpersonal defeasibility account to evaluate if Clark's subpersonal epistemic integration can provide us a concept of epistemic responsibility. I formulate two subpersonal defeasibility approaches but find them both inadequate in describing an epistemically responsible agent.

In chapter 6 I bring worries for the implication of Clark's subpersonal epistemology. I present a strange case of an epistemic agent, who lives with an AI assistant in 2030. The AI has been with the agent for nearly 10 years. They continuously exchange information throughout the day in a way that satisfies the condition for metaphysical cognitive integration. That is, I suggest that the AI is an extended cognitive process of the agent. Problems arise when I propose that we can imagine a mental state that is partially realised by neural states of the agent and partially by the physical states of the AI. Also, I add to this mix that this mental state is a directly formed dispositional belief. That is, the agent formed it subliminally and is not conscious of it. I, then, ask questions like who should this belief be attributed to and why? How can this directly formed dispositional belief be a case of knowledge? Furthermore, whose knowledge will it be? Then, I try to make sense of this case in terms of Clark's subpersonal epistemic integration and find myself in a difficult situation. If our extended cognitive abilities can develop in ways that are not even potentially consciously accessible to us, then we can run into situations where we know so much simply on the basis of subpersonal exchanges between our neural processes and vehicles in the environment. On the same note, it is difficult to demarcate the epistemic agent (who knowledge is creditable to) if we are to understand epistemic integration solely subpersonally.

## Areas for improvement

I am aware that my thesis can be improved on several fronts. Here, I will briefly describe some of the areas where I can develop my arguments better.

Firstly, I have not taken into account any other completely subpersonal epistemologies. My thesis has only evaluated the subpersonal epistemic integration account explained in terms of the precision estimation units. The way the function of these units fits into the predictive brain model makes it inefficient to account for epistemic responsibility. Also, the mechanism of these units is too ubiquitous to grasp epistemic responsibility, as I described in chapter 4. But, there is a possibility that some other understanding of subpersonal integration can explain epistemic responsibility without falling prey to these problems. One place to look for such a notion would be to comprehend Clark's view as a sophisticated process reliabilism, instead of a virtue epistemology. Clark even directs us to treat his epistemology in terms of a 'second-order reliabilist intuition' ((2015), 3770). He points out that this would be in line with proposals from Goldman (2012) and Kornblith (2012). To investigate Clark's view in this light, I would examine the defeasibility theories associated with these epistemologies and analyse how they fare in terms of explaining epistemic responsibility.

Secondly, my treatment of the Alexa case, in chapter 6, has a lot of room for improvement. I tried to show through Alexa that if we understand the generation of knowledge in the same subpersonal terms as the formation of belief, then we erase the *knower*. For this, I had to formulate a case in which the epistemic agent is not conscious of some of her mental states being realised. And, in cases of cognitive extension, these states are partially realised outside the agent's body. If the agent does not even have potential conscious access to the external processes she is employing (or epistemically integrating), which is the case in subpersonal epistemic integration, then how can she have knowledge of these beliefs? This case is supposed to invoke an intuition about knowledge, i.e. that even if the agent is not conscious of her belief-forming process, it should still be potentially (consciously) accessible to her. Specifically, can an agent acquire knowledge via a belief-forming process she does not even have

potential conscious access to? And, if that were possible, would it not follow that the agent may be employing several processes (and forming beliefs and acquiring knowledge through them), spread all over the place, that she has no potential conscious access to.

However, the initial setting of my case is less than foolproof. I think more work needs to be put in to establish that Alexa2030 is indeed a case of cognitive extension and extended mind. Also, I need to ground my assumption that an extended cognitive AI can perform the function that our subliminal perception can and, therefore, help realise a mental state about something that I was not conscious of (i.e. the eggs in the fridge that I did not become aware of). I have simply assumed that something like this can happen because it is merely a result of subpersonal goings-on. It is indeed possible that external processes cannot engage in subliminal perception in the way our innate faculties can. This is something I will look into when I develop this case further.

Thirdly, in chapter 6, I discuss that the continuous reciprocal causation (CRC) requirement (and the Dynamic Systems Theory) is supposed to help us understand why beliefs are attributable to us in cases of extension. For instance, if a process X is metaphysically integrated in my cognitive system via CRC, then the beliefs that are being partially realised by my deployment of X can be attributable to me. However, CRC only describes the nature of information processing between two systems that allows them to temporally evolve into a hybrid system and produce output as one (whole). The CRC requirement does not indicate why the output of the hybrid system should belong to one original system instead of the other. That is, why should the resultant belief be attributed to me instead of X? What about me allows the belief to be attributable to me? I believe there is a lot of room for examination in the concept of belief attributability. And, following this route, it can also be argued that the CRC is insufficient in describing metaphysical cognitive integration, since it does not describe why the output is attributed to one of the systems. It is possible that the demarcation of epistemic agency can be dealt with at the level of belief attributability, provided we take a different account of metaphysical integration to grasp why beliefs are attributed to us in cases of extension.

Furthermore, I am also interested in considering Clark's subpersonal integration in terms of the value of knowledge debate— value of extended knowledge, to be specific. This debate considers why knowledge is more valuable than true belief. Clark's subpersonal epistemic integration puts generation of knowledge at the same place as the formation of beliefs, i.e. the subpersonal realm. In other words, on Clark's view, knowledge can be creditable to an agent based entirely on the subpersonal goings-on in the agent's cognitive system, i.e. just the way belief is attributable to the agent. This removes an important distinction between how beliefs are formed and how knowledge is generated. As a result, we are further away from answering why (extended) knowledge is more valuable than (extended) belief.

## References

- Adams, Frederick R., and Kenneth Aizawa. 2008. *The Bounds of Cognition*. Malden, MA: Blackwell Pub.
- Aizawa, Kenneth. 2018. "Extended Cognition, Trust and Glue, and Knowledge." In *Extended Epistemology*, edited by Duncan Pritchard, Andy Clark, J Adam Carter, Orestis Palermos, and Jesper Kallestrup. Oxford, UK; New York, NY: Oxford University Press. <https://doi.org/10.1093/oso/9780198769811.003.0004>.
- Audi, Robert. 1994. "Dispositional Beliefs and Dispositions to Believe." *Noûs* 28 (4): 419. <https://doi.org/10.2307/2215473>.
- . 2015. *Rational Belief: Structure, Grounds, and Intellectual Virtue*. Oxford ; New York, NY: Oxford University Press.
- Baehr, Jason. 2012. "Credit Theories and the Value of Knowledge." *The Philosophical Quarterly* 62 (246): 1–22.
- Bartels, Andreas, and Semir Zeki. 2004. "The Neural Correlates of Maternal and Romantic Love." *NeuroImage* 21 (3): 1155–66. <https://doi.org/10.1016/j.neuroimage.2003.11.003>.
- Bennett, M. R., and P. M. S. Hacker. 2003. *Philosophical Foundations of Neuroscience*. Malden, Mass., USA ; Oxford: Blackwell.
- Bergmann, Michael. 1997. "Internalism, Externalism and Epistemic Defeat: PhD Thesis." University of Notre Dame.
- . 2005. "Defeaters and Higher-Level Requirements." *The Philosophical Quarterly* 55 (220): 419–36. <https://doi.org/10.1111/j.0031-8094.2005.00408.x>.
- . 2006. *Justification Without Awareness*. Oxford University Press. <https://doi.org/10.1093/0199275742.001.0001>.
- BonJour, Laurence. 1985. *The Structure of Empirical Knowledge*. Cambridge, Mass.: Harvard Univ. Press.
- Bonjour, Laurence. 1980. "Externalist Theories of Empirical Knowledge." *Midwest Studies In Philosophy* 5 (1): 53–74. <https://doi.org/10.1111/j.1475-4975.1980.tb00396.x>.
- Breyer, Daniel. 2010. "Reflective Luck and Belief Ownership." *Acta Analytica*

- 25 (2): 133–54. <https://doi.org/10.1007/s12136-010-0087-8>.
- . 2013. “Ownership, Agency, and Defeat.” *Acta Analytica* 28 (2): 253–56. <https://doi.org/10.1007/s12136-012-0175-z>.
- Breyer, Daniel, and John Greco. 2008. “Cognitive Integration and the Ownership of Belief: Response to Bernecker.” *Philosophy and Phenomenological Research* 76 (1): 173–84.
- Carter, J. Adam, Andy Clark, Jesper Kallestrup, Duncan Pritchard, and Orestis Palermos, eds. 2018. *Extended Epistemology*. First edition. Oxford, UK; New York, NY: Oxford University Press.
- Carter, J. Adam, Andy Clark, and Orestis Palermos. 2018. “New Humans?” In *Extended Epistemology*, edited by J. Adam Carter, Andy Clark, Jesper Kallestrup, and Duncan Pritchard, First edition. Oxford, UK; New York, NY: Oxford University Press.
- Carter, J. Adam, and Jesper Kallestrup. 2018. “Extended Circularity: A New Puzzle for Extended Cognition.” In *Extended Epistemology*, edited by J. Adam Carter, Duncan Pritchard, Andy Clark, Jesper Kallestrup, and Orestis Palermos. Vol. 1. Oxford, UK; New York, NY: Oxford University Press. <https://doi.org/10.1093/oso/9780198769811.003.0003>.
- . 2019. “Varieties of Cognitive Integration.” *Noûs*, May, nous.12288. <https://doi.org/10.1111/nous.12288>.
- Carter, J. Adam, and Orestis Palermos. 2015. “Epistemology and Active Externalism.” Oxford University Press. <https://doi.org/10.1093/obo/9780195396577-0285>.
- Chemero, Anthony. 2009. *Radical Embodied Cognitive Science*. First MIT Press paperback edition. Cambridge, Mass. London: MIT Press.
- Chisholm, Roderick M. 1977. *Theory of Knowledge*. 2d ed. Prentice-Hall Foundations of Philosophy Series. Englewood Cliffs, N.J: Prentice-Hall.
- Clark, Andy. 2001a. *Being There: Putting Brain, Body, and World Together Again*. 1. paperback ed. A Bradford Book. Cambridge, Mass.: MIT Press.
- . 2001b. “Reasons, Robots and the Extended Mind.” *Mind & Language* 16 (2): 121–45. <https://doi.org/10.1111/1468-0017.00162>.
- . 2007. “Curing Cognitive Hiccups: A Defense of the Extended Mind.” *The Journal of Philosophy* 104 (4): 163–92.

- . 2008. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Philosophy of Mind. Oxford ; New York: Oxford University Press.
- . 2010a. "Coupling, Constitution, and the Cognitive Kind: A Reply to Adams and Aizawa." In *The Extended Mind*, edited by Richard Menary, 81–99. Cambridge, MA: The MIT Press. <https://doi.org/10.7551/mitpress/9780262014038.003.0005>.
- . 2010b. "Memento's Revenge: The Extended Mind, Extended." In *The Extended Mind*, edited by Richard Menary, 43–66. Cambridge, MA: The MIT Press. <https://doi.org/10.7551/mitpress/9780262014038.003.0003>.
- . 2013. "Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science." *Behavioral and Brain Sciences* 36 (3): 181–204. <https://doi.org/10.1017/S0140525X12000477>.
- . 2015. "What 'Extended Me' Knows." *Synthese* 192 (11): 3757–75. <https://doi.org/10.1007/s11229-015-0719-z>.
- . 2016a. "Attention Alters Predictive Processing." *Behavioral and Brain Sciences* 39: e234. <https://doi.org/10.1017/S0140525X15002472>.
- . 2016b. "Busting Out: Predictive Brains, Embodied Minds, and the Puzzle of the Evidentiary Veil: Busting Out: Predictive Brains, Embodied Minds, and the Puzzle of the Evidentiary Veil." *Noûs*, April. <https://doi.org/10.1111/nous.12140>.
- . 2017. "How to Knit Your Own Markov Blanket: Resisting the Second Law with Metamorphic Minds." <https://doi.org/10.15502/9783958573031>.
- Clark, Andy, and David Chalmers. 1998. "The Extended Mind." *Analysis* 58 (1): 7–19.
- Colombetti, Giovanna, and Tom Roberts. 2015. "Extending the Extended Mind: The Case for Extended Affectivity." *Philosophical Studies* 172 (5): 1243–63. <https://doi.org/10.1007/s11098-014-0347-3>.
- Dennett, Daniel C. 1969. *Content and Consciousness*. London: Routledge.
- . 2007. "Philosophy as Naive Anthropology: Comment on Bennett and Hacker." In *Neuroscience and Philosophy: Brain, Mind, and Language*, edited by Dennett Bennett M. Columbia University Press.
- Drayson, Zoe. 2012. "The Uses and Abuses of the Subpersonal/Personal Distinction." *Philosophical Perspectives* 26 (1): 1–18. <https://doi.org/10.1>

- 111/phpe.12014.
- Farkas, Katalin. 2012. "Two Versions of the Extended Mind Thesis." *Philosophia* 40 (3): 435–47. <https://doi.org/10.1007/s11406-011-9355-0>.
- Friston, Karl. 2009. "The Free-Energy Principle: A Rough Guide to the Brain?" *Trends in Cognitive Sciences* 13 (7): 293–301. <https://doi.org/10.1016/j.tics.2009.04.005>.
- Gertler, Brie. 2007. "Overextending the Mind?" In *Arguing About the Mind*, edited by Brie Gertler and Lawrence Shapiro, 192–206. Routledge.
- Gettier, Edmund L. 1963. "Is Justified True Belief Knowledge?" *Analysis* 23 (6): 121–23. <https://doi.org/10.1093/analys/23.6.121>.
- Goldberg, Sanford. 2018. *To the Best of Our Knowledge: Social Expectations and Epistemic Normativity*. First edition. Oxford, United Kingdom: Oxford University Press.
- Goldman, Alvin I. 1986. *Epistemology and Cognition*. Harvard University Press.
- . 2012. *Reliabilism and Contemporary Epistemology: Essays*. New York, NY.: Oxford Univ. Press.
- Greco, John. 1999. "Agent Reliabilism." *Nous* 33 (s13): 273–96. <https://doi.org/10.1111/0029-4624.33.s13.13>.
- . 2003. "Further Thoughts on Agent Reliabilism: Replies to Cohen, Geivett, Kvanvig, and Schmitt and Lahroodi." *Philosophy and Phenomenological Research* 66 (2): 466–80. <https://doi.org/10.1111/j.1933-1592.2003.tb00275.x>.
- . 2010. *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511844645>.
- . 2012. "A (Different) Virtue Epistemology." *Philosophy and Phenomenological Research* 85 (1): 1–26.
- Greenwood, Jennifer. 2013. "Contingent Transcranialism and Deep Functional Cognitive Integration: The Case of Human Emotional Ontogenesis." *Philosophical Psychology* 26 (3): 420–36. <https://doi.org/10.1080/09515089.2011.633752>.
- Gross, Steven, and Georges Rey. 2012. "Innateness." In *The Oxford Handbook of Philosophy of Cognitive Science*, edited by Eric Margolis, Richard Samuels,

- and Stephen P. Stich. New York: Oxford University Press.
- Hohwy, Jakob. 2013. *The Predictive Mind*. First edition. Oxford, United Kingdom ; New York, NY, United States of America: Oxford University Press.
- . 2014. “The-Self Evidencing Brain.” *Noûs* 48 (1).
- Kelp, Christoph. 2013. “Extended Cognition and Robust Virtue Epistemology.” *Erkenntnis* 78 (2): 245–52. <https://doi.org/10.1007/s10670-011-9301-3>.
- Kornblith, Hilary. 2012. *On Reflection*. 1st ed. Oxford: Oxford University Press.
- Lackey, Jennifer. 1999. “Testimonial Knowledge and Transmission.” *Philosophical Quarterly* 49 (197): 471–90. <https://doi.org/10.1111/1467-9213.00154>.
- . 2009. “Knowledge and Credit.” *Philosophical Studies* 142 (1): 27–42. <https://doi.org/10.1007/s11098-008-9304-3>.
- . 2010. “Testimonial Knowledge.” In *Routledge Companion to Epistemology*, edited by Sven Bernecker and Duncan Pritchard, 316–25. New York: Routledge.
- Lehrer, Keith. 1990. *Theory of Knowledge*. Boulder: Westview Press.
- Lynch, Michael Patrick. 2014. “Neuromedia, Extended Knowledge and Understanding.” *Philosophical Issues* 24: 299–313. <https://doi.org/10.2307/26611108>.
- McLaughlin, Brian, and Karen Bennett. 2018. “Supervenience.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2018. Metaphysics Research Lab, Stanford University.
- Menary, Richard. 2006. “Attacking the Bounds of Cognition.” *Philosophical Psychology* 19 (3): 329–44. <https://doi.org/10.1080/09515080600690557>.
- . 2007. *Cognitive Integration: Mind and Cognition Unbounded*. Edited by John Protevi. London, UK: Palgrave Macmillan UK.
- . 2009. “Intentionality, Cognitive Integration and the Continuity Thesis.” *Topoi* 28 (1): 31–43. <https://doi.org/10.1007/s11245-008-9044-1>.
- . 2010. “Introduction: The Extended Mind in Focus.” In *The Extended Mind*, edited by Richard Menary, 226–43. Cambridge, MA: The MIT Press. <https://doi.org/10.7551/mitpress/9780262014038.003.0010>.
- . 2012. “Cognitive Practices and Cognitive Character.” *Philosophical Explorations* 15 (2): 147–64. <https://doi.org/10.1080/13869795.2012>.

- 677851.
- Michela, Balconi, ed. 2010. *Neuropsychology of the Sense of Agency*. Milano: Springer Milan. <https://doi.org/10.1007/978-88-470-1587-6>.
- Miyazono, Kengo. 2015. "Does Functionalism Entail Extended Mind?" *Synthese* 194 (9): 3523–41. <https://doi.org/10.1007/s11229-015-0971-2>.
- Palermos, Orestis. 2011. "Belief-Forming Processes, Extended." *Review of Philosophy and Psychology* 2 (4): 741–65. <https://doi.org/10.1007/s13164-011-0075-y>.
- . 2014a. "Loops, Constitution, and Cognitive Extension." *Cognitive Systems Research* 27 (March): 25–41. <https://doi.org/10.1016/j.cogsys.2013.04.002>.
- . 2014b. "Knowledge and Cognitive Integration." *Synthese* 191 (8): 1931–51. <https://doi.org/10.1007/s11229-013-0383-0>.
- . 2016. "Spreading the Credit: Virtue Reliabilism and Weak Epistemic Anti-Individualism." *Erkenntnis* 81 (2): 305–34. <https://doi.org/10.1007/s10670-015-9741-2>.
- . 2017. "Social Machines: A Philosophical Engineering." *Phenomenology and the Cognitive Sciences* 16 (5): 953–78. <https://doi.org/10.1007/s11097-016-9489-4>.
- . 2018. "Epistemic Presentism." *Philosophical Psychology* 31 (3): 458–78. <https://doi.org/10.1080/09515089.2018.1432035>.
- . n.d. "Virtue Reliabilism and Epistemic Defeat." *Unpublished*, 21.
- Plantinga, Alvin. 1993. *Warrant and Proper Function*. Oxford University Press.
- Preston, John. 2010. "The Extended Mind, the Concept of Belief, and Epistemic Credit." In *The Extended Mind*, edited by Richard Menary, 226–43. Cambridge, MA: The MIT Press. <https://doi.org/10.7551/mitpress/9780262014038.003.0010>.
- Pritchard, Duncan. 2007. "Anti-Luck Epistemology," 22.
- . 2008. "Virtue Epistemology and Epistemic Luck, Revisited." *Metaphilosophy* 39 (1): 66–88.
- . 2009a. *Knowledge*.
- . 2009b. "Apt Performance and Epistemic Value." *Philosophical Studies* 143 (3): 407–16. <https://doi.org/10.1007/s11098-009-9340-7>.

- . 2010. "Cognitive Ability and the Extended Cognition Thesis." *Synthese* 175 (S1): 133–51. <https://doi.org/10.1007/s11229-010-9738-y>.
- . 2016a. *Epistemology*. 2nd edition. Palgrave Philosophy Today. New York: Palgrave Macmillan.
- . 2016b. "Epistemic Dependence." *Philosophical Perspectives* 29 (1): 305–24. <https://doi.org/10.1111/phpe.12067>.
- . 2018a. "Neuromedia and the Epistemology of Education." *Metaphilosophy* 49 (3): 328–49. <https://doi.org/10.1111/meta.12295>.
- . 2018b. "Anti-Luck Virtue Epistemology and Epistemic Defeat." *Synthese* 195 (7): 3065–77. <https://doi.org/10.1007/s11229-016-1074-4>.
- . 2018c. "Extended Virtue Epistemology." *Inquiry* 61 (5-6): 632–47. <https://doi.org/10.1080/0020174X.2017.1355842>.
- Pritchard, Duncan, Alan Millar, and Adrian Haddock. 2010. *The Nature and Value of Knowledge: Three Investigations*. Oxford ; New York: Oxford University Press.
- Proust, Joëlle. 2013. *The Philosophy of Metacognition: Mental Agency and Self-Awareness*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199602162.001.0001>.
- Pryor, James. 2001. "Highlights of Recent Epistemology." *The British Journal for the Philosophy of Science* 52 (1): 95–124.
- Roberts, Tom. 2012. "You Do the Maths: Rules, Extension, and Cognitive Responsibility." *Philosophical Explorations: Extended Cognition and Epistemology* 15 (2): 133–45. <https://doi.org/10.1080/13869795.2012.670724>.
- Ross, Don, and J. Ladyman. 2010. "The Alleged Coupling-Constitution Fallacy and the Mature Sciences." In *The Extended Mind*, edited by Richard Menary, 43–66. Cambridge, MA: The MIT Press. <https://doi.org/10.7551/mitpress/9780262014038.003.0003>.
- Rowlands, Mark. 1999. *The Body in Mind: Understanding Cognitive Processes*. Cambridge Studies in Philosophy. Cambridge, U.K. ; New York: Cambridge University Press.
- . 2009. "Extended Cognition and the Mark of the Cognitive." *Philosophical Psychology* 22 (1): 1–19. <https://doi.org/10.1080/09515080802703620>.
- Rupert, Robert D. 2004. "Challenges to the Hypothesis of Extended Cognition."

- The Journal of Philosophy* 101 (8): 389–428.
- . 2009. *Cognitive Systems and the Extended Mind*. Philosophy of Mind. Oxford ; New York: Oxford University Press.
- Schwitzgebel, Eric. 2019. “Belief.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2019. Metaphysics Research Lab, Stanford University.
- Sosa, Ernest. 1988. “Beyond Scepticism, to the Best of Our Knowledge.” *Mind* 97 (386): 153–88.
- . 1991. *Knowledge in Perspective: Selected Essays in Epistemology*. Cambridge [England] ; New York: Cambridge University Press.
- . 1993. “Proper Functionalism and Virtue Epistemology.” Edited by Alvin Plantinga. *Noûs* 27 (1): 51–65. <https://doi.org/10.2307/2215895>.
- Sprevak, Mark. 2009. “Extended Cognition and Functionalism.” *The Journal of Philosophy* 106 (9): 503–27.
- Spurrett, David, and Stephen Cowley. 2010. “The Extended Infant: Utterance-Activity and Distributed Cognition.” In *The Extended Mind*, edited by Richard Menary, 226–43. Cambridge, MA: The MIT Press. <https://doi.org/10.7551/mitpress/9780262014038.003.0010>.
- Stich, Stephen. 1978. “Beliefs and Subdoxastic States.” *Philosophy of Science* 45 (4): 499–518.
- Thompson, Evan, and Francisco J Varela. 2001. “Radical Embodiment: Neural Dynamics and Consciousness.” *Trends in Cognitive Sciences* 5 (10): 418–25. [https://doi.org/10.1016/s1364-6613\(00\)01750-2](https://doi.org/10.1016/s1364-6613(00)01750-2).
- Tollefsen, Deborah, and Rick Dale. 2012. “Naturalizing Joint Action: A Process-Based Approach.” *Philosophical Psychology* 25 (3): 385–407. <https://doi.org/10.1080/09515089.2011.579418>.
- Van Gelder, Tim. 1995. “What Might Cognition Be, If Not Computation?” *The Journal of Philosophy* 92 (7): 345–81. <https://doi.org/10.2307/2941061>.
- Wallisch, Pascal. 2017. “Illumination Assumptions Account for Individual Differences in the Perceptual Interpretation of a Profoundly Ambiguous Stimulus in the Color Domain: ‘The Dress’” *Journal of Vision* 17 (4): 5. <https://doi.org/10.1167/17.4.5>.
- Ward, D. 2012. “Enjoying the Spread: Conscious Externalism Reconsidered.”

- Mind* 121 (483): 731–51. <https://doi.org/10.1093/mind/fzs095>.
- Wheeler, Michael. 2010. "In Defense of Extended Functionalism." In *The Extended Mind*, edited by Richard Menary, 245–70. Cambridge, MA: The MIT Press. <https://doi.org/10.7551/mitpress/9780262014038.003.0010>.
- . 2018. "Knowledge, Credit, and the Extended Mind, or What Calvisius Sabinus Got Right." In *Extended Epistemology*, edited by Duncan Pritchard, Andy Clark, J Adam Carter, Orestis Palermos, and Jesper Kallestrup. Oxford, UK; New York, NY: Oxford University Press. <https://doi.org/10.1093/oso/9780198769811.003.0004>.
- Wiese, Wanja, and Thomas Metzinger. 2017. "Vanilla PP for Philosophers: A Primer on Predictive Processing," 18.
- Wikforss, Åsa. 2014. "Extended Belief and Extended Knowledge." *Philosophical Issues* 24: 460–81. <https://doi.org/10.2307/26611116>.
- Wilson, Robert A. 2004. *Boundaries of the Mind: The Individual in the Fragile Sciences - Cognition*. Cambridge, UK; New York: Cambridge University Press.
- . 2014. "Ten Questions Concerning Extended Cognition." *Philosophical Psychology* 27 (1): 19–33. <https://doi.org/10.1080/09515089.2013.828568>.
- Witzel, Christoph, Chris Racey, and J. Kevin O'Regan. 2017. "The Most Reasonable Explanation of 'the Dress': Implicit Assumptions about Illumination." *Journal of Vision* 17 (2): 1. <https://doi.org/10.1167/17.2.1>.
- Zagzebski, Linda Trinkaus. 1996. *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. Cambridge, UK; New York: Cambridge University Press.