Epistemic Vices in a Non-Ideal World

Daniella Meehan

Submitted in fulfilment of the requirements for the Degree of
Doctor of Philosophy
Philosophy School of Humanities
College of Arts
University of Glasgow

January 2024

Abstract

Recent developments in epistemology have shifted away from idealised perspectives on knowledge acquisition towards an examination of the myriad of ways in which our epistemic practices go astray. This evolution has given rise to the field of non-ideal epistemology, which explores the realities that emerge when individuals and communities falter in their epistemic practices (Barker et al. 2018; Bernecker et al. 2021; Mckenna 2023). This focus extends across various dimensions of applied and social epistemology, addressing issues such as bad epistemic characters, the erosion of trust, epistemic injustice, ignorance, fake news, and corruption.

A significant manifestation of this recent shift in non-ideal epistemology is evident in the burgeoning field of vice epistemology. Epistemic vices are dispositions, attitudes and ways of thinking that make us bad thinkers, in so far as they prevent us from acquiring and sharing knowledge, manifest bad motives, and desires, or disrupt both individual and collective epistemic functioning (Kidd et al. 2020). These vices are harmful to the vice-bearer in so far as they distort and impair cognitive faculties, leading to flawed reasoning, and biased judgement, hindering the attainment of epistemic goods such as genuine understanding and knowledge (Cassam 2019a; Medina 2012, 2020; Priest 2020). These harms also extend beyond the individual, contributing to the perpetuation of misinformation and the erosion of trust within social networks and communities (Baird and Calvard 2018; Fricker 2020; Medina 2020; Sullivan and Alfano 2020).

The acknowledgement of the social nature of epistemic vices is being increasingly recognised. This recognition underscores that epistemic vices can extend beyond individuals and be held by collectives, including educational institutions, online environments, and prisons (Kidd 2019, 2020; Fricker 2020; Medina 2020; Tanesini 2021).

In light of the harms arising from these bad epistemic practices, inquiries into how to respond to and address these wrongs have been crucial. This has led to considerations of responsibility and ameliorative solutions, seeking to rectify these adverse effects (Cassam 2019a; Holroyd 2016; Sherman and Goguen 2019; Tanesini 2021). These ameliorative solutions also extend beyond individual-based strategies to include structural and social solutions.

This thesis contributes to these growing debates by focusing on the harms associated with epistemic vices and exploring ways to address them.

In Chapters 2 and 3, I focus on some foundational aims in vice epistemology, evaluating three prominent accounts of epistemic vice: obstructivism, motivationalism and personalism (Battaly 2016a, 2018a; Cassam 2016, 2019a; Tanesini 2018, 2021). Within these chapters, I also focus on the harmful nature of epistemic vices and whether vice-bearers should be held responsible for their vices, and if so, what form this responsibility would take. In Chapter 4, I evaluate the role of blame as a response to vice more closely, focusing on its epistemic and ameliorative nature.

In Chapters 5, 6 and 7, I turn to assess themes in applied epistemology. Still focusing on the harms of epistemic vices and possible solutions, I examine whether epistemic nudging, a paternalistic method of nudging individuals towards epistemically desirable outcomes, may assist in the mitigation of epistemic vice (Adams and Niker 2021; Grundmann, 2021:213; Miyazono 2023:2). I then focus on how epistemic vices are manifested in online environments, particularly those where information disorder is present (Wardle 2019). Finally, I conclude with an examination of institutional vices, which I argue can act as indicators of institutional trustworthiness.

Published Material

At the time of the submission of this thesis (11th January 2024) portions of Chapters 3, 4 and 5 have been accepted for publication or published in the following journals:

Meehan, D. 2019. Is epistemic blame distinct from moral blame? *Logos and Episteme* 10(2), pp. 183-194.

Meehan, D. 2020. Epistemic Vice and Epistemic Nudging: A Solution? In: G. Axtell & A. Bernal (eds.). *Epistemic Paternalism: Conceptions, Justifications and Implications (Collective Studies in Knowledge and Society).* Maryland: Rowman & Littlefield. pp. 249-261.

Meehan, D. 2023. A social account of the vices of self-assessment. Philosophical Psychology, 36(5), pp. 1033-1036.

Table of Contents

Acknowledgements

I would like to start by thanking my supervisors, Adam Carter, Mona Simion and Emma Gordon whom I owe a great deal of gratitude for their guidance and support throughout my time at Glasgow. Their invaluable contributions have played a pivotal role in the completion of this thesis, and I am deeply thankful for their mentorship and encouragement during these past years and the opportunities they provided me to grow as a scholar.

I would also like to express my gratitude to the various members of the 'vice-squad' for their insightful comments during numerous talks and discussions on the material of this thesis and a seemingly never-ending supply of new material to devour. These interactions have been instrumental in shaping my research and perspectives. A special thanks goes to Ian Kidd who served as a catalyst for my work in vice epistemology during an enlightening talk at the University of Edinburgh.

I am also deeply grateful for the feedback and advice when presenting various elements of this thesis at multiple conferences and workshops. I would also like to thank the Society for Applied Philosophy for their partial funding of my research. Without their scholarship, this research would not have been possible.

My colleagues, friends and family have been a constant source of both professional and emotional support and encouragement during this endeavour. I extend my greatest thanks to Dylan. Words cannot adequately capture how much you have helped me throughout every stage of this process and beyond. You have been an unwavering pillar of support, my most ardent champion and a source of unending motivation. Your belief in my abilities has never wavered, even when I doubted myself, and you have seen me through the highs and lows of this process. I am eternally grateful for you.

Finally, thank you to Ernie and Barney, the best writing companions I could ever ask for. They have kept me company throughout the many stages of writing this thesis and have quite literally been by my side for the entire process.

Author's Declaration

I confirm that this thesis is my own work and that I have: (i) read and understood the University of Glasgow Statement on Plagiarism, (ii) clearly referenced, in both text and the bibliography or references, all sources used in the work; (iii) fully referenced (including page numbers) and used inverted commas for all text quoted from books, journals, web, etc.; (iv) provided the sources for all tables, figures, data, etc. that are not my own work; (v) not made use of the works of any other student(s) past or present without acknowledgement. This includes any of my own works, that has been previously, or concurrently, submitted for assessment, either at this or any other educational institution; (vi) not sought or used the services of any professional agencies to produce this work; (vii) in addition, I understand that any false claim in respect of this work will result in disciplinary action in accordance with University regulations.

I declare I am aware of and understand the University's policy on plagiarism and I certify that this thesis is my own work, except where indicated by referencing, and that I followed the good academic practices noted above.

CHAPTER 1. INTRODUCTION

1.1 Non-ideal Epistemology

'Non-ideal' epistemology is a form of epistemology that focuses on bad epistemic practises and the various ways in which our epistemic lives can go 'wrong' (Barker et al. 2018; Bernecker et al. 2021; Mckenna 2023).[1]

Two recent advancements in mainstream analytic epistemology have paved the way for the conceptualisation of non-ideal epistemology (Barker et al. 2018). The first is an increasing recognition of the social dimension to our epistemic endeavours (Fricker 2007, 2012; Fuller 1988; Goldberg 2010, 2016, 2021; Goldman 1987, 1999; Medina 2013). Until recently, epistemology focused primarily on the epistemic analysis of epistemic goods such as true beliefs and knowledge, neglecting the role of epistemic agents themselves (Alston 2005; Palermos and Pritchard 2013). However, the practices of inquiry are untaken by epistemic agents and can be influenced by our social environments and communities. This recognition of the interplay between epistemic character and our surrounding environments has emerged as a central theme of epistemological discourse in recent decades, notably through the thriving domains of social and virtue epistemology (Code 1987; Goldman and O'Connor 2023; Montmarquet 1987; Zagzebski 1996).

A second development in epistemology involves the examination of how individuals and communities can deviate from the ideal in their pursuit of knowledge and other epistemic goods. This directs attention away from the idealised view of our epistemic lives and focuses instead on the realities that emerge when things go awry. This focus encompasses various aspects of applied and social epistemology, addressing issues such as bad epistemic characters and their resulting vices, the erosion of trust, epistemic injustice, ignorance, and the prevalence of information disorder (Bernecker et al. 2021; Hawley 2017, 2019; Lackey 2014, 2021, 2023; Mills 2007; Tuana 2006).

When our epistemic practices falter in these ways, there exists a potential for harm, extending beyond the realm of epistemology to encompass social, moral, and political spheres. This

[1] This distinction is based on ideal and non-ideal theory in ethics and political philosophy (Mills 2005, 2007; Rawls 1972).

prompts inquiries into how to respond to and address these harms and wrongs, thereby necessitating considerations of responsibility and the exploration of ameliorative solutions aimed at rectifying these flawed practices (Fricker 2007, 2016; Holroyd et al. 2017, 2020; Kidd 2020; Medina 2012, 2020; Sherman and Goguen 2019; Tanesini 2021).

1.2 Vice Epistemology

One area in which this recent interest in the non-ideal manifests itself is in the field of virtue and vice epistemology. Virtue epistemology is distinguished by its focus on the evaluation of epistemic agents, specifically the exploration of what qualities make someone an excellent epistemic agent (Baehr 2011; Code 1987; Montmarquet 1987; Roberts and Wood 2007; Sosa 2007, 2015; Zagzebski 1996).

Only in recent years has attention been directed towards epistemic *vices*, traits such as arrogance, dogmatism and closed-mindedness (Battaly 2016a, 2018a; Cassam 2016, 2019a; Tanesini 2018, 2020).[2] Epistemic vices are dispositions, attitudes and ways of thinking that make us bad thinkers, in so far as they prevent us from acquiring and sharing knowledge, manifest bad motives and desires, or disrupt both individual and collective epistemic functioning (Kidd et al. 2020).

The aims of contemporary vice epistemology have been helpfully divided into three key themes (Kidd et al. 2020:6):

1. Foundational work on the structure and features of epistemic vices and their impact on knowledge.
2. Analyses of specific epistemic vices.
3. Case studies in applied vice epistemology.

In accordance with the first theme, vice epistemologists have focused on the structure of epistemic vices and what makes them epistemically bad (Battaly 2014, 2016, 2018a; Cassam 2016, 2019a; Tanesini 2018, 2021). This includes a focus on the ontology and metaphysics of

[2] Epistemic vices were studied in various forms before this point, however, it was not named 'vice epistemology' until Cassam's (2016) paper of the same name.

epistemic vices, exploring what constitutes an epistemic vice and how they can be individuated from one another (Baehr 2020; Cassam 2020).

Epistemic virtues and vices can be classified in accordance with two perspectives, responsibilism and reliabilism. On the reliabilist view, epistemic virtues are stable dispositions that reliably produce true beliefs e.g., hard-wired faculties like reliable vision. Resultingly, epistemic virtues need not be acquired, need not be praiseworthy, and need not be personal qualities, though they must be reliable (Greco 2010; Sosa 2007). In contrast, responsibilism holds that epistemic virtues must be character traits, over whose acquisition or operation we exert some control over and for whose possession we are (partly) responsible. These epistemic virtues include open-mindedness and intellectual humility (Battaly 2016a:99).[3]

This same distinction also applies to epistemic vices. From a reliabilist analysis, an epistemic vice is a stable, unreliable disposition. From a responsibilist analysis, epistemic vices are blameworthy character traits which may be unreliable.

A related distinction can be made between vice monism and vice pluralism/heterogeneous accounts of epistemic vice. Vice-monism holds that epistemic vices are just one kind of thing e.g., character traits. Alternatively, vice pluralism contends that many different qualities can be considered epistemic vices e.g., thinking styles, sensibilities, and attitudes (Cassam 2020:37).

However, these distinctions are not as straightforward as they might seem. Some virtue and vice theorists argue against the need to choose exclusively between these conceptions of virtue and vice. Instead, they advocate for a form of personalism, wherein both unreliable/reliable faculties and good/bad character traits can be considered epistemic virtues and vices (Baehr 2011, 2020; Battaly 2016a).

Additionally, some vices might relate to one's epistemic character in a broad sense and not just be confined to traits e.g., epistemic vices can be 'dispositions to act, think, and feel in particular (rational or excellent) ways' (Baehr 2020:21). They may also be defined as 'a deep quality of a person' that can encompass any trait, behaviour, attitude or so forth that deeply reflects their self-hood (Zagzebski 1996:119). Similarly, some reliabilist accounts of knowledge include

[3] Virtue responsibilists disagree as to whether virtues also require reliability (Battaly 2018a).

notions of epistemic agency that carry implications for attributions of responsibility (Greco, 2010; Sosa 2011).

Furthermore, we can be responsible for vices beyond character traits. For example, we may be responsible for vices in so far as they result in bad epistemic consequences e.g., a lack of knowledge (Cassam 2016, 2019a). Likewise, some have argued that there are character traits vices for which we are not blameworthy (Battaly 2016a, 2018a; Kidd 2020).

There are also competing analyses on the 'badness' of epistemic vices. Broadly construed, a motivational perspective argues that motivations are integral to epistemic vices. These can include motivations towards an epistemic bad (e.g., indifference to truth) or away from an epistemic good (cognitive contact with reality) (Battaly 2016a, 2018a; Tanesini 2018, 2020).

Conversely, consequentialist accounts of epistemic vice maintain that vices are bad because of their bad effects or failure to generate good epistemic effects. One prominent account under this view is 'obstructivism' which holds that vices are bad because they systematically obstruct the acquisition, transmission, and retention of knowledge (Cassam 2019a:1).

There are also objections to both of these analyses, whilst others have argued that both explanations can be accepted (Crerar 2018, 2020; Fricker 2020).

Likewise, some vice epistemologists have argued that epistemic vices can be vicious in some contexts but virtuous in others e.g., closedmindedness and open-mindedness (Battaly 2018b, 2021). This speaks to a debate between vice externalism and internalism (Battaly 2022; Simion, 2023).

In addition to these key debates, vice epistemologists have also detailed various other factors of epistemic vices. These range from their 'stealthy' nature which means they cannot be known to the vice-bearer and their systematic occurrence, which means they must be displayed consistently (Cassam 2015, 2016, 2019a).

As a response to the detrimental nature of epistemic vices, questions of responsibility and amelioration also arise. Three key debates emerge from these considerations. The first is whether or not blame is a fair and appropriate response to vice. Consequently, this has

implications for the responsibilist claim that blame necessarily follows as a response to epistemic vice (Battaly 2019).

On the stronger end of the view, some vice epistemologists have argued that responsibility (not necessarily blame) is integral to the definition of a vice (Cassam 2016, 2019a; Code 1984; Fricker 2020; Zagzebski 1996). On the weaker position, others have argued that blame is not integral to the definition of a vice. The reasons for this argument vary, with some arguing that blame is appropriate under certain conditions, but just not in all instances (Battaly 2016a, 2018a; Kidd 2016, 2020; Tanesini 2018, 2021).

Second, there is a debate as to what type of responsibility is best suited to epistemic vices. Whether or not we are in control over our epistemic vices leads to a distinction between attributability responsibility and accountability responsibility.

Attributability responsibility holds that we are responsible for qualities that can be properly attributable to us, regardless of whether we are in control of it (Hieronymi 2008; Sher 2006 Shoemaker 2015; Talbert 2012; Watson 2004). An attribute can be said to 'belong' to us in the appropriate way if it reflects our motivations, values, or attitudes. In so far as our actions can reflect some aspect of ourselves, it can provide grounds for negative or positive appraisals. This appraisal attaches to the person *qua* agent and can take the form of praise or blame (Hieronymi 2008:358; Talbert 2008, 2019; Zheng 2016:65). As this form of responsibility does not require control to be exercised, it is a non-voluntary form of responsibility. This means vice-bearers can be blamed for their vices in this sense, despite potentially lacking control over their acquisition or retention of it.

Accountability responsibility, on the other hand, does require control, making it a voluntarist position. According to this view, individuals are held responsible for an action only when they have a realistic chance, either directly or indirectly, to refrain from breaching the standards for which they are being held accountable (Watson 2004:276).

Thirdly, there is a debate as to whether this form of responsibility amounts to blame or criticism, and what this can consist of e.g., distancing yourself from a vice-bearer, punishment, or anger (Cassam 2019a; Fricker 2020; Tanesini 2021). This also opens up a discussion on the distinct epistemic nature of blame (Boult 2020, 2021; Brown 2017, 2020).

Despite the disagreements across these three debates, there is a general consensus that holding vice-bearers responsible for their epistemic vices should aim to have an ameliorative effect (Cassam 2020; Kidd 2016). Theorists have also offered different strategies for ameliorating epistemic vices, ranging from policy changes to changes in our education curricula (Cassam 2019a; Holroyd 2016; Sherman and Goguen 2019, Tanesini 2021).

Focusing on the analyses of specific epistemic vices, vice epistemologists have identified a variety of specific epistemic vices. This ranges from the study of epistemic injustice and intellectual humility (Fricker 2007; Medina 2013), arrogance, timidity, and servility (Tanesini 2018, 2020a, 2020b), closedmindedness (Battaly 2018a, 2018b, 2021) and epistemic malevolence (Baehr 2010; Meyer 2023).[4]

Finally, vice-epistemologists have also focused their attention to the applied nature of epistemic vices and their practical application. Epistemic vices are present in the vast majority of epistemic individuals and structures, thereby causing real-world issues which need addressing.

Accordingly, several theorists have examined the presence of epistemic vices within political discourse, online discourse, healthcare practises and educational structures (Battaly 2013; Cassam 2019b; Kidd 2019, 2021, 2021, 2021 et al.; Nguyen 2018; Lynch 2019). This application also applies to the institutional domain, with some scholars arguing for the existence of institutional vices (Fricker 2010; Lahroodi 2018).

1.3 Thesis Overview

As previously outlined, epistemic vices are harmful, both epistemically and morally, to the vice-bearer and their surrounding community. At the individual level, they can cause harm by distorting and impairing the cognitive faculties of the vice-bearer, leading to flawed reasoning, biased judgement and a diminished capacity to engage in intellectual inquiry. They can also hinder the attainment of epistemic goods such as genuine understanding and the pursuit of knowledge (Cassam 2019a; Medina 2012, 2020; Priest 2020).

[4] See also epistemic insouciance (Cassam 2018), intellectual laziness (Kidd 2017) and epistemic self-indulgence (Battaly 2010).

As a result, when the epistemic capabilities of the vice-bearer are compromised, they can inflict harm on those around them. This harm may take the form of perpetuating misinformation or even instilling epistemic vices in others (Tanesini 2016).

Epistemic vices can also contribute to the erosion of trust within social networks and communities (Baird and Calvard 2018, Fricker 2020; Medina 2020; Sullivan and Alfano 2020) Likewise, institutions and networks themselves can be epistemically vicious, resulting in harms to those exposed or within the institution/network (Fricker 2020; Lahroodi 2018). Some systems and environments can be epistemically corruptive, meaning they damage an individual's epistemic character and facilitate the development and exercise of epistemic vices (Kidd 2019, 2020, 2021, 2021 et al., 2022).

All of these harms have a distinctively epistemic dimension in so far as they impede effective inquiry but can also have a moral dimension if they also result in moral failings. For example, vicious testimonial injustice can lead to moral wrongs and political injustices, stemming from unfair deflations of the testimonial credibility of agents (Congdon 2017; Pohlhaus Jr. 2014).

A recurring theme through my thesis is what we can do to address these harms. This can be dissected into three primary objectives. Firstly, I aim to evaluate the 'viciousness' of epistemic vices, evaluating the three main perspectives of obstructivism, motivationalism and personalism (Battaly 2016a, 2018a; Cassam 2016, 2019a; Tanesini 2018, 2021). Secondly, I aim to evaluate whether individuals are epistemically responsible for their vices and whether this responsibility can be distinctively epistemic in nature and ameliorative in spirit. Thirdly, I aim to assess how institutions and environments can either encourage the development of epistemic vices or exhibit epistemic vices themselves as a collective group.

1.4 Chapter Summaries

With these broad aims in mind, let us now turn to an overview of each chapter and its key objectives.

I start, in Chapter 2, with an examination of Quassim Cassam's (2016, 2019a) account of epistemic vice referred to as 'obstructivism'. I divide my evaluation of Cassam's account into two components.

First, I assess his 'obstructivist' claim, examining how character traits, attitudes and thinking styles can systematically obstruct knowledge and other epistemic goods. An objection is raised towards the systematic clause of this account which requires vices to obstruct knowledge 'more often than not' (Cassam 2019a:38). I question whether this is compatible with Cassam's distinction between high and low-fidelity vices, which pertains to the frequency of a vice's occurrence (Alfano 2013:32). I argue for a shift in focus, suggesting that the systematic clause should concentrate on the frequency of harms rather than the frequency of the vice itself occurring.

Shifting focus, I then assess the normative component of obstructivism, namely that epistemic vices are blameworthy or criticisable. I direct my attention to Cassam's distinction between criticism and blame, examining his rationale for considering criticism to be the most fundamental responsibility component for epistemic vice (2019a:127-128). This leads to an evaluation of Cassam's position on control and responsibility, as well as his argument that vice-bearers are not responsible for the acquisition of their epistemic vices (ibid.:128). Finally, I critique Cassam's preference for 'revision responsibility', a form of responsibility that requires vice-bearers to have control over their vice in its current state (ibid). I argue that it is unclear how this form of blame aligns with Cassam's aim for responsibility to serve an ameliorative purpose.

In Chapter 3, I turn my attention to motivationalism, critically examining two prominent motivationalism accounts of epistemic vice offered by Heather Battaly (2016, 2018a) and Alessandra Tanesini (2018, 2021). The focus begins with Battaly's account of personalism. I explore both the reliabilist and responsibilist features of this account which define vices as personal qualities for which we are not necessarily blameworthy for (2016:99-100, 2018a:115). I focus my attention primarily on Battaly's initial argument that vice-bearers should not be blamed for vices that they lacked control over (2018a:120-121).

Following this, I assess Battaly's later preference for attributability responsibility, a form of responsibility without control. I evaluate what the blame component of vice is responding to,

and how this view now contradicts her earlier rejection of the responsibilist claim that blame is not a necessary requirement of epistemic vice.

Shifting the focus, I turn to evaluate Tanesini's account of epistemic vice, referred to simply as 'motivationalism'. Tanesini claims that epistemic vices consist of self-deceptive epistemic motives towards epistemic bads and away from epistemic goods (2018:350). Her account is more optimistic about the prospects of blaming epistemic vices, as vice-bearers can be held attributability responsible for their vices (2021:171). However, Tanesini rejects the view that vices are inherently blameworthy for their vices due to a variety of practical and moral concerns with blaming vice-bearers (ibid.:182-183). I respond to each of these concerns, arguing that an ameliorative account of blame escapes these worries. Finally, I demonstrate that the form of responsibility that Tanesini is most optimistic towards, referred to as 'taking responsibility', contradicts her position that epistemic vices are undetectable to the vice-bearer (ibid.:186).

In Chapter 4, I turn to focus exclusively on the role of blame for epistemic vice. I present three interconnected objectives in this chapter. First, I aim to motivate an epistemic and ameliorative account of blame. I outline how epistemic blame can be defined as a response to a violation of an epistemic norm, that aims to reduce bad epistemic conduct and bring about epistemic goods (Boult 2021; Fricker 2016; Piovarchy 2021; Sliwa *forthcoming*). Drawing from feminist perspectives on responsibility and functional and communicative accounts of blame (Ciurria 2021), I explore the various ways in which blame can be directed towards ameliorative aims.

My second objective is to respond to the 'argument from lack of control' (Cassam 2019a; Battaly 2016a, 2018a; Kidd 2016, 2020). This is a reoccurring argument presented by many vice epistemologists in their rejection of blame for epistemic vice. In short, this argument claims that we should not be blamed for epistemic vices that are outside of one's control i.e., vices formed as the result of environmental factors which we could not change.

In response to this position, I advocate for attributability responsibility, demonstrating how this is a form of responsibility that does not require control, and coupled with my earlier arguments, can be a fair response to epistemic vice. At this stage, I also respond to an objection presented by Battaly (2019) towards this form of responsibility as a non-voluntary form of blame.

Turning to my third and final aim, I defend the responsibilist position that responsibility is a necessary feature of epistemic vices. I argue this follows from the acceptance of attributability responsibility as the appropriate form of blame for epistemic vices. This is because the qualities which are worthy of blame under attributability responsibility align precisely with epistemic vices.

Moving beyond the foundational theme of the first half of the thesis, I turn my attention in Chapter 5 to a potential ameliorative solution to the presence of epistemic vices. The solution in question concerns epistemic nudging, a paternalistic method of nudging individuals towards epistemically desirable outcomes (Adams and Niker 2021; Grundmann 2021:213; Miyazono 2023:2).

Despite initial signs of plausibility as an ameliorative solution, I argue that epistemic nudging only provides a superficial mitigation of epistemic vices. This is because its effectiveness lies primarily in 'masking' vices rather than addressing their fundamental causes and 'deep' nature.

I then argue that more concerningly, epistemic nudging might contribute to the creation of further epistemic vices, specifically those of epistemic injustice and intellectual laziness (Kidd 2017; Riley 2017). I also review the concern that epistemic nudging may violate our intellectual autonomy (Riley 2017). I challenge the assertion that the benefits of attaining epistemic goods outweigh the restriction on one's autonomy, given my prior argument that epistemic nudging can also lead to the creation of epistemic vices. Finally, I turn my attention to 'weak epistemic paternalism' (Miyazono 2023). I argue that this form of epistemic nudging may address concerns to do with autonomy and is unlikely to result in further displays of vice. However, this is too weak to make any significant changes to our epistemic behaviours.

In Chapter 6, I turn my attention to epistemically corruptive environments that promote epistemic vices. I argue that environments where information disorder is present, encompassing issues such as fake news and related phenomena, can be epistemically harmful to one's intellectual character by promoting intellectual vices (Wardle 2019).

Following Ian Kidd's (2019, 2020) framework for a successful 'corruption criticism', which is to label something as epistemic corrupting, I assess how online media environments can be epistemically corrupting. Within the context of information disorder, I identify three distinct

epistemic vices that it tends to foster: prejudice, conspiratorial thinking, and epistemic capitulation (Battaly 2017b; Begby 2013; Cassam 2019a, 2019b; Fricker 2007). I then outline how these vices are brought about by information disorder in accordance with five modes of corruption, focusing on how vices can be intensified, propagated, and created.

I conclude by examining the different ameliorative solutions to these identified vices and their corruptive environments. I outline three broad approaches that one could take here. Firstly, there are individualistic solutions, directed at the vice bearer and their vice. Secondly, there are structural solutions, directed at the systems and structures that facilitate and create vices Thirdly, there are social solutions, directed at epistemic communities and networks. I argue that virtue-centric ameliorative solutions are not restricted to individual approaches, and instead, a coordinated approach that aims to develop individual and social virtues can go some way in addressing the outlined vices.

Finally, in Chapter 7, I turn my attention to institutional epistemic vice and how they can affect our trust in institutions. I argue that institutions can exhibit epistemic vices as a collective by appealing to Margaret Gilbert's plural subject theory (1987, 2000, 2002, 2013) and Miranda Fricker's (2010, 2020) argument for an 'institutional ethos' (2020:90).

I propose two modifications to Fricker's account of institutional epistemic virtue and vice. Firstly, I add a consequentialist modification to explain how institutional vices are vicious due to their bad motivations or bad epistemic effects. Secondly, I drop the self-awareness requirement of Fricker's account which claims that the institution's members must be aware of their commitments or motives (2010:245). I argue that this conflicts with our understanding of epistemic vices as primary 'stealthy' and undetectable vicious motives (Cassam 2015, 2016, 2019a; Tanesini 2021).

Having outlined how institutions can possess epistemic virtues and vices, I then focus on how these virtues and vices can be used in our evaluation of trustworthy institutions. I focus on how an institution can be trustworthy/untrustworthy as an institution, by displaying the communicative attributes of transparency and honesty (Byerly 2022a; King 2021; Wilson 2018). I then explain how these attributes (and their counterpart vices) are virtues and vices of institutions pertaining to trustworthiness. Finally, I conclude this chapter by raising and responding to the potential objection that transparency and honesty are not always beneficial

attributes of an institution, nor indicators of a trustworthy institution (John 2018; Nguyen 2021).

CHAPTER 2. VICE CONSEQUENTIALISM: A CRITICAL ANALYSIS OF OBSTRUCTIVISM

2.1 Introduction

This chapter focuses on one of the prominent consequentialist accounts of epistemic vice, known as 'obstructivism', proposed by Quassim Cassam (2016, 2019a). Obstructivism defines epistemic vice as '…blameworthy or otherwise reprehensible character traits, attitudes, or ways of thinking that systematically obstruct the gaining, keeping, or sharing of knowledge.' (Cassam 2019a:23). I divide my analysis into two parts, focusing on the two halves of Cassam's obstructivist account. The first section will assess the claim that epistemic vices are character traits, attitudes or thinking styles that must systematically obstruct knowledge. I refer to this as the 'obstructivist claim'. I raise an objection to this part of the account, namely that Cassam's systematic requirement is incompatible with his further claim that epistemic vices can be low-fidelity vices, meaning a vice requires a low threshold of behavioural consistency. I also offer a way to resolve this concern by modifying the systematic requirement on vice.

The second section of this chapter assesses the normative claim that epistemic vices must be blameworthy or at least reprehensible. I focus on Cassam's distinction between criticism and blame and his preference for a form of revision responsibility for epistemic vice, where the attribution of blame depends on whether the vice bearer exercised control over it (2019a:124). I also focus on Cassam's reasons for rejecting acquisition responsibility which states that a person is responsible for a vice if they are responsible for acquiring or developing it (ibid.) I challenge Cassam's endorsement of revision responsibility for epistemic vices and argue that his reasons for rejecting acquisitional responsibility are uncompelling. By raising these objections, I aim to further our understanding of the nature of epistemic vices, specifically whether we can be held responsible for them, and what form this responsibility may take.

2.2 The Obstructivist Claim

Within the framework of obstructivism, epistemic vices are identified as character traits, attitudes, or ways of thinking that obstruct knowledge. In this sense, obstructivism offers a

consequentialist perspective of epistemic vice in so far as it characterizes epistemic vices in terms of their negative consequences for effective epistemic inquiry (Cassam 2019a:5). Following Christopher Hookway (1994), Cassam highlights inquiry as the primary aim of epistemology. It is the endeavour 'to find things out, to extend our knowledge by carrying out investigations directed at answering questions, and to refine our knowledge by considering questions about things we currently hold true' (Hookway 1994:211). In general, an inquiry is deemed successful when it is both responsible and effective, while an unsuccessful one is characterized by irresponsibility and/or ineffectiveness (Cassam 2019a:7). Building on from this, Cassam argues that makes a trait a virtue or a vice is the consequences it has for the kinds of inquiries we conduct, and whether they successfully result in knowledge or not.

By the definition of obstructivism, any character trait, thinking style or attitude must occur *systematically* in order to suffice as an epistemic vice. That is to say, a character trait, thinking style or attitude must systematically 'impede effective inquiry' in order to count as an epistemic vice (ibid.). By building in this requirement, Cassam follows Julia Driver (2001:82) in allowing for a distinction to be made between instances of luck or accidents and virtues and vices. For example, consider a gullible individual who believes a statement that just so happens to turn out true. At first glance, this individual's gullibility has seemingly led them to an epistemic good. However, under obstructivism, their gullibility could not be defined as an epistemic virtue as it does not s*ystematically* lead to positive epistemic effects (in most cases, it likely results in epistemic harms). Whilst it just so happens that the individual's gullibility epistemically paid off in this instance, we can easily imagine many other instances where it did not, meaning it was a matter of luck that the belief transpired to be true. To avoid cases such as these, the requirement that a behaviour must occur systematically, meaning consistently, is built into the definition of vice. In sum, under obstructivism, we cannot classify traits as epistemic vices unless they *systematically* result in bad epistemic effects, where bad epistemic effects are defined by how they impede inquiry.

Cassam also offers a heterogeneous account of epistemic vices, meaning there can be different varieties of vice such as emotions or cognitive biases (2020:40). Cassam classifies epistemic vices into three primary categories: character traits, attitudes, and ways of thinking. Let us turn to examine each of these categories more closely.

2.2.1 Character Traits

Cassam defines character traits as 'stable dispositions to act, think, and feel in particular ways' (2019a:31). Historically, the concept of character traits has predominantly focused on moral qualities as opposed to epistemic ones e.g., honesty and cruelty.[5] Cassam aims to draw from these moral traits in order to understand their epistemic dimension. He focuses on two dimensions of moral traits that can map onto epistemic character traits; the behavioural and psychological (Doris 2002).

The behavioural component of moral traits requires that they are exercised consistently. For example, an honest person will not just be honest in certain situations when it is easy or convenient to do so, but also when it is inconvenient and difficult to do so. Crucially, this consistency of behaviour must have its foundation in the individual's values, desires, and motives. The individual must be consistently honest because they have the proper desires and motives of an honest person. In turn, these desires and motives must be rooted in one's values, they desire to do the right thing because they value honesty (ibid.:31-32).

Returning to character traits, we can understand behavioural consistency by distinguishing between high and low-fidelity traits (Mark Alfano 2013:31-32). This is the level of behavioural consistency required by different varieties of vice. Cassam argues that most epistemic character vices are what Alfano (2013) calls 'high-fidelity' traits which require near-perfect consistency (2019:38-39). For example, a person who steals but is occasionally honest is not faithful. Contrastingly, low-fidelity vices do not require near-perfect consistency. For example, an individual who is occasionally cruel displays the behaviour enough times to be considered cruel (ibid.:33).

According to Cassam, closed-mindedness is considered to be the paradigm character trait that serves as a representation or exemplar of many other character traits. A closed-minded individual tends to be rigid in their thinking, resistant to considering new information and intolerant of opposing views (ibid.:34). These tendencies are behavioural dispositions, where 'behavioural' is now understood in a broad epistemic sense to include intellectual/epistemic

[5] For further literature on moral virtues and vices see Driver (2001); Hurka (2001); Merritt et al. (2010) and Zagzebski (1995, 1996, 2004).

conduct. What qualifies closed-mindedness as an epistemic trait is the fact that it is associated with various epistemic dispositions '…to think, reason, and respond to new information in certain ways' (ibid.:11). In this context, closed-mindedness primarily pertains to one's disposition; being closed-minded involves having specific intellectual tendencies and harbouring motives and values that align with these inclinations.

There is also a psychological element to many intellectual character traits. For example, with closedmindedness, there is usually a need for cognitive closure. Closure can be non-specific, such as the desire for a confident judgement, or specific, like a desire for a particular answer to a question. Not every vice has this psychological component, however. For example, there is no motive or desire that relates to stupidity in the way that the need for closure relates to closed-mindedness (ibid.:39).

Cassam considers closemindedness to be a reasonably accurate guide to the nature of character vices. This means, in general, character traits are stable dispositions to act, think, and feel in particular ways. They tend to be high-fidelity traits meaning that they require a high consistency across contexts and are not subject-specific. Finally, they often have a psychological component which is the need for closure of some sort.

2.2.2 Thinking Styles

Cassam's second category of epistemic vice concerns thinking vices. Thinking vices are epistemically vicious ways of thinking that are closely related, but still distinct from, character vices (ibid.:57).

In order to understand the difference between thinking vices and character vices, we must first understand Cassam's separate distinction between the qualities of a thinker and the qualities of a person's thinking e.g., what is the difference in *being* closed-minded versus *thinking* closed-mindedly?

To help explain this distinction, we can borrow Cassam's example of widely publicized example of a miscarriage of justice, referred to as the 'Birmingham Six' (ibid.:53-54). In November 1974, a bomb exploded in the Mulberry Bush pub in Birmingham leading to the

arrest of six men. During their trial, their defence argued that the forensic evidence against them was unreliable and that the confessions had been beaten out of the defendants. Despite this, the judges ruled that the men were guilty, for if their claims were true then there would have to be an admission of guilt on the police's side. In 1991, the Court of Appeal ruled that the original convictions were unsafe and unsatisfactory and that the forensic evidence at the first trial had been demonstrably wrong. Consequently, the Birmingham Six were released.

When assessing how the judges arrived at such a poor decision, one explanation is that they displayed an array of epistemic vices (ibid.:56). Specifically, they displayed a lack of humility and acted from closed-mindedness and gullibility. However, these same judges also had displayed previous instances of virtues. For example, one of the judges, Lord Denning, was responsible for some (then) progressive and open-minded rulings e.g., that a divorced wife is entitled to an equal share of her husband's wealth, and that non-married, cohabiting couples have rights to each other's property (Burrell 1999: Dyer 1999; Freeman 1993).

Cassam acknowledges that the open-mindedness and progressiveness of Lord Denning's prior rulings seem to suggest that he is not the closed-minded individual we may have believed him to be with regard to his decisions of the Birmingham Six. What this example illustrates is the need to thereby distinguish between the qualities of a *thinker* and the qualities of *thinking* on a particular occasion. Although the thinking that resulted in the Birmingham Six being wrongfully imprisoned may have been closed-minded, it cannot be inferred that the judges themselves were inherently closed-minded (2019a:57). As we've seen, closedminded is a high-fidelity vice, a persistent flaw, and only thinkers who consistently exhibit this quality can be considered closed-minded. The evidence of the judge's previous behaviour and decision-making suggests that they were not consistently engaged in closed-minded thinking. In turn, this casts doubt on the notion that their behaviour in the case of the Birmingham Six is attributable to their overall epistemic conduct.

Another way to explain the wrongness of the judge's behaviour is to therefore appeal to thinking vices. The closed-mindedness being displayed is a quality of this instance of thinking, a thinking style, and in this instance a thinking vice. It is '…thinking that is inflexible and unreceptive to evidence that conflicts with one's pre-existing views, regardless of its merits' (ibid.:58).

In differentiating closed-mindedness as a thinking vice from a character trait, Cassam observes that closed-mindedness as a character trait depends on first comprehending what it means to *think* closed-mindedly. In other words, to understand closed-mindedness, one needs to grasp how individuals with this trait think and process information. However, the reverse is not necessarily true; merely knowing that closed-minded thinkers engage in closed-minded thinking does not fully elucidate what closed-minded thinking actually entails.

Another way to express this is that character vices build off thinking vices due to their 'explanatory basicness'. This means that 'X is explanatorily more basic than Y just if X can be explained without reference to Y but Y can't be explained without reference to X' (ibid.:59). In this context, the closed-mindedness thinking vice is explanatorily more basic than the closed-mindedness character vice.

Drawing upon the distinction between thinking fast and thinking slow, Cassam argues that thinking vices can occur in both slow thinking (goal-directed, deliberate, conscious) and fast thinking (rapid, automatic and effortless) (ibid.:60).[6] Slow thinking can be epistemically vicious, as we can see through the example of the judge's conclusion in the case of the Birmingham Six trial – he refused to give appropriate weight to the possibility that the accused were telling the truth. Fast thinking can also be vicious. For example, we can suppose that the judge's bias in favour of the police was so strong that the fact the prison doctor was telling the truth about the injuries of the Birmingham Six, did not even cross his mind. His dismissal of the relevant possibilities happened instantaneously and automatically rather than as a result of conscious reflection. Cassam also notes that many cases of cognitive biases are examples of fast thinking and can also be considered epistemic vices (ibid.:66).

In sum, thinking vices are epistemically vicious ways of thinking or thinking styles. Thinking vices are distinguishable from character vices in the sense that they are more explanatory basic than character traits. This means that one can exhibit vicious thinking on an occasion without possessing the associated vicious character trait. Thinking vices can also occur in both slow thinking (goal-directed, sequential, effortful) and fast thinking (automatic, effortless).

[6] For more on the distinction between fast and slow thinking, see Baron (1985) and Cassam (2010).

2.2.3 Attitudes

The third species of vice discussed by Cassam are attitude vices. An 'attitude' can be broadly defined as a perspective or evaluation of an object (2019a:81).[7] These objects are referred to as 'attitude objects', defined as any object which it is possible for someone to have an attitude towards (ibid.). In this broad sense, attitude objects can encompass a variety of things, from people to buildings or political parties. Straightforwardly, a basic positive attitude is liking, and a basic negative attitude is disliking.

In explaining further what is meant by an 'attitude', Cassam distinguishes between two varieties: stances and postures. Two key features of postures are that they are *affective* and *involuntary* (ibid.:82-83). Considering the attitude of contempt, Cassam notes that contempt is not just a matter of belief or opinion, but something that is felt towards an object (ibid.). This feeling of contempt is an *affective* quality of an attitude. The affective quality of an attitude motivates behavioural manifestations of contempt such as refusing to shake someone's hand. Attitudes are involuntary in the sense that they are not generally matters of choice. This mirrors how we often consider feelings to behave e.g., you cannot choose to feel contempt towards someone.[8]

The paradigm example of an attitude vice that takes the form of an affective epistemic posture, is epistemic insouciance. This is defined as an indifference or lack of concern as to whether claims are grounded in reality or evidence (ibid.:79). Epistemic insouciance is also an *attitude* of indifference towards the truth, which is what separates it from a thinking style or character trait (ibid.). The attitude object of epistemic insouciance is the object of inquiry, as the epistemically insouciant individual 'views the business of acquiring knowledge via investigation as a meaningless and tedious chore which doesn't warrant one's full attention' (ibid.:86). Epistemic insouciance is also a posture attitude. This is because it is involuntary – one does not choose to have an indifferent attitude toward inquiry. The affective quality of epistemic insouciance is characterised by the lack of concern for epistemic goods such as evidence or truthfulness, and also by the presence of contempt for these epistemic goods too (ibid.:79).

[7] For more on the psychological definition of attitudes, see Maio and Haddock (2015).
[8] There is a difference between choosing to show respect or contempt and having respect or contempt (Cassam 2019a:83).

Unlike postures, 'stances' are voluntary attitudes which one can choose to adopt or reject (Fraassen 2004). An attitude vice of this sort is epistemic malevolence, defined as an opposition to knowledge (Baehr 2010). Cassam notes that malevolence in this form is an attitude vice but a stance rather than a posture. It is voluntary in so far as it involves a decision to act in a way that undermines knowledge, such as spreading doubt about the validity of scientific knowledge. It can also lack the affective quality that is essential for posture-based attitude vices.

Attitude vices are more explanatory basic than character trait vices since character trait vices must make reference to the relevant attitudes (ibid.:99). To recap, this means that an attitude can be explained without reference to the character trait, but the character trait cannot be explained without reference to the attitude. Regarding the relationship between attitude and thinking vices, neither the attitude nor the way of thinking is more basic than the other, (though both are more basic than the character traits to which they correspond) (ibid.:98).

To conclude, Cassam understands attitude vices as orientations or postures towards something that comes in two varieties: stances and postures. Postures are involuntary and affective whereas stances are voluntary and lack an affective element. Attitude vices are distinct from character vices in the sense that attitudes are more basic than character traits. Finally, they are distinct from thinking vices in the sense that attitudes involve thinking and ways of thinking cannot be properly explained without reference to the attitudes they manifest, leaving neither more basic than the other.

2.3 Systematic versus Low Fidelity Vices

Having laid out the three classifications of things Cassam identifies as vices, I will now raise an objection concerning Cassam's claim that vices must occur *systematically* in order to be defined as vices. The concern arises from the potential inconsistency with Cassam's assertion that certain vices can be low-fidelity. I also propose a solution to address this concern by revising Cassam's interpretation of 'systematically'.

As discussed above, Cassam holds that a character trait, thinking style or attitude must systematically obstruct knowledge in order to be defined as an epistemic vice (alongside the normative component of the definition). Cassam defines 'systematic' as akin to consistency,

meaning that vicious behaviour must 'normally' (ibid.:4) and more than 'sometimes' (ibid.:35) get in the way of knowledge. Cassam borrows this criterion from Driver's (2001) account of moral virtues and vice, which discusses the systematic nature of vices and virtues in more depth. Driver seems to understand systematicness as a majority '…any account of virtue must be able to tolerate some actual mistakes, and not mere haziness, as long as those mistakes systematically promote the good more than not.' (Driver 2001:70). Applied to epistemic vices, we can better understand systematic as meaning that vices occur when they obstruct one's attempt to form knowledge 'more often than not' (ibid.:38).[9]

Cassam introduces the systematic clause in order to distinguish epistemic vices from one-off displays of bad epistemic behaviour, instances of luck, or seemingly vicious behaviours that result in good epistemic ends (2019a.:12). For example, if an agent only acted arrogantly a handful of times, they would not be said to hold the vice of arrogance as their behaviour was consistent, despite it resulting it epistemic harms. Likewise, if an agent displays closedmindedness consistently, but their closedmindedness did not result in any bad consequences, they would not be said to hold the vice of closedmindedness. It is clear then, that for Cassam, the combination of both systematic and bad effects must be held for the behaviour to be categorised as a vice.

However, whilst the addition of the systematic clause might allow Cassam to explain why instances of less than-consistent behaviours do not count as vices, there is some confusion over its compatibility with another important feature of Cassam's account, namely the fidelity status of vices.

As we have seen, fidelity is also a measurement for determining the behavioural consistency of vices. Returning to Alfano's distinction between low and high-fidelity virtues, we can refer to his 'saturation metaphor' to explain this requirement (Alfano 2013:32). Using a metaphor of a blue piece of paper, Alfano argues that properties such as colour are gradable. For example, a piece of paper could be powder blue in some places, ultramarine in others and even have splotches of white here and there. The paper's saturation of blue has a depth dimension – how deep is the blue? – and a breadth dimension – how deep is the blue in each region? Alfano compares this type of saturation to virtue properties. For example, a saint may be virtuous

[9] Cassam describes his account as the 'epistemic analogue' (2019a:12) of Driver's (2001) in his discussion of the systematic requirement.

through and through. This would correspond to the paper being entirety ultramarine. The average person may be mostly virtuous through and through. This would correspond to the paper being baby blue from edge to edge (2013:29). The breadth metaphor applies as well; someone may be deeply virtuous in some respects but vicious in others. This corresponds to the paper's having splotches of ultramarine with splotches of baby blue, powder blue, and white. Just as we can say that a piece of paper is blue even if it has a few splotches of white, we can say that a person is virtuous despite the fact that they are not virtuous through and through.

High-fidelity virtues are only attributable to someone who has the property in both its full depth and breadth dimensions. These include the virtues of chastity, fairness, fidelity, honesty, justice, and trustworthiness (ibid.:31). Using percentages to explain this, Alfano gives the example of someone who does not steal in 70% of cases when they could have. This does not mean they possess the virtue of honesty (ibid.).

Conversely, low-fidelity virtues may be attributable even to someone who has the property in some depth and breadth, but nowhere near full depth and breadth. These include charity, diligence, friendliness, generosity, industry, magnanimity, mercy, tact, and tenacity (ibid.:32). The percentage example here would be that someone who gives to charity 20% of the time can have the virtue of charitability.

Broken down in this respect, we can understand the distinction between low and high-fidelity virtue/vice as a majority percentage. 100% is moral saint or ultramarine. Sometimes traits need to be displayed as close to that as possible to be regarded as virtue, and sometimes they need only make a 20% threshold. Overall, however, what this distinction is concerned with is measuring instances of behaviours, to know whether they can be considered a virtue or vice.

However, immediately, a problem arises when we try to marry systematic and low-fidelity vices. How can a trait need to occur systematically in order to qualify as a vice (amongst other things) yet also be low fidelity in nature? There is clearly a contradiction in these two behavioural consistency requirements.

The simple solution would be to drop one of these requirements. Yet this would prove a troublesome outcome for Cassam. On one hand, too many vices are classified by Cassam as

low-fidelity vices, meaning getting rid of this quantifier would mean grouping vices under high-fidelity that clearly do not fit. Take dishonesty for example. It seems true that only a few occasions of dishonesty are enough to label one as dishonest, and low-fidelity allows for this to be true. Alternatively, the systematic requirement is integral in preventing one-off displays of bad traits or occurrences from being qualified as instances of epistemic vices. Getting rid of this would allow for one-off instances of bad behaviour to count as a vice.

The best approach then, would be to try and make sense of how low-fidelity and systematic conditions can both feature in Cassam's account of vice. I argue that this may be possible if we refine what exactly we are trying to measure. Specifically, whether we are measuring the number of times a character trait is displayed or the number of epistemic harms it results in.

2.3.1 Refining the Systematic Requirement

One suggestion is that by modifying the systematic requirement in the context of Cassam's consequentialist account of vice, we can understand it as being primarily concerned with measuring the *consequences* of behaviours, specifically the resulting epistemic harms. This is then distinct from fidelity, which is concerned with how many times the behaviour itself occurs, and what threshold needs to be applied. On this interpretation, the systematic requirement is concerned with the number of times the behaviour in question resulted in an epistemic harm in proportion to how many times it was exercised. The epistemic harms would be understood as 'obstructed the gaining, keeping or sharing of knowledge' for Cassam (2019a:23). If a behaviour resulted in epistemic harms 'more often than not', and if the other conditions for vice were met, it would be deemed epistemically vicious (ibid.:38).

To now combine the two requirements, we can consider low-fidelity as a measurement of how many instances the trait in question was displayed by an agent. Alternatively, the systematic condition is concerned with how many times the trait resulted in epistemic harms. If it did so consistently, then it can be classified as an epistemic vice. This then allows for both measurements to work together in Cassam's obstructivist account of vice.

This allows a trait to occur only 'occasionally' and be classified as a low-fidelity trait. Despite occurring occasionally, if it is found to systematically result in bad epistemic effects, this will

be enough to categorize it as a low-fidelity vice. Going back to our example vice of cruelty, we get the result that agent A who is cruel only 30% of the time (a low-fidelity trait) but has an epistemic harm 'success rate' of 60% (via the systematic clause), possesses the epistemic vice of cruelty.[10] An additional ramification and benefit of this adjustment to Cassam's vice measurement is that we get different results in what is classified as a vice, that are arguably better suited to a consequentialist account of vice.

In the original reading of systematic, an agent may exercise cruelty 60% per cent of the time, meaning they count as holding the vice of cruelty (again, assuming all other conditions for a vice attribution are met). In my modified reading, despite acting cruelly 60% of the time, the agent may have only caused epistemic harm in 30% of those occasions, meaning they have not met the requirement for the vice of cruelty.

This adjustment not only allows for compatibility between Cassam's low fidelity and systematic requirements but arguably gives us better results when it comes to primarily concerning ourselves with the epistemic harms of vice, in keeping with Cassam's consequentialist aims.

Consider the below example that highlights this.

Conspiracy Thinkers

Test 1

Alex takes a vice personality quiz to determine whether he holds the trait of conspiratorial thinking. The vice quiz uses Cassam's systematic requirement, understood only as the number of times potentially vicious behaviour occurs, to determine whether the subject has a vice or not.

Alex is given a weighting of 60% for the vice of conspiratorial thinking. The types of conspiracy theories that Alex listed he believed in include the belief that Avril Lavigne died in

[10] Likewise, we can also allow for instances where a trait is exercised for the majority of the time but does not result in enough epistemic harms to be ruled out as an instance of epistemic vice. We would also need to ensure the 30% is indexed to the times that it would be appropriate to manifest cruelty (i.e., not when one is sleeping).

2003 and was replaced with a body double called Melissa Vandella, the Disney film 'Frozen' was created to hide the fact that Walt Disney's body was cryogenically frozen, and that the Moon is just a light projection.

Alex's friend, Sammi, also takes the same test. Sammi is also a believer of many conspiracy theories, but overall, only believes more than a handful. She is assigned a score of 30% for the trait of conspiratorial thinking. However, Sammi believes in extremely dangerous and harmful theories. These include the belief that the Sandy Hook shooting was a hoax and that 5G pylons caused the coronavirus. The conspiracy theories that Sammi believes in are far more epistemically harmful than Alex's. She spreads misinformation on online forums trying to 'inform' people that Sandy Hook victims were paid actors, and she also is hesitant to go to school as it's located near a 5G pylon that she fears will give her COVID-19. Every instance of Sammi's conspiratorial thinking results in epistemic harm.

Test 2

Alex takes another vice personality quiz; this time it uses the term 'systematically' to focus on whether the harms caused by a trait in question occurred more often than not. Alex gets the result of 60%, for the vice of conspiratorial thinking because all of Alex's conspiracy theories are epistemic harmful; they prevent him from the knowledge that Avril Lavigne is very much alive, that Walt Disney is not cryogenically frozen and that the Moon is not a projection.

Sammi also takes the revised test. This time she gets the output of 100% for the vice of conspiratorial thinking. This is because despite only believing in a handful of conspiracy theories, all of them systematically led to epistemic harms.

If we are only concerned with the number of times the trait of conspiracy thinking occurs, then we get the result that Alex possesses the vice of conspiratorial thinking, but Sammi does not. However, if we are also concerned with the amount of epistemic harm that results from the trait, as I previously described, then we also get the result that Sammi holds the vice of conspiratorial thinking. This is because her 'success rate' for harm is extremely high, even if the occurrence of the behaviour itself is not.

Why would we want to consider Sammi's behaviour as vicious? Using Cassam's reasoning, there is something epistemically wrong with Sammi's behaviour that appears to be best defined by her possession of the conspiratorial thinking trait. Secondly, there is value in being able to label Sammi's actions as vicious. As we will discuss in the next section and further chapters, using the language of vice allows for appropriate measures of criticism or blame to be made and various self-improvement ameliorative steps to be taken (Cassam 2019a; Kidd 2020, 2022; Tanesini 2021). As Miranda Fricker notes, the vocabulary of virtues and vices is a 'proper part of our contemporary normative equipment for ethical evaluation' (Fricker 2020:90). For the same reason that motivates Cassam's and other popular accounts of vice, labelling Sammi's behaviour as vicious would be a useful and important normative tool.[11]

To summarise, by modifying the systematic requirement as being primarily concerned with measuring the *consequences* of behaviours, specifically the resulting epistemic harms, we overcome the concern that it is incompatible with his other form of vice measurement, low/high fidelity, and we get a better result on what to classify as vices.

2.4 The Normative Claim

Having explored what types of things are epistemic vices and how we can distinguish between the three species of character traits, thinking styles and attitudes, we can now focus on the normative part of Cassam's account. Where is the *badness* of a vice located, and what makes a given trait, thinking style, or attitude an epistemic *vice* as such?

As we have seen, obstructivism places emphasis on the consequences of epistemic vices. What makes vices vicious on this view is that they tend to systematically result in bad epistemic effects, or a failure to result in good epistemic effects (whether that be towards the individual, other agents, collectives, or the environment). For obstructivism specifically, the badness is located in how vices obstruct the gaining, keeping and sharing of knowledge (2019a:5).

In what sense then, do character traits, thinking styles and attitudes obstruct knowledge in a way that constitutes an epistemic vice?

[11] I speak more on why blame can be an important ameliorative tool for vice in Chapter 4 of this thesis.

Starting with the paradigm character vice of closedmindedness, straightforwardly we can see how being reluctant to consider novel information can impede effective inquiry. Inquiry involves extending our knowledge by carrying out investigations, assessing new evidence and answering questions. Someone who is reluctant to consider new information or evidence will inevitably be prevented from acquiring the resulting epistemic goods of inquiry. Cassam notes that closedmindedness implies an 'unwillingness to go wherever the evidence leads' (ibid.:35) making it a clear example of an obstacle to knowledge, whether this is via acquisition, transmission, or retention.[12]

Turning to thinking vices, vicious thinking can hinder our ability to acquire knowledge and to evaluate evidence objectively. In general, vicious thinkers tend to be resistant to new information, rely too heavily on their preconceptions and ignore or dismiss counterevidence. This can lead to beliefs that are unsupported by evidence or that are even contrary to the available evidence (ibid.:67).

More specifically, Cassam claims that vicious thinking can lead to systematic errors which get in the way of knowledge. Take the paradigm thinking vice of wishful thinking. This can lead to error as the self-fulfilment of a person's wishes is usually given more weight than evidence (ibid.:59). Consider someone who enters the lottery and frequently daydreams about how they will spend their winnings and all the luxuries they will be able to afford. They have not won the lottery before, despite entering multiple times, but they believe today their 'luck is in' and their numbers will be drawn. This wishful thinking ignores the statistical odds and the fact that winning the lottery is largely a matter of chance and not the result of willing it to happen. Here wishful thinking has prevented the player from knowledge about the real statistical chance of winning the lottery.[13]

Cassam also details how thinking vices can prevent us from acquiring knowledge by lowering our confidence in our beliefs (ibid.:67). For one to know the truth of some proposition P one must be reasonably confident that P and one must have the right to be confident. However, if

[12] There is a debate on the epistemological benefits of closed-mindedness, and consequently, whether it should be categorised as an epistemic virtue as opposed to a vice (Battaly 2018b). To speak to this debate, Cassam argues that closedmindedness *systematically* obstructs knowledge, making it an epistemic vice (2019a:38).

[13] The fact that the belief was based on wishful thinking is what resulted in the loss of knowledge, not just that wishful thinking was a casual antecedent for the belief. In some instances, wishful thinking may causally contribute to knowledge e.g., if it causes you to enter a contest (believing that you will win) in which you gain some knowledge. What matters in this example is that wishful thinking obstructs knowledge when the belief is based on the wishful thinking.

it is wishful thinking rather than the evidence that leads a person to believe P, then they do not possess the right to be confident that P even if, by some chance, P turns out to be true after all. This is because individuals may have the right to be confident only if their belief is reasonable and formed using a reliable method (Bonjour 2001; Foster 1985).

Turning our attention to attitude vices, one way that attitude vices can get in the way of knowledge is by making one's beliefs less likely to be evidence-based (2019a:94). For example, someone who is epistemically insouciant does not care for evidence, meaning they are unlikely to form informed, true beliefs. Likewise, an individual who is epistemically insouciant and does not care about the evidence will make little effort to ground his beliefs in evidence.

Additionally, attitude vices can obstruct knowledge by undermining one's confidence in their belief, in the same way that thinking vices lower confidence in beliefs. Consider the example of a strategy employed by the tobacco industry in the face of research which discovered the concerning correlation between smoking and cancer. In response to this research, they founded the 'Tobacco Industry Research Committee' to cast doubt over the link between smoking and cancer (Oreskes and Conway 2010). This doubt was enough for people to be sceptical of the findings and, crucially for the tobacco industry, to continue buying tobacco. Instilling doubts about the link between smoking and cancer deprived many people of knowledge as they became less confident that the correlation was genuine (Cassam 2019a:89). The vice at hand here which led to a reduction of confidence in true beliefs, was the attitude vice of epistemic malevolence (ibid.:90).

By lowering one's confidence in beliefs and by reducing the number of evidence-based beliefs one forms, thinking styles, when consistently present, can obstruct knowledge in a way that is constitutive of epistemic vice.

We now have a clearer picture of Cassam's account of epistemic vice, obstructivism. Epistemic vices are character traits, thinking styles or attitudes that systematically obstruct knowledge, in the various ways outlined above. This answers the normative question of where the badness of these vices is located – they are bad in so far as they obstruct individuals and communities from accessing epistemic goods. In so far as traits, thinking styles and attitudes result in these

epistemic harms, Cassam argues that vice-bearers are responsible for their vices. So much so, that being responsible for said vice is an essential component to the definition of vice (ibid.:6).

2.5 Understanding Responsibility

As we have seen, the viciousness of these traits, thinking styles and attitudes are explained through a consequentialist lens under obstructivism. Vices are the result of epistemically (and sometimes morally) bad consequences, predominantly understood as the obstruction or loss of knowledge. What follows from these harmful failings is the belief that we must also be held responsible for these epistemic harms (ibid.:123). Cassam considers two dimensions of responsibility: acquisitional and revisional responsibility.

Cassam places restrictions on the conditions for which someone is responsible for their vices. In general, Cassam states that for a person S to be blameworthy for a vice it must be the case that the vice is epistemically harmful, and we are in some relevant sense, responsible for the vice (ibid.:124). As we have seen, this harm condition is spelt out via the effects that vices can have when they obstruct knowledge or other epistemic goods. What then, are the relevant ways in which a vice bearer can be deemed responsible for their vice?

One type of responsibility discussed by Cassam is *acquisition responsibility,* which states that a person is responsible for a vice if they are responsible for acquiring or developing it (ibid.). On this view, a vice-bearer is responsible and blameworthy for their vice because they are responsible for the past actions or decisions that led to the development of the vice.

However, Cassam does not consider this type of responsibility to be compatible with vice, claiming that it does not paint a plausible picture of true vice acquisition and is only concerned with the actual or imagined origin of one's vices (ibid.:125). Take for example, Heather Battaly's example of a man whose vice of dogmatism is acquired through the '...bad luck of being indoctrinated by the Taliban' (Battaly 2016a:100). In this instance, the vice-bearer had no control over its formation, yet we still want to criticize his dogmatism and claim that it is a vice (Cassam 2019a:19). Cassam also contends that the formation of one's epistemic character usually occurs in early childhood, something that agents have no control over due to a lack of

maturity and sophistication (2019a:128). We are therefore rarely blameworthy for the acquisition of our epistemic vices, meaning we must be blameworthy for them in some other sense.

Instead, Cassam argues that one way we could hold the indoctrinated vice-bearer responsible is to argue that he may not be responsible for *becoming* dogmatic but may be responsible for *being* that way now (ibid.). This is based on the distinction between permanent attributes, and attributes which are malleable and open to revision or modification through one's own effort (ibid:129). If an agent possesses a vice which they have the ability to modify or eradicate, then they crucially have control over them and thus can be responsible for them. This view of responsibility that Cassam is more optimistic about is called 'revision responsibility' and it focuses on what an agent can and cannot change or revise (ibid.:124)[9].

According to Cassam, the vital component of revision responsibility is the ability to control or modify the vice in question. To this end, Cassam details three different varieties of control: voluntary, evaluative, and managerial, arguing that it is managerial control which allows us to have the most effective control over our character vices.

Cassam acknowledges that we have no voluntary control over our character traits or attitudes since we cannot change them at will (ibid.:125). This applies particularly to beliefs, for example, I cannot will myself to believe that the sun is shining if it is not true. One way in which we can possess control over our beliefs, however, is through evaluative control. We can possess evaluative control over our beliefs in the sense that we can evaluate and re-evaluate what we take to be true.[14] Additionally, Cassam states that we can possess evaluative control over complicated attitudes such as hatred and contempt (ibid.:126). This is because if my contempt for someone reflects my beliefs about them then I am responsible for my contempt as long as I have evaluative control over those beliefs. If my beliefs are unjustified, and I have no other basis for my attitude, then I can be condemned for my contempt.

Finally, Cassam introduces a third form of control; managerial control. This is another form of control put forth by Hieronymi which is a kind of control present '…when we manipulate some ordinary object to accord with our thoughts about it' (Hieronymi 2006:53). For example, I

[14] Using an example offered by Hieronymi (2006:54), we can possess the belief that it takes 45 minutes to drive to the airport. However, if we leave at rush hour, we can reconsider our belief and change it accordingly. In this sense, we are said to be 'controlling' our beliefs.

control the direction and speed of my car when I turn the wheel or shift gear. It is also possible to have managerial control over our beliefs (2019a:126). For example, I might want to believe that I am unwell, so I look for evidence to support this claim. In the absence of evidence, I may resort to other methods of belief about belief management, such as self-hypnosis or positive thinking, with the aim being to manipulate my mind to produce in me the desired belief. It is also worth noting that Cassam understands managerial control as *indirect* when it concerns our character vices (ibid.:129). Whilst we can directly move a steering wheel to change directions, we modify our character vices or other traits indirectly e.g., by limiting our exposure to diverse perspectives and sources.

Cassam notes that these considerations highlight the complicated relationship between blameworthiness for vices and managerial control. To the extent that we have effective control over our vices, that type of control is managerial as opposed to voluntary or evaluative. Furthermore, managerial control is the type of control necessary for blameworthiness, unless one has a culpable lack of managerial control. Finally, managerial control over our vices is usually sufficient for us to count as revisionary responsible for them, meaning we can be blamed for them.

Cassam also believes that there may be cases where an agent possesses an epistemic vice which is not a malleable character trait. Take, for example, gullibility or foolishness, both character traits which are epistemically harmful in the sense that they obstruct knowledge. One may argue that these vices are not malleable because they are hard-wired in such a way that people who possess them cannot do anything about them. What would be the appropriate response to these types of epistemic vices? In response to this concern, Cassam appeals to a distinction between blameworthiness and criticism.

Cassam appeals to a distinction made by George Sher to carve his distinction between blame and criticism.[15] According to Sher, we can distinguish between blameworthiness and a trait reflecting badly on someone (2006:58). Whilst individuals can only be blamed for what reflects badly on them, it is not the case that individuals can be blamed for *everything* that reflects badly on them. Cassam uses this distinction to argue that when blame is not appropriate, but something reflects badly on the agent, criticism can instead be warranted.

[15] Cassam also uses the term 'reprehensible' in exchange with criticism (Cassam 2019a:23).

To explain further, Cassam claims that when it comes to the identification of intellectual or epistemic failings, what counts is not whether they define the type of person one is but whether or not they define the kind of thinker or knower one is. On this reading, gullibility is a deep intellectual defect and casts a negative shadow over those who suffer from it, even if they are not blameworthy for it. It reflects badly on someone to say that he is gullible in so far as these are negative traits that define the kind of intellectual or epistemic agent he is. These traits are not separate from him; they are a part of him and of who he is. Furthermore, to say that traits such as gullibility reflect badly on someone is to say that they can be criticized for them even if they cannot be blamed for them. This is what allows Cassam to categorise these and other such traits as epistemic vices regardless of whether they are blameworthy. If it turns out that they are also blameworthy then the classification of them as epistemic vices is even more straightforward, but the issue of blameworthiness cannot be settled without also settling the question as to whether these traits are malleable.

Finally, traits, attitudes and thinking styles do not have to be malleable in order to count as epistemic vices meaning they can be criticised even if an agent lacks control over them (2019a:134). Cassam's distinction between blame and criticism demonstrates why under his account, character traits, attitudes or thinking styles can be called vices still even if they lack blameworthiness as opposed to mere 'cognitive defects' (ibid.:127-128).

It is important to emphasize this distinction between blame and criticism under Cassam's account and how is criticism, not blame, that is integral to his definition of vice. An agent can still possess a vice even if they are not blameworthy for it, as long as it is an intellectual failing that warrants criticism (and meets the additional obstructivist requirements). In this sense, responsibility is inherent to the definition of an epistemic vice according to obstructivism. For V to be defined as an epistemic vice, it must be at least criticisable.

To summarize, Cassam proposes the following taxonomy for epistemic vices and responsibility. To begin with, there are cognitive defects for which neither blame nor criticism is appropriate. These are 'mere' cognitive defects as opposed to epistemic vices. An example of a cognitive defect is a handicap, such as blindness which prevents one from knowledge. It would of course be extremely inappropriate to criticize or blame someone for this. Secondly, there are intellectual failings that for one reason or another are not blameworthy but are open to criticism. These are epistemic vices. Lastly, there are intellectual failings that are not just

reprehensible but are also blameworthy, in the epistemic understanding of blame. This blame is based on revision responsibility, which states that for an agent to be blamed for V they must be in control of it in order to be able to revise or modify it.[16]

2.6 Acquisitional versus Revisional Responsibility

The subsequent sections of this chapter will focus on Cassam's endorsement of revision responsibility for epistemic vices and his rejection of acquisitional responsibility. I challenge his preference for the former and argue that his reasons for rejecting the latter are uncompelling.

Firstly, to understand the broader context of revision responsibility as a type of blame, it is necessary to consider the distinction between two concepts of moral responsibility: responsibility as 'attributability' and responsibility as 'accountability'.[17] Attributability concerns the relation between an agent and their action, belief, or trait. Informed by metaphysics and action theory, attributability responsibility holds that we are responsible for qualities that can be properly attributable to us (Hieronymi 2008; Sher 2006; Shoemaker 2015; Talbert 2012; Watson 2004). An attribute can be said to 'belong' to us in the appropriate way if it reflects our motivations, values, or attitudes. In so far as our actions can reflect some aspect of ourselves, it can provide grounds for negative or positive appraisals. This appraisal attaches to the person *qua* agent and can take the form of praise or blame (Hieronymi 2008:358; Talbert 2008, 2019; Zheng 2016:65).

As is evident then, we are responsible in this sense when something is properly 'attributable' to us. There are a number of conditions in which actions cannot be attributed to agents, referred to as 'excusing conditions'. These conditions include whether an agent was under coercion when they acted or if they acted accidentally, as well as conditions in which a person is not acting as their 'full agent' i.e., children and non-animals who have not developed an epistemic character (Scanlon 1998: 278–85, Sher 2006, 2009; Smith 2005).

[16] Cassam notes that since revision responsibility for an epistemic vice can vary from person to person, the same epistemic vice can be blameworthy in some cases without being blameworthy in every case. In this sense, vices are personal, and not necessarily universal (2019a:22).

[17] Shoemaker (2015:88) argues that there are three kinds of moral responsibility: attributability, accountability, and answerability. I explore this third form in more detail in Chapter 3 of this thesis.

Accountability, on the other hand, is a more practical form of responsibility which originated from political philosophy (Shoemaker 2015; Watson 1996, 2004). According to this view, individuals are held responsible for an action only when they had a realistic chance, either directly or indirectly, to refrain from breaching the standards for which they are being held accountable for (Watson 2004:276).[18]

Returning to Cassam's revision responsibility, it is apparent that Cassam understands blame akin to accountability. He explains the condition for which 'S is accountable is one that S has the ability to control or revise by her own efforts. When this is the case S has revision responsibility for V' (Cassam 2019a:124). Furthermore, he states that 'responsibility is (a matter of) accountability: to view a person as responsible for being foolish or gullible is to regard them as accountable for their vice' (ibid).

What about acquisitional responsibility? As we have seen, Cassam rejected this form of responsibility for vice due to the claim that vice-bearers often lack control over the acquisition of their vice (ibid.) Attributability responsibility, however, holds that we are responsible for attributes, even if we lacked control over their formation (Scanlon 1998:278–85, Sher 2006, 2009; Smith 2005).

If a form of responsibility for vice is possible that does not require control, what are Cassam's reasons for rejecting it? We can recall that Cassam objections were based on the claim that one's epistemic character is something that is formed early in one's life. As children lack a degree of maturity, it would be wrong to deem them blameworthy for acquiring their epistemic character traits, even if they are epistemically vicious (ibid.:128).[19]

Three responses can be given to this claim. Firstly, one's character can change over time, developing beyond your childhood into your adult years. It is therefore not necessarily the case that vice-bearers are too immature to be held responsible for the formation of their traits or character.[20] Secondly, there is a distinction to be made between exercising control over the

[18] There is some variation between how accountability and attributability responsibility are described and their differences (Fischer and Tognazzini, 2011). I therefore define both forms of responsibility broadly, returning to the distinction in more detail in Chapter 3 of this thesis.
[19] See Battaly (2019) for an objection to Cassam's responsibility condition for vice. I address Battaly's broader concerns for non-voluntarist interpretations of vice-responsibility in Chapter 4 of this thesis.
[20] Driver discusses how moral character, and therefore the moral virtues one possesses, can change with time (2001:84-85).

formation of one's character traits and control over one's overall epistemic character.[21] This means that a lack of control over our overall character does not extend to a lack of control over the formation of individual traits. As we are concerned with character vices, we should be focused on the latter, making this a weaker claim to accept – individuals can possess some element of control over the character traits they form. Thirdly, as noted above, attributionist accounts of responsibility can be subject to 'excusing conditions' that excuse one from responsibility, including '…a lack of well-formed character or the capacities required to reflectively deliberate and choose ends' (Zheng 2016:65).[22] These excusing conditions may accommodate Cassam's concerns, meaning one can accept that a lack of control over the formation of one's vice is still compatible with acquisitional responsibility.

It seems then, that we should not rule out the possibility that acquisitional responsibility is incompatible for epistemic vices on the basis that we may lack control over the initial stages of vice acquisition. Cassam also needs to present a stronger case for why control is a necessary element for responsibility over epistemic vices if it can be argued that responsibility does not require control (Hieronymi 2008; Sher 2006; Shoemaker 2015; Talbert 2012; Watson 2004).

Alongside his rejection of acquisitional responsibility, we can also object to Cassam's arguments in support of revision responsibility. As a reminder, revision responsibility has stronger control conditions as it requires the vice-bearer to be able to control or revise their vice by their own efforts (2019a:124). One key concern arises in situations where it is unjust or unsuitable to expect an agent to revise their vices, which presents a challenge to Cassam's position.

Drawing from Sher (2006) on control and moral vice, Cassam quotes a variety of ways in which one may attempt to revise their vices, such as 'reflect[ing] on his past lapses, forc[ing] himself to do what does not (yet) come naturally, imitate exemplary others, and avoid those whom he knows to be bad influences' (2006:55). However, there are many cases where agents do not possess the intellectual capabilities to overcome their epistemic vices through self-improvement or reflection. One reason may be down to the options available to the agent, so

[21] One concern here may be that one's overall character is nothing over and above the collection of singular character traits. If this is true, there could be no distinction between the two. This seems to be a claim that Cassam would be unlikely to accept, however, seeing as he believes one's character can also consist of attitudes thinking styles and cognitive biases. Instead, we can think of epistemic vices are defects of one's character, and virtues are excellences of one's character (Baehr 2020:24).
[22] See Strawson (1962) for further details on other excusing conditions.

despite being aware of their epistemic vices and wishing to revise them, they cannot. This could be the result of a host of social or situational factors, such as a lack of access to good education.[23] Taking these considerations into account, blaming agents for these vices (based on Cassam's favoured account of revision responsibility) might seem unfair when agents are unable to revise their vices due to practical reasons.

Cassam could address this issue by suggesting that in such instances, the suitable course of action would be to criticise the vice bearer instead. He may recognise that agents do not always have the capacity to alter or amend their vices, and the ability to do so is a prerequisite for revision responsibility. However, Cassam believes that blaming or criticising vice-bearers should serve an ameliorative aim (2019a:49). Yet, what constructive outcomes can be derived from critiquing agents who aspire to become better epistemic agents but lack the means to achieve it? It is therefore unclear in what sense vice-bearers of this kind can be responsible according to Cassam.[24]

Cassam also encounters challenges with revision responsibility when agents are *not* aware of their vices. This draws on Cassam's categorization of 'stealthy vices', epistemic vices which block their own detection to the possessor 'to the extent that it nullifies or opposes the very epistemic virtues on which active critical reflection depends' (ibid.). Take, for example, a dogmatic Brexiter who refuses to listen to arguments in favour of the UK remaining in the EU. How do we go about holding this individual responsible for their epistemic wrong? It seems a fair assessment to say that the agent has control over their vice, so blame in the form of revision responsibility is viable. However, the agent is unaware of their vice as it is stealthy, meaning in practice, they will not be able to update or revise their behaviour.

Cassam's resolution to address concerns about compatibility issues between stealthy vices is to argue that vice-bearers can be deemed culpable for their self-ignorance over the possession of their vice (ibid.:166).

In brief, Cassam argues that at first glance the vice-bearer's self-ignorance can mitigate their responsibility and culpability for their epistemic vices if their ignorance is not itself culpable.

[23] See Kidd (2020) and Medina (2012) for detailed discussions on how vices relate to one's social status.
[24] At the very least, Cassam needs to expand on the ameliorative role that blame can play in such instances. See Chapter 4 of this thesis for an evaluation of this view.

To be culpable for your self-ignorance you must be responsible for it, and this is unclear in instances of stealthy vices. If the vice-bearer of the stealthy vice has no knowledge of their vice, they cannot be culpable for their ignorance. This is because the vice-bearer lacks the relevant managerial control over their vice. They cannot make appropriate revisions to their behaviour if they are unaware of what their behaviour is in the first instance.

However, Cassam asserts that we need to look at the reasons as to why the vice-bearer is ignorant of their vice and assess whether these reasons are culpable ones. If it's because of a mental disorder such as depression, we are blameless for this ignorance. However, with stealthy vices, the reason we are ignorant is because of the vice itself. It is causing us to not revise our behaviour or accept negative information about our vicious ways. This brings us back to focusing on whether this vice is culpable or not.

Cassam (ibid.:166) presents an example of Donald Trump and his seemingly stealthy vice of epistemic incompetence to explain this further. Trump is blindly unaware of his incompetence. It seems then that he cannot be culpable for his vice, as he cannot exercise the relevant managerial control. However, this blind unawareness is directly caused by his vice itself. If his epistemic incompetence is culpable, which we believe it is, then Trump can be blamed for his self-ignorance and vice. As Cassam says 'It is no excuse that he (Trump) is so incompetent that he can't get the measure of his incompetence.' (ibid.).

To summarise, when a vice-bearer is ignorant over their possession of their vices, Cassam argues that we need to look at the reasons for one's self-ignorance and trace culpability back to these very reasons. For stealthy vices, the reason for self-ignorance is the vice itself, making it culpable.

To evaluate this response, Cassam seems to now dramatically weaken his conditions on control for responsibility. In detailing such as case where the vice-bearer is culpably self-ignorant for their vice, Cassam states that 'If she is culpably self-ignorant, and her self-ignorance accounts for her lack of managerial control over V, then she is potentially blameworthy for V despite her lack of managerial control' (ibid.:130).

To recap, the three key conditions for managerial control over a vice were one, to have a trait which is open to manipulation, two, to know one has the trait, and three, to be motivated to

change it. For stealthy vices, condition two is now dropped, the agent need not be aware of their vice, and as a result condition three is also redundant. The vice-bearer cannot be motivated to change their vice if they are unaware of it. An objection that emerges as a result of this shift is that Cassam seems to forsake the control condition in the context of responsibility for stealthy vices (Beaton et al. 2019:54). It now seems that we can now be blameworthy for our ignorance even if we have *no control* over the ignorance itself (Cassam 2019a:166). This is clearly a significant shift from Cassam's earlier commitment to control for responsibility which weakens his original stance.

On an alternative view, Cassam may avoid this contradiction via his claim that 'if the only thing preventing one from knowing one's vices is those very vices, then one's ignorance is culpable' (ibid.). It appears that in advocating for responsibility over stealthy vices, Cassam is relying on a 'tracing strategy' for responsibility. This strategy is used to explain how individuals, who in the moment of action, do not meet the control requirements for responsibility (Talbert 2019). Despite this, they still seem responsible for their actions. In these instances, the agent's responsibility may be traced back to an earlier occasion where they did meet these control conditions. For example, an individual may be so intoxicated that they lack appropriate control over their actions, making them seemingly blameless. However, the individual is responsible for choosing to freely intoxicate themselves. Here we can trace the responsibility back to a moment where the control conditions were met. Cassam's argument may follow a similar line of thinking. If one is not responsible for their stealthy vices due to their self-ignorance, we can then trace the responsibility back to the causation of the ignorance which is the vice itself. From there we can assess whether the vice-bearer was in control of their vice through the conditions of managerial control.

However, tracing the self-ignorance back to the vice does not seem to go far enough. Considering the vice is still a stealthy vice, we end up facing the same concerns where the managerial control conditions have not been met. To trace it back to a moment where it was not stealthy using the tracing strategy would plausibly be the moment of acquisition, for the vice may not yet be undetectable to the vice-bearer. However, as we are aware, Cassam contends that we are not in control or responsible for our vice acquisition.

Additionally, if Cassam maintains that control is still relevant for responsibility over stealthy vices, as many prominent moral theorists point out, dropping the awareness requirement for control may be problematic. Neil Levy for example, states that agents cannot be in control of causing alterations if they do not know that they are doing so '…if moral responsibility requires control, then it requires that we know what we are doing' (Levy 2005:5). Whilst the awareness requirement on our vices may be disputed, it is a significant challenge for accounts of vice that stipulate control is a necessary requirement. If Cassam is consistent and claims that vice-bearers have control over their stealthy vices, then he must explain the compatibility between control and the lack of awareness. If he abandons the control condition, he exposes an inconsistency in his account of vice.

To summarise, Cassam runs into multiple concerns when explaining how vice-bearers can be responsible for stealthy vices or instances where one is aware of their vice but cannot revise it. It seems that a form of blame that requires awareness and active revision or control over one's behaviour, is again ill-suited to these types of vice.

2.7 Conclusion

This chapter has examined Cassam's account of epistemic vice termed obstructivism and explored various objections against this view. I focused first on the 'obstructivist claim', examining how character traits, attitudes and wishful thinking can systematically obstruct knowledge and prevent other epistemic goods from being acquired. I raised an objection concerning the compatibility of the systematic clause and low-fidelity nature of many vices, arguing that Cassam should focus on whether the epistemic harms caused by vices were systematic, as opposed to whether the vice itself occurred systematically. This is distinct from fidelity, which is concerned with how many times the behaviour occurred and what threshold needs to be applied. If a supposed instance of vice resulted in epistemic harms 'more often than not', and the other conditions for vice were met, the behaviour could be deemed epistemically vicious (2019a:38).

Subsequently, I shifted the focus towards evaluating the normative claim that vices are blameworthy or at least criticisable. I discussed Cassam's distinction between acquisitional and revisional responsibility and his distinction between blame and criticism. I then critiqued

Cassam's reasons for rejecting acquisitional responsibility, which was based on the claim that we lack control over the formation of our vices, meaning blaming vice-bearers for vices on this basis would be unfair. I responded to Cassam's claim that a lack of control does not necessarily mean a lack of responsibility, as highlighted by attributability responsibility (Hieronymi 2008; Sher 2006; Shoemaker 2015; Talbert 2012; Watson 2004). I also raised three concerns with his specific claim that we lack control over the formation of our vices as they are formed in our early lives. I argued that one, our epistemic character can change over time, being formed and shaped in our adult years, two, we can be held responsible for individual traits as opposed to our entire epistemic character, and three, there are many 'excusing conditions' that could accommodate Cassam's concerns, meaning we can accept that vice-bearers may lack control over the formation of their vice, but still be considered blameworthy for it.

Finally, focusing on Cassam's support for revision responsibility, I argued that stealthy vices, or instances where one is aware of their vice but cannot revise it, created problems for Cassam's claim that we are responsible for vices if they are under our control and can be revised. I argued that these scenarios contradicted his ameliorative goal of holding vice-bearers responsible and led to a contradiction in his commitment to managerial control.

CHAPTER 3. UNVEILING VICIOUS MOTIVES: THE MOTIVATIONALIST PERSPECTIVE ON VICE

3.1 Introduction

This chapter will focus on motivationalism, defined broadly as the view that good or bad motives are constitutive to epistemic virtues and vices (Baehr 2010; Battaly 2016a, 2018a; Montmarquet 2000; Tanesini 2018, 2021; Zagzebski 1996).[25] This contrasts with consequentialist views, which hold that epistemic vices are bad because of their harmful effects (Cassam, 2016, 2019a; Driver 2001).

To get a clear picture of the general motivationalist view, we can use Linda Zagzebski's distinction between 'motives' and 'ends' (1996:179). According to Zagzebski, an agent's end identifies a specific objective that they wish to pursue, whilst their motive refers to the driving force that arises from that end. Epistemic virtues are therefore partly understood as a disposition to be motivated by a particular set of ends e.g., the epistemically humble individual will care about and reflect on their ontological commitments, beliefs, and biases. Charlie Crerar (2018:755) refers to these as the proximate ends of an epistemic virtue. He observes how being motivated by these ends alone is not enough to make you epistemically virtuous. These proximate ends need to be grounded in a fundamental motivation for epistemic goods such as knowledge, truth and understanding, referred to as the 'ultimate' motive (ibid.). For example, an epistemically humble agent is motivated towards reflection because of their ultimate motivation for epistemic good.

Within vice epistemology, the motivational view assumes the same broad position but in reverse. Crerar elaborates on this in two ways. Firstly, epistemic vices might be characterised by the presence of bad epistemic motivations, such as being motivated by epistemic bads or away from epistemic goods. Crerar refers to this as the 'presence conception' (ibid.). Secondly,

[25] Not all responsibilists use the term 'motivation' to describe their stance. For example, Baehr talks of a 'love' of epistemic goods (2011:101). Broadly, however, they all concur on the notion that virtue entails a favourable orientation towards epistemic goods, which 'motivation' captures.

epistemic vices might involve the absence of good motivations, by failing to value epistemic goods or the ultimate ends. This is referred to as the 'absence' conception' (2018:758).[26]

Heather Battaly and Alessandra Tanesini's accounts of epistemic vice are most closely aligned with the presence conception, arguing that vices are 'partly composed of bad epistemic motives' (Battaly 2017a:7) or 'are guided by motives to turn away from epistemic goods' (Tanesini 2021:21). We can now turn to evaluate each account, referred to as 'personalism' by Battaly (2016a:99) and simply the 'motivational account' by Tanesini (2021:22).

The plan for this chapter is as follows. To start, I will evaluate Battaly's account, which advocates for a more moderate form of motivationalism as vices are only partly explained via their bad epistemic motivations. Her account is presented as a medium between two competing analyses of epistemic virtue and vice: reliabilism and responsibilism (Code 1987). I begin this chapter by outlining the key features of both of these conceptions of epistemic virtue and how Battaly combines selected features of them to create her account of vice. In the following section, I argue that personalism has limited scope and fails to successfully demonstrate that vice-bearers are not responsible for their vices. Finally, I will explore how attributability responsibility, a form of responsibility that Battaly seems later inclined to, is incompatible with the rest of her account.

In the second half of this chapter, I turn to examine Tanesini's (2018, 2021) account of epistemic vice, which argues for a robust form of motivationalism informed by psychological and empirical research. I focus on the different classifications of vice, including sensibilities, thinking styles and character traits, alongside their relevant motivational components. I also examine the role of responsibility and blame in Tanesini's account and her argument in favour of attributability responsibility. In my criticisms of Tanesini, I assess her objections to holding vice-bearers accountable for their vices, both epistemically and morally. I also evaluate Tanesini's endorsement of 'taking responsibility' for our own epistemic vices, which I argue is undermined by her view that most vices are stealthy.

In summary, this chapter aims to examine motivationalism as a theoretical framework for understanding epistemic vices as presented via the two, aforementioned motivationalist

[26] Crerar argues that there are also instances where agents possess good motives but warrant ascription vice. This underpins his inversion thesis, the claim that epistemic virtue and vice are not complete opposites (2018).

accounts of vice. I will focus specifically on evaluating the responsibility component of these accounts and whether they consider vice-bearers to be blameworthy for the formation and expression of epistemic vices.

3.2 Battaly's Personalism

Battaly introduces her account as a medium between both virtue responsibilism and virtue reliabilism (2016a, 2018a)[27]. She identifies five of the key features from both categories and subsequently outlines which aspects from each her account incorporates.

First, reliabilist virtues and vices need not be acquired qualities as they can include hard-wired faculties (Greco 2010; Sosa 2007). For example, vision, if reliable, is a hard-wired virtue in the sense that our brains are wired to produce beliefs based on visual inputs.[28] Second, we need not be responsible for reliabilist qualities in order to be epistemically virtuous or vicious. In particular, we do not need to be responsible for possessing the virtues or vices we hold, nor do we need to be responsible for the operation of them. Resulting from this is the claim that we need not inherently be praised nor blamed for the possession of reliable faculties (virtues) or unreliable ones (vices). Third, epistemic virtues and vices are not necessarily personal qualities but are instead 'sub-personal' (Battaly 2018a:116).[29] This is because personal qualities express one's character – one's epistemic values, motivations, judgements and so forth. Conversely, hard-wired epistemic virtues or vices such as vision are sub-personal as they tell us nothing about the epistemic character of the agent[30]. Fourth, reliabilism claims that intellectual virtues must demonstrate reliability by having a predisposition to generate a majority of accurate beliefs. Greco (2010) and Sosa (2007) emphasize these virtues need to possess this disposition specifically in the typical conditions we encounter. For example, one's vision may not provide reliable outcomes in the dark. However, in situations where we do anticipate reliability, it should assure us that what we perceive exists. Fifth, the value of reliabilist epistemic virtues and vices is often instrumental and need not be intrinsically so. They are valuable because they

[27] See also Battaly (2016b, 2017a) and Slote and Battaly (2016).
[28] Not all visual perceptual competencies are 'hard-wired' for virtue reliabilists. For example, consider a birdwatcher who possesses the ability to distinguish between a chaffinch and a goldfinch does not possess a hard-wired trait.
[29] For more on how personal qualities express the type of thinker we are, see Baehr (2011) and Montmarquet (1993).
[30] Battaly notes that Greco (2010) and Sosa (2015) have argued that some reliabilist qualities involve a motivation to seek truths and avoid falsehoods. However, Sosa does not believe that this motivation for truth must be a personal quality informed by one's epistemic values or motivations. Battaly also emphasises that reliabilism claims that epistemic virtues *can* be personal but *need not* be (Battaly 2018a:117).

consistently get us true beliefs, which are fundamentally or intrinsically valuable too. Similarly, epistemic vices, like unreliable vision, are dis-valuable because they consistently get us false beliefs (2018a:117).

Moving onto responsibilism, Battaly also identifies five of its important features. First, epistemic virtues and vices must be acquired qualities; they cannot be hard-wired faculties. This is because we cannot be praised or blamed for possessing hard-wired faculties, as their possession is outside of our control (ibid.:118).

What follows from this is the second feature, that we must be responsible for our epistemic virtues and vices, in so far as they reflect some acquired trait that we are in control of. Epistemic virtues must therefore be praiseworthy and epistemic vices must be blameworthy.[31]

Third, epistemic virtues and vices must be personal and closely connected to one's epistemic character. Specifically, Battaly states that in order to express an individual's epistemic character, intellectual virtues must be partly constituted by internal psychological features such as epistemic motivations and value commitments (ibid.:118-119).[32] Battaly also notes here that on the responsibilist picture, agents must have some control over the development of their epistemic motivations and values.

Fourth, responsibilists agree that reliability is 'conceptually insufficient' for epistemic virtue but disagree on whether virtues conceptually require reliability (ibid.:119). For example, Zagzebski (1996) argues that virtues require success, specifically a motivation for producing true beliefs and success in attaining that end (reliability). On the other hand, James Montmarquet (1993) and Jason Baehr (2011) argue that reliability is not necessary for intellectual virtue because intellectual virtue is (sufficiently) subject to our control, whereas reliability is often down to luck. Our motivations and values are (sufficiently) subject to our control, but reliability is not, and thus is not required for intellectual virtue.

Fifth, responsibilists contend that epistemic virtues should possess some intrinsic value since they are partly formed by motivations and commitments that themselves have intrinsic value, such as the desire for genuine knowledge (Battaly 2018a:119; Sher 1992:93; Zagzebski

[31] Many virtue and vice epistemologists interpret this responsibility requirement differently. For example, Montmarquet (1993) advocates for responsibility over the *operation* of virtues and vices, whilst Zagzebski (1996) focuses on responsibility for the *possession* of virtues and vices.

[32] Regarding epistemic value commitments, responsibilists argue that intellectual virtues require true (or at least justified) beliefs about what is and is not epistemically valuable. This view is proposed by Baehr (2011:102) who argues that an intellectually virtuous agent will love what they consider epistemic goods and hate epistemic bads.

1996:80). Similarly, intellectual vices will be intrinsically dis-valuable. A disposition will be deemed an epistemic vice if it is shaped by inherently negative motivations and commitments, such as the inclination to choose the easiest answer. There is disagreement between responsibilists as to whether virtues and vices must also be instrumentally (dis)valuable. For example, Zagzebski (1996) argues that intellectual virtues must be both intrinsically and instrumentally valuable as they require good motivations and reliability. Conversely, for Montmarquet (1993) and Baehr (2011), they must be intrinsically (or fundamentally) valuable but need not be instrumentally valuable (Battaly 2018a:120).

As highlighted, Battaly positions her perspective on vice as a bridge between responsibilism and reliabilism. After outlining the key attributes of both viewpoints, we can now turn to explore the specific features that Battaly embraces within her account of vice.

First, Battaly argues that motivations play a constitutive role in the definition of epistemic virtues and vices. Borrowing here from responsibilism, personalism holds that intellectual virtues and vices must be personal, in the sense that they express an individual's epistemic character and are therefore constituted by epistemic motivations and value commitments (2018a:120). Subsequently, personalism asserts that intellectual virtues and vices must possess intrinsic value. This means that an epistemic character trait cannot be considered an intellectual virtue unless the epistemic motivations and commitments which contributing to its formation are intrinsically good.

Whilst personalism favours this responsibilist feature, it also favours the second reliabilist feature, that agents do not need to be responsible for their epistemic virtues or vices.[33] More specifically, vices do not necessarily warrant praise or blame. This feature is motivated by Battaly's claim that individuals might not exercise control over the possession of their character traits, meaning they should not be responsible, particularly in the voluntarist sense, for their epistemic virtues or vices (2018a:120).[34]

[33] Tanesini observes that Battaly is unclear whether her focus is on moral or epistemic responsibility (2021:179).

[34] Battaly distinguishes between operation and possession personalism (2016:106). Operation personalism claims that epistemic agents are not required to be responsible for the operation of their virtues and vices. Battaly highlights its attractiveness to free will sceptics like Derek Pereboom (2014), who argue that many individuals lack the necessary free will for ordinary responsibility due to their limited control. Possession personalism holds that individuals need not be responsible for having epistemic virtues or vices. This is based on the idea that these traits are often shaped significantly by one's environment as opposed to conscious efforts and intentions.

Battaly presents an example of vices that are the result of indoctrination (2016a:108, 2018a:121). Consider a child who has been raised by the Hitler Jugend or ISIS. In this case, the child becomes conditioned to exhibit closed-mindedness, mirroring the attitudes prevalent in their surrounding community and adopting the relevant closed-minded epistemic motivations and value commitments. Battaly highlights that this particular vice is personal, as it expresses their epistemic character and stems from their bad epistemic values and motives. Additionally, the indoctrinated vice-bearer is not responsible for becoming closed-minded in the voluntarist understanding where blame requires control. This is because the vice was involuntarily acquired by being part of a vicious environment. This is an example of a personalist vice, one that is personal to the vice-bearer, but they are not necessarily responsible for it. What makes the display of closedmindedness vicious is its bad motivational component, not the fact that it is blameworthy.

Finally, Battaly notes that personalism can be filled out in a variety of ways. For instance, personalists can disagree as to whether intellectual virtues require reliability (though there is consensus that reliability is not sufficient for intellectual virtue), whilst others may require that epistemic virtues and vices are to be acquired (2018a:120).

The two key claims of personalism are therefore as follows. One, responsibility and character can come apart, meaning we are not necessarily blameworthy for our vices or praiseworthy for our virtues. This aligns with the reliabilist claim. Two, virtues and vices are personal, meaning they reflect one's motives and value commitments**.** Crucially, these motives and values are intrinsically good or bad, which explains how traits can be deemed virtuous or vicious. This aligns with the responsibilist claim.

Battaly pre-emptively identifies and responds to objections directed at personalism (2016a, 2018a). I will examine a selection of these objections, including some additional ones, with a primary emphasis on the responsibility and motivational components of vice that form the central components of Battaly's personalist account.

3.3 The Scope of Personalism

One initial point to briefly consider is the uncertainty surrounding the alignment of some epistemic vices and the personalist definition.

Cassam raises an objection when attempting to distinguish motivations from viciousness, citing vices such as stupidity that appear to lack a motivational component entirely (2019a:16). If stupidity is devoid of a motivational component, then Battaly would have to commit herself to the view that it is not an instance of epistemic vice. However, this stance seems counterintuitive and inconsistent with her classification of other traits as epistemic vices.[35]

Additionally, there are some vices where it is unclear if the motive is intrinsically bad. Cassam presents closedmindedness as an example, which involves a need for closure or a firm answer. He observes that this motivation does not seem inherently bad, despite being characterised as a vice, and can even result in potential benefits[36]. What is a more reliable guide then, is whether the vice results in bad epistemic effects and not whether the motivation is bad (Cassam 2019a:16-17).

In an anticipated response, Battaly may argue that vices such as stupidity may be better explained through a different analysis of vice (Battaly 2018b:28). For example, stupidity may be an 'effects' vice, defined as vicious due to its consistent bad effects or lack of good effects. Personalism is therefore just one of several ways to interpret vices, and it may be unsuitable for this specific vice.

Determining which vices will be analysed through the personalist lens will depend on the scope of personalism. This leads to a further objection discussed by Battaly herself, which is that personalism might be limited to vices cultivated through indoctrination (2018a:121-122). The paradigmatic examples of personalist vices are those acquired as the result of indoctrination. However, if we wish to acknowledge vice formed outside of these means, Battaly considers extending personalism to encompass many other instances of vice. She argues that in order to know just how far personalism extends, we need to know how much control one has over their character formation (ibid.). Non-voluntarists such as George Sher (2006, 2009) and Miranda

[35] Crerar also outlines three ways in which epistemic traits can still be categorised as vices despite being orientated towards epistemic goods (2018). This further underscores the divide between motivations and viciousness.
[36] Battaly also discusses how closedmindedness can be virtuous in some epistemic contexts. Interestingly, what differentiates closedmindedness as a virtue from closedmindedness as a vice, is the effects it has on one's surrounding environment and not whether its corresponding motives are good or bad (Battaly 2018b:29).

Fricker (2007, 2016) argue that we have limited control over the development of our character traits. Sher (2006) suggests that we rarely consciously shape our traits, especially during childhood. Even as adults, it's challenging to predict how actions influence traits. Fricker (2007, 2016) believes that we typically inherit our character traits from society, often acquiring vices due to societal issues. Drawing on these claims, Battaly argues that many of our virtues and vices would fall under personalism. This implies that personalism might have a broader scope than responsibilism, meaning 'responsibilism could be the exception, and personalism the norm.' (Battaly 2018a:122).

However, if Battaly does commit herself to the view that personalism encompasses most varieties of vice, it follows that we are not responsible for most displays of vice either, as the control conditions of personalism dictate that we are often not responsible for what is outside of our control (2018a:120). Battaly does not provide an argument for this bolder claim, however, and as we will discuss momentarily, the absence of control does not necessarily equate to the absence of responsibility.

In other words, a result of expanding the scope of personalism is the claim that we are not responsible for most instances of epistemic vice. This conclusion follows from Battaly's claims that we should not be blamed for what is outside of our control and that we are not in control of most of our traits. Let us now turn to evaluate Battaly's arguments concerning responsibility and control.

3.4. Motivations and Responsibility

In another anticipated objection, Battaly considers whether the indoctrinated, closed-minded individual is truly vicious or just possesses a 'bad' epistemic character trait that can be defined in some other way (2016a:111, 2018a:122). The rationale behind this perspective is that for something to qualify as a vice, it must inherently possess some responsibility condition, such as blameworthiness. In order words, a vice contains a built-in responsibility component, and to deny this component, as personalism does, would be to describe some new quality altogether. These qualities may be better described as impairments, defects or aptitudes (Cassam 2019a).

This line of reasoning is proposed by Zagzebski (1996:118-121) who argues that virtues fundamentally require the individual who possesses them to be deserving of praise. To illustrate this point, Zagzebski employs Nozick's transformation machine, a hypothetical device that can provide any desired experience (1974:44). Attempting to generate virtues or vices via this machine would be unsuccessful according to Zagzebski. Virtues require experience and motivation which are only possible through human development. This means that we would not praise an agent for possessing the machine-virtue, as they have failed to possess it in any meaningful way. This quality also would not be a virtue as it is not praiseworthy.

Battaly can respond to this objection by reiterating her reasons for rejecting this responsibilist claim. Firstly, we can sufficiently explain the 'viciousness' of vices without relying on a responsibility component (2016a:111-112, 2018a:122). Secondly, blame is not always a suitable response to vices, indicating that it cannot be an intrinsic feature (2016a:107-108, 2018a:121).

Let us start with the first of Battaly's claims and return to the second in the subsequent objection.

Battaly contends that the badness of a vice can be traced back to its motivational component. This means that to determine whether a vice is bad, we do not need to know whether it is inherently blameworthy, but rather understand the motives and values behind the trait (2016a:111-112).

One example that highlights this line of reasoning is Gary Watson's (2004) example of a man named Robert Harris, who, at the age of 25, murdered two people in San Diego in 1978. Harris was consistently abused as a child by his parents and was later abandoned by them. As a result, he spent much of his childhood incarcerated in a juvenile detention centre. Watson argued that Harris possesses the vice of cruelty. In Watson's view, Harris's upbringing does not undermine the judgement that he was '…brutal, vicious, heartless, mean' rather it '…provides a kind of explanation for his being so' (Watson 2004:242). Watson argues that Harris's cruelty was a result of his unfortunate circumstances over which Harris had no control, and for which he was not accountable. In Watson's view, Harris had the vice of cruelty, though he was not accountable for possessing it, consequently meaning he was not blameworthy for the vice too.

Battaly suggests that it is unsurprising that Harris ended up with the vice of cruelty. His upbringing moulded distinctive characteristics in him, leading to a skewed perception of value. His skewed view placed importance on suffering and actively pursued it. These were the essential belief and motivational elements of his vice of cruelty. These qualities were bad enough to demonstrate Harris was a bad person. Not only did his cruelty result in bad consequences, but more importantly it was because of the beliefs and motivations involved. These motivations are intrinsically bad, 'even if we can't act on them, and had no control over their acquisition.' (Battaly 2016a:112).

Based on the case study of Harris, Battaly argues that it is the motivations that show us Harris was a bad person. Harris's 'cruelty was bad because of the beliefs and motivations it involved' (ibid.). This is a sufficient vice-explanation. There is no need to appeal to responsibility to understand the vice, particularly as in Battaly's view Harris's upbringing undermines his responsibility for possessing his vice.

This example reinforces Battaly's claim that the viciousness of a vice can sufficiently be explained through its bad motivations without inquiring into whether the vice is inherently blameworthy. Consequently, we can reject the responsibilist claim that vices are inherently blameworthy, as it is not essential for explaining the badness of a vice. As Battaly puts it 'why would vices require the additional dis-value that comes from being blameworthy for their acquisition, when the dis-value of their belief and motivational components already accounts for one's being bad qua person' (ibid.).

In response to Battaly's argument, I argue that it is unclear why the motivationalist claim must replace the responsibilist one. Whilst responsibility may be an inherent feature of some accounts of vice, it does not function as an explanation for the vices' badness. Instead, it operates as a reaction to this badness (Cassam 2016, 2019a; Zagzebski 1996). Blame, therefore, is a reaction to some kind of epistemic harm or wrongdoing, whether this wrongdoing is spelt out as bad epistemic consequences or bad epistemic motivations.

This means that offering a motivational explanation for the viciousness of vice does not automatically negate the responsibility component of vice. It is perfectly reasonable that an account of vice can locate the viciousness of epistemic vices in bad motivations, of which the vice-bearer is also responsible for (Tanesini 2021). Therefore, simply providing a motivational

explanation for the viciousness of a vice does not serve as an objection to the responsibilist claim that we are inherently responsible for our vices. In other words, by demonstrating that the motivational component of vice is sufficient to account for its viciousness, Battaly does not present a compelling argument against the responsibilist claim that blame is not a necessary response to vice. A bad motivation can explain the viciousness of vice, and this bad motivation can still be inherently blameworthy. In order to successfully challenge this responsibilist claim, Battaly would need to establish that blame does not inevitably follow as a response to a bad motivation.

To summarise, Battaly's claim that motivations can explain away the responsibility component of vice does not constitute a substitute explanation of the badness of epistemic vices. It is conceivable that motivations can explain the viciousness of a trait, and blame be a response to said trait or motivations. In this sense, the responsibility component of vice remains very much intact.

3.5 Changing Direction: Attributability Responsibility

Another concern that Battaly anticipates is that she might be letting vice bearers such as the Hitler Jugend or Harris off too easily by not blaming them for their vices (Battaly 2018a:124). Alongside her argument for the motivational explanation for vice, Battaly's further argument is that these vice-bearers were not in control over their vice, meaning they cannot be fairly blamed for it. However, to some, this may seem to 'excuse' the badness of the vice-bearer's behaviour.

In response to this, Battaly argues that personalism *can* accommodate a responsibility condition for vice, namely in the form of attributability responsibility.[37] Battaly refers to this form of responsibility as non-voluntary and argues personalism can accommodate it in at least three of its forms. Firstly, it is compatible with Watson's understanding of attributability responsibility as an expression of one's 'real self' and the motives and values that they endorse (Watson 2004:270). Secondly, it is compatible with Sher's view that an agent is responsible for traits

[37] Battaly considers attributability responsibility for epistemic wrongs to be a form of blame. Discussing the example of Harris and the Hitler Jugend, she argues that '…if attributability responsibility is viable, they will be blameworthy for their vice's nonetheless' (2016a:114).

that 'reflect badly on her' whether or not these traits express her 'real self' (Sher 2006:57). And thirdly, it is compatible with Fricker's view that an agent is blameworthy for bad traits that have their source either in the agent's epistemic character or epistemic system (Fricker 2016:41). All of these are forms of attributability responsibility are suited to personalism, in so far as they are non-voluntary (Battaly 2018a:124-125).

Battaly (2016a:113) outlines three overarching conditions that an agent must meet to be considered responsible for a trait within the framework of an attributability-based understanding of responsibility:

I. The trait must be a personal quality, expressing the subject's 'real self'; i.e., her evaluative judgements and corresponding motivations.
II. The subject must be generally responsive to reasons.[38]
III. The subject must have the capacity to recognize the trait as their own and be able to evaluate it.

Returning to the earlier machine transformation case, Battaly argues that if we assume Nozick's machine can produce full-blown personal qualities, the attributability condition for vice can be met. Condition one is fulfilled since the trait is a personal quality of ours. Condition two is satisfied when we exhibit responsiveness and possess the capability to act freely, and condition three is met as we are able to recognise and assess our traits (Battaly 2016a:114). What follows from this is that we would be praiseworthy for our virtue if the machine-produced open-mindedness, and blameworthy for our vices if the machine-produced dogmatism.

Regardless of whether personalism adopts this particular type of responsibility or the less stringent ones suggested by Fricker (2007, 2016) and Sher (2006, 2009), what is important to Battaly is that personalism permits the assignment of responsibility to individuals for their virtues and vices. However, this act of assigning responsibility is not an inherent component of what defines a virtue or vice. In other words, the definition of vice does not intrinsically involve the requirement that one holds this or any, form of responsibility. This enables Battaly to address the objection that personalism exonerates individuals too readily, as well as the

[38] Battaly understands reason-responsiveness in a similar vein to Pereboom (2014:136), Sher (2006:58) and Smith (2008:383,388).

earlier argument suggesting that traits cannot be genuinely considered virtues or vices if individuals cannot be blamed or praised for them.

Whilst Battaly still contends that blame does not *need* to be a component of vice, from this response, it is clear that she does now consider it to be an appropriate response in some instances. This is a stark contrast to her previous position, and one that I argue undermines her rejection of the responsibilist claim that blame is an integral feature of vice. Let's assess this form of responsibility that Battaly is now open to and how it impacts some of her prior arguments.[39]

One concern with Battaly's adoption of attributability responsibility is the ambiguity in attributing a vice to one's real self within the framework of personalism. As we have seen, according to condition one of the above attributability framework, a trait reflects the agent's real self via their motives and values. These are reflective enough of the individual to attribute a vice and therefore hold the agent responsible.

However, it cannot be the case that bad motives alone equate to blame under Battaly's view, as we saw individuals such as Harris possess bad motivations and not be deemed blameworthy. Despite these traits reflecting some features of Harris and being personal to him, Battaly argues that these bad motivations do not warrant blaming Harris for his vice of cruelty. Consequently, if Battaly intends to justify her previous claim then vice-bearers must be responsible for the badness of their vice in some other sense. It is unclear in Battaly's account what this could be.

Should Battaly now opt to argue that Harris and other similar vice-bearers are blameworthy, albeit in the attributability sense, she risks contradicting her earlier position.

Here we can return to Battaly's second reason for objecting to holding vice-bearers like Harris, responsible for their vices. Battaly argued that blame is not always a suitable response to vices, indicating that it cannot be an intrinsic feature. Crucially, it is often unsuitable when vice-bearers have lacked sufficient control over the formation of their vices, due to a challenging upbringing or unfortunate circumstances (2016a:107-108, 2018a:121).

[39] More recently, Battaly raised an objection to accounts of vice that advocate for attributability responsibility, including her own (2019). See Chapter 4 of this thesis for a discussion of this argument.

However, this argument appears to be redundant if Battaly is now open to acknowledging that vice-bearers can be held responsible for their vices despite the absence of control over their formation.[40] On one hand, Battaly argues that if we lack control over our epistemic vices, then we should not be blamed for them, refuting the responsibilist claim. However, Battaly now considers attributability responsibility, a form of blame which does not require control, a potentially suitable reaction to vice. Whilst this does not mean that the responsibilist claim is therefore true, it does undermine one of Battaly's reasons for primarily objecting to it.

This same objection also applies to Battaly's wider advocacy for personalism, as it is her rejection of the responsibilist claim that spurs motivationalism. By her own admission, a form of blame that is suitably apt for the likes of Harris, is possible, if we are willing to consider a non-voluntary form of blame (2018a:125). What, then, propels her argument that blame is not a necessary feature of vice? If a form of blame is suitable for instances where individuals lack control, it is perfectly possible that vices can be blameworthy in instances where control over their possession was possible (voluntary accounts of blame) or when agents had no control (non-voluntary accounts of blame).

It seems then, by acknowledging and endorsing attributability responsibility, a form of responsibility that does not require control, Battaly weakens her initial claim that responsibility need not be an inherent feature of vice, given the lack of control one usually or sometimes has over its formation. This, coupled with the prior objection to Battaly's first claim - that a motivational explanation need not substitute the responsibility component of vice - creates problems for her fundamental proposition that individuals need not be responsible for their vices.

Having raised this final objection, let us briefly recap personalism and the concerns I have expressed with it. Borrowing features from both virtue reliabilism and responsibilism, Battaly presented an account of vice with two core claims. One, that vices are personal qualities that are instantiated via the vice-bearers' motivations, and two, that vice-bearers are not necessarily responsible for their vices.

[40] Tanesini raises a similar objection aimed at Battaly, contending that Battaly does hold Harris responsible for his vice of cruelty (2021:178). She also argues that Battaly confuses attributability with answerability and that Battaly believes we might be responsible for intellectual vices in Tanesini's own understanding of attributability (2021:179).

I raised concerns over the scope of personalism, given that it seemed limited to instances of vices formed via indoctrination. If Battaly wishes to extend personalism to other varieties of vice, she would have to commit to the view that we are not in control or responsible for the majority of our vices, a claim she provided limited support for.

As we saw, Battaly also argued that motivations can explain the viciousness of vices, and responsibility need not play a part. I argued that these two features of vice were incompatible with one another, as the responsibility condition on vice is a reaction, not an explanation, of the badness of vices. I also discussed how some vices could not be explained by the personalist framework and that it had potentially limited scope (Cassam 2019a; Crerar 2018).

Finally, I examined Battaly's argument for favouring attributability responsibility for vices. I demonstrated that this undermined her previous arguments for personalism, arguing that it was unclear how one's values and motives can truly reflect one's deep self in the sense needed for attributability responsibility. One option was that the motivations possessed by vice-bearers are intrinsically bad, and this is what the blame attaches to. However, I noted how this cannot be the case seeing as Harris possessed bad motives and was not blameworthy. Another contender was that Battaly revises her view and argues that Harris and other similar vice-bearers, were now blameworthy in the attributability sense. However, this would undermine her argument that responsibility is not a necessary feature of vice, as it is not fair to blame people who lack control for their vice formation.

3.6 Tanesini's Motivationalism

In continuing our evaluation of motivationalism as a theoretical framework for epistemic vice, our focus now turns to another prominent, motivationalist account of epistemic vice offered by Tanesini (2018, 2021).

3.6.1 Sensibilities, Thinking Styles and Character Traits

Tanesini argues that epistemic vices '…involve non-instrumental motives to oppose, antagonize, or actively avoid things that are epistemically good in themselves' (Tanesini

2018:350). In agreement with Cassam (2016, 2019a), Tanesini understands vices as heterogeneous, meaning they can encompass a variety of kinds. However, three species are particularly prominent: sensibilities, thinking styles and character traits. All of these species reflect aspects of one's intellectual character.

Sensibilities are 'dispositions to use one's perceptual capacities in distinctive ways in the service of epistemic activities' (Tanesini 2021:27). Virtuous sensibilities are, in part, comprised of complex tendencies to have strong feelings about certain aspects of our surroundings, which then makes those aspects stand out as important. Among these virtues is the virtue of being observant (Hookway 2003). Alternatively, vicious sensibilities involve a form of insensibility to what matters. Vices of this nature include testimonial injustice (Fricker 2007) and wilful ignorance (Tuana 2006). These are also sensibilities that are neither virtuous nor vicious.

Motivation is also crucial in shaping intellectual sensibilities according to Tanesini. With wilful ignorance for example, some people are motivated to divert their attention away from their racial privilege and not notice their discriminatory behaviours (Tanesini 2021:28). Tanesini remarks that the motivations that lead to the cultivation of wilful ignorance such as this are often hidden to the agents who develop this skill. Such individuals are also prone to wishful thinking or self-deception, as their desire not to know the facts can result in them either refraining from forming any opinions or forming false beliefs (ibid.:29,45).

What follows then, is Tanesini's following criteria: A sensibility is epistemically virtuous 'only if (a) it is a skill that generally promotes the achievement of the subject's domain-specific epistemic goals, and (b) it is developed as a result of a general motivation to acquire epistemic goods.' (ibid.:29). It is epistemically vicious only if '(a) it systematically frustrates the achievement of some of the subject's domain-specific epistemic goals, and (b) it is developed as a result of a motivation to turn away from epistemic goods' (ibid.).

Turning now to thinking styles, these are characterised as dispositions towards adopting specific thought processes and favouring them over other options (ibid.). Virtuous thinking styles include the tendency to find pleasure in thinking or being open to new ideas (Kahneman 2012; Mercier and Sperber 2017), Vicious thinking styles include prejudice or a tendency to relentlessly pursue certainty. Motivations also play an integral part in virtuous and vicious thinking styles. Agents may be motivated by a love of learning or knowledge or an

overwhelming desire for cognitive closure. Again, we get the following criteria: 'Thinking styles are virtuous only if (a) they are driven by motivations that are epistemically good and (b) generally promote the agent's epistemic goals. They are vicious, rather than mixed, only if (a) they are driven by motivations to turn away from what is epistemically good, and (b) typically result in the frustration of at least some of the agents' epistemic goals (Tanesini 2021:31).[41]

Finally, we have character traits, defined as dispositions to favour certain methods of engaging in intellectual activities over other approaches. Virtuous examples include open-mindedness, epistemic humility, and courage. Vicious examples include closedmindedness, epistemic arrogance, and cowardice. Motivations play a key role in character traits too. For example, open-mindedness is driven by a motivation to explore alternative and novel ways of thinking, and intellectual arrogance involves an attitude driven by the need to preserve a high opinion of oneself (ibid.:32).

We arrive at the final condition. Epistemic character traits are virtuous if they '(a) are driven by motivations that are epistemically good because directed at what is intrinsically epistemically good, (b) systematically facilitate the agent's setting of epistemic goals that are commensurate to her abilities and that promote the attainment of epistemic goods, and (c) typically foster the achievement of the agent's epistemic goals' (ibid.:34). Character traits are vicious if they '(a) are driven by motivations that are intrinsically bad because they involve turning away from what is epistemically good; (b) they also generally frustrate the agent's setting of epistemic goals that are commensurate to her abilities and promote the attainment of epistemic goods, and (c) typically hinder the achievement of the agent's epistemic goals' (ibid.:34-35).

Having gained insight into the types of things Tanesini categorises as epistemic vice, let's now turn to discuss the motivational element in more detail.

3.6.2 The Motivational Component

[41] By 'mixed' Tanesini refers to a trait that contains both virtuous and vicious values (2021:26).

As we have seen, for Tanesini, motivation is an essential component of epistemic virtue and vice. Crucially, she considers motivation to have an important role in the explanation of action and belief formation.

Tanesini appeals to three types of explanations used in the philosophy of action: justifications, rationalizations, and mere explanations (ibid.:42). Actions are justified when they are supported by normative reasons. They are rationalised by presenting the reasons that the individual believes supported their action, and they can be explained by referring to the psychological state that made the action. Turning to motivations, Tanesini notes that an open-minded individual would not rationalize her epistemic conduct in terms of her open-mindedness (this would appear arrogant). Instead, open-mindedness is the driving force that has influenced her behaviour and motivated her to seek our reasons and explore viewpoints other than her own. Importantly then, epistemic virtues and vices are best thought of as the deep roots of epistemic conduct rather than as the *conscious* reasons used by agents to rationalize their views and conduct.

According to Tanesini, current views in vice epistemology focus on the wrong kind of motivations, the motivations of the former sort, where they are conscious and therefore available to agents to rationalize their behaviour (Cassam 2016, 2019a; Crerar 2017). When an action or belief stems from bad motives, an acknowledgement of these bad motives and an attempt to rationalise them is contradictory to the bad motives. For example, a truly arrogant individual would not be aware of their arrogant motives nor be able to explain their behaviour in these terms. Tanesini appeals to Cassam's (2015) understanding of 'stealthy vices' to support this view. Individuals rarely 'discover' their epistemic vices as they often attempt to rationalize their harmful behaviours as acceptable and are therefore prone to self-deception (Tanesini 2021:45).

Finally, motivations in part make up 'attitudes', specifically the functional component of attitudes that act as the psychological basis for virtues and vices (ibid.:66). Tanesini considers attitudes as akin to likes and dislikes, and they have a causal role in shaping the behaviour that exemplifies virtues and vices.

To use Tanesini's example, if an individual aims to boost their self-esteem, they might be driven to protect their ego from anything that could harm their positive self-image.

Consequently, their evaluation of situations may be shaped by the criterion of whether they pose a threat to or bolster their self-worth. Their attitudes would then be formed primarily by information tied to these self-esteem considerations. Additionally, due to the significant risks associated with confusing something innocuous for a potential threat, the individual's assessments tend to be defensive and hypervigilant. These judgements are condensed into their attitudes, which now primarily serve to defend the individual's ego. Therefore, the function of attitudes is strongly connected to the predisposition to have certain underlying motives. These motives are often not conscious to the individual (2021:66).

To summarise, Tanesini argues that the negative quality of vices stems from their bad epistemic motivations. An individual can be motivated away from epistemic goods or motivated towards epistemic wrongs. A side-effect of this is that vicious individuals are also self-deceptive, which further explains their badness. Motives are central to the psychological framework of vice in so far as they define the function of attitudes.

3.7 Motivations and Responsibility

Tanesini details the various harms that motivational epistemic vices cause. These range from harms to the vice-bearer, including harms to self-knowledge or self-trust via deception, to harms to others, including denying others the credit they deserve via displays of arrogance (ibid.:154,159). Tanesini assesses who can be held responsible for these harms and wrongs. Setting aside issues of control and responsibility, she focused on 'responsibility responses', defined as reactive attitudes including praise and blame (Tanesini 2021:170). Following Shoemaker (2015) Tanesini surveys three forms of responsibility: attributability, answerability, and accountability.

For Shoemaker, agents are attributable-responsible for features of their character when they reflect the agent's 'deep self' (Shoemaker 2015:38). This is understood as encompassing the psychological attributes that reflect the person's most fundamental values and commitments. On this account, an agent is responsible for their deep self and the behaviour/beliefs that stem from this. In other words, individuals are responsible for what can be properly attributed to

them.[42] Tanesini remarks that the 'responsibility responses' (also referred to as reactive attitudes) that are characteristic of attributability responsibility include admiration and esteem or disesteem and disdain. Positive emotions amount to praise, and negative emotions amount to blame, despite not invoking punishment or resentment (2021:172).

Answerability responsibility holds people responsible for their actions and beliefs based on the quality of their choices, beliefs, and judgements (Shoemaker 2015:72-73). The responsibility-responses characteristic of answerability includes approval or pride e.g., we approve the quality of someone's judgement and disapproval or regret e.g., we disapprove of others when they display poor judgement. Again, these positive and negative emotions constitute ways of praising and blaming, meaning answerability responses are responsibility responses (ibid.).[43]

Finally, individuals can be held responsible in the accountability sense when they have the capacity for an empathetic understanding of how situations appear and different viewpoints (Shoemaker 2015:88).[44] The responsibility-responses characteristic of accountability includes gratitude e.g., we are grateful to those who are kind to us, or anger and resentment e.g., we are angry with people whose actions demonstrate a disregard for our beliefs. Like the previous two accounts, these responses amount to a form of blame or praise.

Tanesini dismisses answerability as the type of responsibility fitting for our epistemic virtues and vices. She contends that individuals are not wholly answerable for their intellectual vices because answerability responsibility requires individuals to be able to provide reasons for their beliefs and choices that they can justify and assess. However, Tanesini states that this ability presupposes that the individual can have an awareness of the alternative viewpoints and actions, yet these are the very abilities that are often impaired in those who possess the intellectual vices of self-evaluation (2021:174-175). For example, a closed-minded individual is unable to give fair weight to alternative views.

Attributability responsibility has more promise, according to Tanesini, in elucidating how vice-bearers can be held responsible for their vices. She argues that individuals are attributability responsible in a moral and epistemic sense for their virtues and vices in so far as they are among

[42] For further views informed by the 'deep/real self' see Frankfurt (1971); Taylor (1976); Watson (1996) and Wolf (1990).
[43] Hieronymi (2008,2014); Scanlon (1998,2008) and Smith (2015) use 'answerability' to refer to a view more like the attributionist perspective (Talbert 2019).
[44] Others define accountability responsibility more broadly, meaning individuals are responsible when it is appropriate for others to hold them to a particular standard. Doing so entails appropriate sanctions such as praise or blame (Levy 2005; Watson 2004; Zheng 2016).

the components of people's character or deep self – they are attributable to one's character. For example, an arrogant individual is driven by the desire to feel good about themselves. This motivation for self-enhancement reflects one of their deepest cares and is therefore a reflection of their 'deep self' or character.

Finally, Tanesini argues that vice-bearers can sometimes be accountable for their vice, but not in an epistemic sense. This distinction arises from the assertion that the reactions constitutive of accountability e.g., resentment or anger, are not suitable for addressing epistemic failings or harms. Tanesini defines epistemic blame as something that '…attaches to beliefs and forms of inquiry where the inquirer is at fault.' (ibid.:173). However, she contends that this form of blame is ill-suited to accountability, as it lacks the '…reactive attitudes that are characteristically accountability responses' (ibid.). Consequently, Tanesini 'strongly suggests that epistemic responsibility is not a matter of accountability' (ibid.)

Shifting the focus specifically to the moral dimension of responsibility, Tanesini observes that the criteria for attributability and answerability responsibility are the same in both the epistemic and moral dimensions (2021:179). This means that agents are both epistemically and morally responsible for their epistemic vices in the attributability sense, but not fully answerable, as the previous arguments also apply in the moral domain.

Despite its unsuitability to epistemic harms, accountability responsibility may be a sufficient response to the moral harms that stem from epistemic vices. (ibid.:173). For example, we might hold someone accountable for their ignorance not because it resulted in a lack of knowledge, but because of their disregard for others or rudeness. Similarly, we might blame a timid individual as their silence meant they complied with an unethical decision.

However, on a practical note, Tanesini observes that it may rarely be useful to blame vice-bearers in this way. Let us now turn to assess these reasons.

3.7.1 Blame

To summarise, through her assessment of attributability, answerability, and accountability responsibility, Tanesini argues that epistemic vices are attributable to individuals in so far as they reflect the individual's character or deep self. This opens vices up to the appropriate attributability responses, such as esteem and disesteem. Individuals are not answerable for their epistemic vices because agents do not reflectively endorse their vices. Finally, with regards to accountability, as this is a predominately moral responsibility-response (e.g., anger or resentment) it is not fitting to epistemic harms and wrongs.

When our epistemic vices have moral consequences that need addressing, accountability responsibility may be a viable response. However, it is seldom beneficial to actively blame agents in this moral accountability way, as there are numerous prudential and moral reasons to refrain from doing so (ibid.:182)

The first such reason is that labelling someone as vicious might only serve to enhance their vicious behaviour or act as a self-fulfilling prophecy (Alfano 2013:88-96). Tanesini notes that if an individual comprehends their assigned label and finds it plausible, it is highly likely that their future behaviour becomes more compatible with the label. For example, labelling someone as arrogant may only serve to make them act more arrogantly.

Secondly, it is challenging to know whether a vice-bearer genuinely possesses the alleged vice, and if their behaviour qualifies as an instance of vice as opposed to an out-of-character action. Tanesini observes that we might lack the required evidence to properly attribute vices and therefore accurately blame vice-bearers, particularly as it is difficult to know other's motives. Levying an ill-informed vice-charge also runs the risk of generating negative consequences, as it may create resentment if one is blamed undeservingly.

In addition to these pragmatic considerations against blaming vice-bearers, Tanesini argues that many individuals lack the moral standing required to hold blameworthy vice-bearers accountable. Those who are equally as vicious as the other vice-bearer would be hypocrites, and those not in possession of the same vice must recognise their privilege and fortune for not doing so. If the individual casting blame would also exhibit the same vice under the same set of circumstances, then they cannot be in a position to blame (Tanesini 2021:183).[45] In these

[45]Tanesini argues that one of the most effective ways to tackle epistemic vices is at a societal level rather than a personal one (2021:185). This goes beyond the practices of holding individuals responsible or blameworthy and involves targeting

instances, we should observe that we are merely luckier or more privileged than the vice-bearer, meaning we should be more forgiving and charitable.

Despite these concerns motivating Tanesini's hesitancy to morally blame vice-bearers, she does consider that 'accountability responses are not the only kind of blaming attitudes one can adopt' (ibid.). As attributability responsibility is applicable to the moral domain, we still have the responses associated with this to utilize, such as esteem or disesteem. These are also forms of blame and we can manifest them by distancing ourselves from the vice-bearer and encouraging others to do so too.

Alongside these attributability responses, we can also 'take responsibility' for our own epistemic vices. Taking responsibility involves reflecting on our character and traits and making attempts to change or strengthen them, such as acknowledging one's own servility and attempting to stand up for yourself more (ibid.:185). This is a form of responsibility that is forward-looking, meaning it focuses on the individual's own shortcomings and what can be done about them. This practice requires an understanding of what is possible, deciding what actions you need to take, making yourself answerable to your choices and a commitment to follow through with your decision to take action (Card 1996:28).[46] We can therefore acknowledge that our vices may be partly the result of unfortunate circumstances but still take steps to modify our behaviour. For example, a viciously timid individual might reflect on their character and acquire the motivation to change it, gaining self-esteem.

The prudential and moral concerns associated with holding vice-bears accountable for their vices suggest that it is rarely useful to blame individuals in this manner. Epistemically, it does not make sense to speak of accountability blame. Morally, blame may seem fitting, but is rarely useful. Individuals may develop further harmful behaviour as a result of having their vice labelled, possess a lack of knowledge and accuracy over who possesses vices, and a lack of moral standing to blame others for vices that individuals may also possess or not possess due to privileged circumstances.

oppressive and dominant structures. I discuss an objection to this in Meehan (2023). Nevertheless, Tanesini maintains that individual responsibility remains an effective approach to overcoming our epistemic vices.

[46] This is similar to Cassam's understanding that vice-bearers can be responsible for their vices in a revisional sense if they are able to exercise managerial control over their actions (2019a:124).

These concerns strengthen Tanesini's position that attributability responsibility is the best-suited form of responsibility for epistemic vices. Blaming vice-bearers in this sense, both epistemically and morally via the associated responsibility responses, is therefore a suitable practice.

Finally, alongside the appropriate attributability responses, we can also 'take responsibility' for our own vices by reflecting on and bettering one's character (Tanesini 2021:185).

We can now turn to discuss some objections to Tanesini's account of vice, primarily focusing on the responsibility condition. I will express concerns over Tanesini's conceptualization of epistemic blame, her prudential reasons for dismissing the moral accountability for vice, and her argument for 'taking responsibility' for our own vices.

3.7.2 Epistemic Blame

As we have seen, Tanesini holds that the accountability responses of anger, resentment and punishment are restricted to the moral domain, and thereby not suitable to address epistemic harms and wrongs. If and when these attitudes are employed, it is likely that we are reacting to some moral harm that the epistemic vice caused, and we use these accountability responses to address that moral aspect. Consequently. Tanesini contends that it does not make sense to speak of 'epistemic blame' in the accountability understanding of blame (ibid.:173).

An immediate concern here is that it is not entirely clear why some accountability responses cannot be considered appropriate responses to distinct epistemic harms. Consider a well-known example of a blameworthy belief discussed by Jessica Brown (2020:390) in her exploration of epistemic blame.[47] The example is as follows: Maud possesses a reliable clairvoyant power and uses it to form the belief that the President is in New York. However, she also possesses strong evidence that the President is not in New York, after seeing live footage of the President in Washington on the TV. Furthermore, Maud has no evidence to believe her clairvoyant power exists, and plenty of evidence to be sceptical of it. Maud dismisses the TV broadcast and the

[47] This example was first presented by Bonjour (1980:61).

evidence against her clairvoyance. She thereby dogmatically believes the President is in New York.

In response to Maud's bad belief, Brown considers her to be epistemic blameworthy, with the blame directed specifically at Maud's bad belief. However, Brown also details how Maud's friends may rightly express this blame through reactions such as anger or rebuke (Brown 2020:391).[48] Additionally, Brown references similar scenarios where one may respond to a blameworthy belief with resentment. For instance, if a close friend believes that I lied to them I may harbour resentment towards them for believing so. My resentment is attached to the epistemically bad belief, just as it is with Maud's dogmatic belief.

There are other instances where anger seems like a plausible response to epistemic vices. For example, consider a colleague who consistently withholds information from you, causing harm to your projects and hindering your ability to make informed decisions. This colleague can be said to hold the vice of dishonesty or epistemic insouciance. It seems like a fitting response to blame your colleague and respond with anger. This anger is also specifically directed at the epistemic vice and its resulting epistemic harms.

These examples cast doubt on Tanesini's claim that some accountability responsibility-responses are rarely appropriate reactions to epistemic wrongs. Anger or resentment may, in fact, be suitable responses to the epistemic harms associated with epistemic vice.[49]

Furthermore, given this critique reading the appropriateness of accountability responses to epistemic harms, it is surprising that Tanesini does not extensively discuss the suitability of attributability responses to such harms. We can recall that Tanesini advocated for the attributability responses of disdain, disesteem, and revulsion as the most suitable reactions to blameworthy epistemic vices. However, it is not entirely clear in what sense these attitudes are suitable reactions to epistemic harms or how they align with Tanesini's definition of epistemic blame as something that '…attaches to beliefs and forms of inquiry where the inquirer is at

[48] My point here is to acknowledge that anger is a fitting response to epistemic failings. See Chapter 4 for a discussion on what makes responses such as these distinctively epistemic.
[49] Determining whether individuals expressing blame for epistemic vices experience anger would require empirical research beyond the scope of this thesis and Tanesini's work. The debate here is whether accountability responses such as anger are suitable responses. Drawing on these examples, I argue that they indeed are.

fault.' (Tanesini 2021:173). This is particularly concerning given Tanesini's rejection of accountability responsibility due to its incompatibility with epistemic harms.

3.7.3. Blaming Ourselves and Others

Shifting the focus now to moral responsibility, we can recall that even when moral accountability seems fitting for the moral harms resulting from epistemic vice, Tanesini contends that it is rarely apt to blame vice-bearers in this manner. She cites various prudential and moral reasons to avoid doing so. Firstly, attaching labels to vices can become a self-fulling prophecy, resulting in further epistemic harm. Secondly, the lack of evidence to know if someone truly possesses the blameworthy vice poses a challenge. And thirdly, we often lack the appropriate moral standing to blame due to hypocrisy or privilege. I will now assess each of these three claims, arguing that these concerns also pose challenges for Tanesini's attributability responsibility.

We can respond to Tanesini's first concern regarding vice-labelling in two ways. Firstly, from Tanesini's understanding of accountability, it remains unclear how the various responsibility-responses of anger, punishment or resentment amount to mere vice-labelling. Whilst it seems true that merely labelling vices is an unproductive and potentially harmful practice, it is not clear how these accountability responses amount to this. For example, responding to someone's vice with anger or resentment does not necessarily amount to mere vice-labelling. This type of blame may even be non-verbal if it is gestural or behavioural e.g., I cannot stand to be around you, so I leave the room (Fricker 2016:171).

Secondly, and potentially more concerningly, is that this problem does not appear to be exclusive to accountability responsibility. How do attributability responses such as disesteem, disdain or revulsion avoid becoming mere vice labels?

Whilst this objection makes the important point – simply labelling vices is not an effective way of assigning blame – this concern appears to apply broadly to any superficial form of blame, categorised as accountability, attributability or beyond. Therefore, as long as our practices of assigning blame are intentional and do not devolve into mere name-calling, we can avoid this particular worry.

We can turn now to address Tanesini's second concern, that we often lack the necessary knowledge to accurately attribute a vice and, consequently, assign blame. Once again, however, this concern does not seem isolated to moral accountability and appears to undermine Tanesini's own advocacy for attributability responsibility.

If we encounter difficulties in identifying vices in others with the purpose of assigning blame, a similar objection arguably arises when seeking to determine whether traits are a reflection of one's 'deep-self', as required for attributability responsibility. If we cannot determine this, it also appears that we cannot attribute vices to an individual's epistemic character.

One way to avoid this problem for attributability responsibility is to argue that individuals do not need to know that they are blaming others for an epistemic vice per se, but just the bad trait. In other words, individuals may blame others for their persistent displays of arrogance or closedmindedness, but without needing to recognise these traits as epistemic vices. This avoids the above concern as we would not need to accurately attribute a vice to the wrongdoer, but just ensure that they are truly arrogant, or truly closed-minded. This shifts the required knowledge for blame from identifying the behaviour as a vice to acknowledging that the behaviour is genuinely present.[50]

Despite this potential response, the crucial observation here is that again, Tanesini's concern is not isolated to accountability blame and creates potential concerns for attributability blame too.

Moving onto Tanesini's third objection, there is a concern that we often lack the appropriate moral standing to blame others for their epistemic vices and their resulting harms. This lack of moral standing could stem from an individual sharing the same vice that they are accusing others of, or by benefitting from privilege that shields them from possessing the same vice.[51]

The first part of this concern is often referred to as the 'non-hypocrisy condition' and is a widely recognised concern for various accounts of blame (Fritz and Miller 2018, 2019; Isserow and Klein 2017; Roadevin 2018; Todd 2019; Wallace 2010). This condition states that an individual cannot blame another for something that they too are at fault for (Fritz and Miller 2019). I

[50] See Kelp (2019) for an argument on the knowledge norm of blaming.
[51] See Bell (2012) for a critique of the view that blame is only appropriate when the blamer has standing to blame.

concur with Tanesini that in such instances it would be morally inappropriate to blame the vice-bearer.

However, Tanesini's second reason that we lack moral standing to blame is less convincing. She argues that individuals who, owing to good fortune or privilege, do not possess a particular vice, are not justified to cast blame. For example, I should not blame my uneducated friend for their closed-minded beliefs if, given the same circumstances and limited educational opportunities, I too would have developed this vice. The moral inappropriateness of such instances arises from the unfairness of blaming or the potential for other negative consequences (Tanesini 2021:183).[52] Even though I recognise the closedmindedness in my friend, I should also recognise my privilege and not blame them for their vice as doing so would be unfair.

This places a strong condition on our blaming practises, given its potentially wide applicability. Many vices are at least partially the result of unfortunate circumstances that others have been fortunate enough to avoid.

The first point to acknowledge here is that blame need not be inherently harmful or unfair. For example, accounts of blame that are sensitive to these aetiological concerns may be suitable, such as one that takes the primary aim of blame to be communicative. From this approach, blame consists of letting the wrongdoer know that they have wronged you with the hope of making them perceive or acknowledge their wrong. (Fricker 2016:171-173).[53] If blaming primarily consists of communicating the wrong to the wrong-doer, then we can still condemn our friend's closed-minded beliefs and even request they change their behaviour. However, this blame does not inherently come with the harmful connotations of being unfair, high-minded or morally inappropriate. To perceive of blame in such a way speaks to the 'bad reputation' that blame has acquired, which ignores its nuances (Fricker 2016:169).[54]

A second point for consideration is that again, this concern is equally as applicable to attributability responsibility and not just our moral accountability responses. If the responses

[52] In Chapter 4 I discuss how the unfairness to blame in instances such as these is due to a lack of control over the formation of vice.
[53] Fricker contends that aiming to get the wrongdoer to acknowledge their wrong may not be a motive that is known to the blamer (2016:173).
[54] See Chapter 4 for a more detailed discussion of the ameliorative role of blame.

of anger, resentment and punishment are deemed morally inappropriate, it seems likely that as will Tanesini's detailed attributability responses of disesteem, disdain or revulsion.

However, Tanesini believes attributability responsibility is not subject to these same concerns. She considers disesteem to be a morally appropriate response as it is not subject to the same hypocrisy charges, making it acceptable to express disesteem for a vice I also possess. This is because, unlike accountability responsibility, disesteem does not call for a response from the vice-bearer e.g., an apology (2021:184). This means the moral standing of the blamer is less important.

Not all accountability responses demand a response, however. For example, anger may have the purpose of allowing the wronged to express emotions and vent without expecting anything in return from the wrongdoer (Cogley 2013). If attributability responses escape the hypocrisy charge on this condition, accountability responses also seem to as well.

What these concerns on moral standing demonstrate is that we should be sensitive in our blaming practices. This is not a criticism that can be solely directed towards moral accountability, however. Any form of blame that is insensitive to the wider context or fails the non-hypocrisy condition is subject to this criticism, and this can equally include certain attributability responses of disesteem, disdain, or revulsion too.

We can conclude this section with a brief remark on Tanesini's advocacy for 'taking responsibility', an alternative form of responsibility for epistemic vices. This involves understanding your own behaviours, recognizing your bad tendencies, and evaluating your epistemic character (Tanesini 2021:185-86). By recognising these shortcomings and attempting to change them, individuals assume responsibility for their own character. In this sense, 'taking responsibility' is a form of self-blame.

Tanesini herself raises a potential issue with this approach, however. The concern is that this practice is only effective when the individual's vices are not 'stealthy' (2021:190). Stealthy vices are ones that cannot be observed or therefore overcome via self-reflection (Cassam 2015) It is therefore unrealistic to expect individuals to take responsibility for vices that they are unaware of and consequently cannot modify.

However, despite this concern, Tanesini considers this to be an important form of responsibility, one that allows vice-bearers to acquire the 'measure of oneself', developing self-respect and self-esteem (Tanesini 2021:186). [55]

It is questionable just how feasible this practice is, however. We can recall Tanesini's claim that the motivations constitutive of epistemic vices are typically concealed from the vice bearer. If this is true, it appears that most vices, according to her account, will be stealthy (ibid.:44-45) Consequently, it appears that only a limited number of vice-bearers will be able to take responsibility for their vices, rendering this method nearly redundant. If anything, the prevalence of stealthy vices only underscores the importance of interpersonal responsibility methods. With stealthy vices, we need to rely on others to point out our vices, as recognising them in ourselves may prove exceedingly challenging if they are undetectable to us.

Let us now reassess Tanesini's stance on the moral dimension of blameworthy epistemic vices. As we have seen, Tanesini rejected the suitability of moral accountability as a response to epistemic vices, given the various prudential and moral reasons discussed above. To avoid implicitly legitimatising vicious behaviour, Tanesini argues that other attributability responses such as esteem or disesteem are still fitting ways to blame the moral dimension of vice. Additionally, we can 'take responsibility' for our own vices and their harms, by reflecting on our character and making attempts to better it.

Given my above responses to the prudential and moral concerns, it appears that attributability responses are equally susceptible to some of the objections raised by Tanesini. I have argued that the real target of her criticisms appears to be any hypocritical, insensitive or superficial account of blame, whether it be a form of accountability or attributability responsibility. Tanesini's objections to moral accountability are therefore too general, and risk ruling out cases where blame of this form may be appropriate. I also argued that the existence of 'stealthy vices' undermines the effacing of 'taking responsibility' as a responsibility method for vice. We cannot hold ourselves responsible for vices that elude our detection, making it impossible to reflect or improve upon them.

[55] Tanesini holds that other interventions that do not fall under the bracket of 'taking responsibility' may be applicable for stealthy vices too, such as self-affirmation techniques (2021:193).

These concerns, alongside my previous objections regarding the epistemic dimension of both accountability and attributability responsibility, suggest that blame may be a justified response to a broader range of epistemic vices and contexts than Tanesini originally acknowledged. For instance, if anger proves to be a suitable reaction to epistemic failings, or we acknowledge that blame is still justified in aetiologically sensitive contexts, blame seems like a feasible option in cases that were previously excluded. This suggests the need for a more comprehensive understanding of the ameliorative purpose of blaming epistemic vices, which is the topic of my next chapter.

3.8 Conclusion

The focus of this chapter has been to critically examine two prominent motivationalist accounts of epistemic vice: personalism and motivationalism. Beginning with personalism, I outlined the responsibilist and reliabilist features of Battaly's account, focusing particularly on the claims that vices are personal qualities, stemming from one's personal motives and values and that vices are not necessarily blameworthy.

I then raised concerns over the scope of personalism, given that it seemed limited to instances of vices formed via indoctrination. I also discussed concerns that the motivational component of Battaly's account was underdeveloped or irrelevant for some instances of vice such as stupidity (Cassam 2019a; Crerar 2018).

Turning my focus predominantly to the personalist stance on responsibility and blame, I examined Battaly's argument in support of attributability responsibility. I argued that this undermined her previous arguments for personalism, as it raised doubts about how a vice-bearer's values and motives can genuinely reflect one's deep self in the sense needed for attributability responsibility. If the motivations possessed by vice-bearers are intrinsically bad, and this is what the blame attaches to, Battaly would be contradicting her earlier claim that Harris and other indoctrinated vice-bearers were not blameworthy for their vices. However, if Battaly updated her view and argued that these vice-bearers were now blameworthy in the attributability sense, this would contradict her initial position that responsibility is not an obligatory feature of vice, as it is not fair to blame people who lack control for their vice formation.

After raising these objections, I proceeded to examine Tanesini's account of vice, referred to as motivationalism. Tanesini argued that vices consist of self-deceptive epistemic motives towards epistemic bads or away from epistemic goods. Additionally, vice-bearers can be attributably responsible for their vices, as vices are part of one's character and 'deep self'. Despite this, Tanesini also claimed that it is rarely appropriate to blame individuals for their vices. This is because blaming and therefore labelling someone's vice, might only serve as a self-fulfilling prophecy. This, alongside, showing resentment and other negative reactions towards a vice bearer might also be counter-productive to the practice of blame. Additionally, many of us lack the moral standing required to blame others without being considered hypocrites. Tanesini also argued that we often lack the confidence to know if someone truly possesses the vice that we believe them to have in order to blame them.

Moving onto my criticisms of Tanesini's account, I focused on how accountability responses may sometimes be suitable to epistemic contexts, particularly with regard to anger. I also discussed how the epistemic dimension of attributability responsibility was lacking.

I subsequently addressed Tanesini's prudential objections to holding vice-bearers morally accountable. Regarding the first issue of vice-labelling, I argued that this worry applies to any shallow form of blame and not exclusively to moral accountability. I then addressed the second concern, that we often lack the knowledge to accurately attribute a vice and, consequently, assign blame. Once more, this concern extended beyond moral accountability and undermined Tanesini's endorsement of attributability responsibility, creating concerns for the possibility of attributing vices to an individual's 'deep self', a fundamental aspect of attributability responsibility.

The third concern was that we often lacked a moral standing to blame others for their vices, either because we share the same vice and would therefore be hypocrites, or because it would be unfair to blame others for vices that only privilege prevents us from acquiring. Referring to the first part of this concern, I discussed how Tanesini considers attributability responsibility. to be immune from this objection, as responses such as disesteem do not require anything back from the wrongdoer such as an apology. I argued that this reasoning also applied to accountability responsibility thereby letting it avoid this concern. Regarding the potentially unfair instances of blame, I argued that the core issue lies with any harmful or high-minded blaming practice, irrespective of whether they fall under moral accountability or other forms

of blame. Mitigating the concern involves adopting a blame approach that is ameliorative and fair to both the vice-bearer and the wronged party.

Finally, I argued that the form of responsibility that Tanesini appeared most optimistic towards – taking responsibility for your own vices – contradicted her earlier statement that most motivationalist vices are 'hidden' to the vice-bearer. If most vices are indeed stealthy, this suggests that we should blame others for their vices, in a way that is productive and fair.

In my evaluation of both accounts of vice, I addressed the critiques put forth by both Battaly and Tanesini regarding the appropriateness of blame as a response to epistemic vices. Overall, both accounts leaned more favourably towards blaming vice-bearers within an attributability responsibility framework. In the upcoming chapter, I will further support and advocate for this position, as well as emphasizing its distinct epistemic dimension.

CHAPTER 4. BLAMEWORTHY VICES: UNDERSTANDING THE ROLE OF BLAME FOR EPISTEMIC VICE

4.1 Introduction

As discussed in previous chapters, a debate is emerging in vice epistemology as to whether it is appropriate to hold agents responsible for their epistemic vices, and crucially whether, by their very nature, epistemic vices are such that their possessors are blameworthy for them (Battaly 2016a, 2018a, 2019; Cassam 2019a; Kidd 2016, 2020; Tanesini 2018, 2021). This thesis gains prima facie plausibility from the observation that epistemic vices cause a variety of epistemic and moral harms which are damaging to one's overall character and surroundings, and that frequently we want to hold people accountable when they fail to acquire the truth due to their gullibility or naivety, or when their closed-mindedness or arrogance leads to false beliefs. However, despite the intuitive plausibility of the constitutive blame thesis, many vice epistemologists cast doubt as to whether vices are blameworthy, and the stronger claim that they are consistently and inherently so.

This chapter explores if and how we can be held responsible for our intellectual vices and the derivate behaviour that stems from them, and what this form of responsibility looks like. What does it mean to say someone is responsible or blameworthy for an epistemic vice? What understanding of responsibility and blame best suits an account of vice? This chapter has three objectives. The first objective is to illustrate that an epistemic and ameliorative form of blame can be suitably assigned to epistemic vices. The second objective is to argue that epistemic vices are attributability responsible. The third and final objective is to argue that attributability blame is constitutive to epistemic vice.

Doubts about the appropriateness of blame for vice comes in the form of two sceptical positions. First, some vice-epistemologists deny that we have control over our vice acquisition, making blame inappropriate or unfair (Cassam 2019a; Battaly 2016a, 2018a; Kidd 2016, 2020). I refer to this position as the 'argument from lack of control'. From this, the second claim arises that blame cannot be an integral feature of epistemic vice (Battaly 2016a, 2018a; Cassam

2019a; Kidd 2016, 2020; Tanesini 2018, 2021). I refer to this as the denial of the 'constitutive blame thesis'.

The plan for this chapter is as follows. I will begin by motivating my first objective, examining four different accounts of epistemic blame and whether they are suitable responses to epistemic vice. I argue that the fourth account is promising, drawing on the ameliorative nature of blame discussed in feminist approaches to responsibility (Ciurria 2021; Fricker 2016; Hutchison 2018; Mackenzie 2018; Oshana 2016).

I then turn to my second and third objectives. I will identify the argument from lack of control, outlining the objection to the appropriateness of blame for vice and the constitutive blame thesis. I then propose two solutions. Firstly, I argue that a form of responsibility without control, attributability responsibility, explains how agents can be blameworthy for vices that they acquired in environments outside of their control. I then argue that if we accept this form of responsibility, it necessarily follows that blame is integral to the definition of epistemic vice.[56]

4.2 The Nature of Epistemic Blame

As discussed in the previous two chapters, contemporary accounts of epistemic vice have explored whether agents are blameworthy for their epistemic vices, focusing on what type of responsibility if any, is appropriate for vice and the bad behaviours that flow from them. In general, most vice-epistemologists have expressed a degree of scepticism on the appropriateness of blaming agents for their vices. These arguments have taken the form of either denying that blame is a fair response to vice (Kidd 2016, 2020; Tanesini 2016, 2021) or opting for a weaker form of responsibility or lower compatibility conditions (Battaly 2016a, 2018a; Cassam 2019a). Most accounts were open to the possibility of epistemic blame, despite being presented vaguely. For Cassam, what was epistemic about the type of responsibility (both criticism and blame) is that it is blame directed at an epistemic failing (Cassam 2019a:123). For Tanesini, epistemic blame was defined as a response attached to 'beliefs and forms of

[56] Despite only focusing on how individuals are blameworthy for their vices, I believe my argument also extends to collectives and institutions that can exhibit vices, for which they will also be blameworthy. See Chapter 7 of this thesis for a discussion on how institutions can exhibit virtues and vices.

inquiry where the inquirer is at fault' and instances 'where epistemic defects reflect badly on the agent' (Tanesini 2021:173). She spends some time detailing what epistemic is *not*, stating it does not involve the kind of reactive attitudes that we usually associate with moral responsibility i.e. resentment, anger, and punishment.

Leaving these arguments aside for now, we can turn to literature on blameworthy beliefs and epistemic norms for a more detailed understanding of the epistemic form of blame. In discussing the nature of epistemic blame, four key views have been presented: the desire-based view (Brown 2020), the relationship-based view (Boult 2020, 2021), the emotion-based view (McHugh 2012; Nottelmann 2007; Rettler 2017), and the agency-cultivation view (Piovarchy 2021).[57] Let us now examine each approach and its potential appropriateness for blame for vice.[58]

4.2.1 The Desire-Based View

The desire-based view is defended by Jessica Brown (2020) who appeals to George Sher's (2006) 'two-tiered' account of moral blame to formulate her own account of epistemic blame. Sher's account of blame claims that blame 'consists of a characteristic set of affective and behavioural dispositions that are organized around a characteristic type of 'desire-belief-pair' (Sher 2006:14-15). The relevant dispositions are those that we normally associate with blame e.g., the disposition to reproach, feel anger and apologize.

However, Sher argues that it would be wrong to view blame consisting solely as relevant dispositions, as this would fail to explain what unifies the relevant dispositions and blame. Instead, Sher suggests that what unifies these dispositions is that they consist of a particular belief-desire pair, namely the belief that someone had acted badly/has a bad character, and the desire that they had not acted in this way/have a bad character (ibid.:102-103). Blame therefore consists of our reactions to when people culpably violate norms that we had desired they not.

[57] These views are categorised by Boult (2021:4).

[58] Most accounts of epistemic blame also accept that both epistemic and moral blame may be appropriate responses to the same target (Boult 2021:2). However, as Cassam argues, it would be excessively moralistic to argue that all epistemically blameworthy conduct is also morally blameworthy (Cassam 2019a:18). With this in mind, it is reasonable to consider that if blame is an appropriate response to epistemic vices, it can be either epistemic, moral, or both. In this chapter, I will concentrate exclusively on the relatively underexplored epistemic dimension.

Brown applies Sher's theory to the epistemic domain and adjusts the relevant belief-desire pair. Accordingly, epistemic blame consists of a characteristic set of dispositions (e.g. reproach, upset, verbally request reasons) unified by their causal connection to a certain belief-desire pair (Brown 2020:399). The relevant belief would be that the agent has believed badly, where believing badly is understood as violating an epistemic norm without justification. The relevant desire pair is akin to the moral one, a desire that the agent had not believed badly. Similarly, frustrated desires lead to negative reactions such as resentment or anger. Finally, the alternative desire is that the believer has appreciated the relevant epistemic reasons that they previously ignored (ibid.:396).

How appropriate is the desire-based understanding of blame for vice? At first glance, it appears to cover most instances of vice. I believe someone has acted badly/holds a vice and desire that they did not (perhaps because of the badness that it resulted in, or the bad motivations it consists of).

However, there may be some exemptions. These take the form of instances where we want to blame the vice-bearer for their vice, and this blame is not constituted by a desire that the vice-bearer had not acted badly. In fact, I may be pleased that they acted in such a way, but I wish to blame them still.

Consider, for example, a friend who expresses a strong distrust of medical experts. Their distrust leads them to shut off any medical advice given to them by doctors or take seriously medical research on a range of topics, from vaccinations to a healthy diet. After begrudgingly attending a GP appointment, they are misdiagnosed with a heart condition and prescribed medication. Luckily, due to their excessive distrust, they do not believe the GP and chuck the medication in the bin. If they had taken it, it could have been highly dangerous.

Our response here is not that we wish our friend did *not* act closed-mindedly and believed the GP. This would have resulted in them taking the dangerous medication. However, we do still want to hold them responsible for their closed-mindedness, as they likely will not be so lucky next time. In this instance though, there is no belief-desire pair to constitute blame.

Boult also raises an example of blame lacking the desire that the blamed party had not acted in a blameworthy way (Boult 2021:6). This can occur when we blame an epistemic rival, as we may be glad of their mistake.

These examples illustrate that blame does not solely arise from a belief-desire pairing. There are some occasions where I do not desire that the wrongdoer had acted differently, yet I still feel the inclination to blame them. Given this inconsistency, let us turn to an alternative account of epistemic blame.

4.2.2 The Relationship-Based View

Next, we have the relationship-based view. Here, blame is connected to a different motivational component known as the relationship modification. Boult draws on T.M Scanlon (2008, 2013) to argue that members of an epistemic community stand in an 'epistemic relationship' with each other. This relationship involves a mutual set of intentions and expectations that are directed towards each other's epistemic agency. Members of an epistemic community expect others to meet certain epistemic conditions and default to trusting in each other unless they have good reason not to. When it is revealed that one agent may fall short of these epistemic expectations, for example, they are prone to conspiratorial thinking, others can then modify their trust in the agent. This modification is just what epistemic blame consists of.

One worry that arises with this type of blame, which is often pitted against Scanlon, is whether this form of blame is strong enough, as a number of authors have accused Scanlon of 'leaving the blame out of blame' (Wallace 2013:349). The lowering of one's epistemic expectations towards the vice-bearer may go some way in expressing blame, but not enough.

Of course, for some, the weakness of this form of blame may be the appeal of such a view. For example, lowering your expectations seems akin to Tanesini's preferred responsibility response of distancing oneself from the vice-bearer (Tanesini 2021:178). It might seem true then, that both of these amount to instances of epistemic blame.

However, another worry is that like the desire-based view, this account does not extend to all instances of blameworthy epistemic vice. Consider, for example, a scenario where I have no

grounds to trust my vicious neighbour, as I know they are dishonest. When my neighbour lies to me again, I am inclined to blame them. However, I will not do so by lowering my epistemic expectations of them, as they have already been lowered. Consequently, on this relationship account of blame, it remains unclear how I could assign blame to the vice-bearer if I am not lowering my expectations towards them.

4.2.3 The Emotion-Based View

The third view to examine is modelled on the popular emotional view of moral blame. For instance, Strawson (1962) provides an emotion-based account of moral blame which focuses on reactive attitudes (particularly resentment, indignation and guilt), whilst other emotional theories of blame are broader, including 'hostile attitudes', contempt or anger (Shoemaker 2015, 2017; Wolf 2011). Emotional theories of moral blame need not have widespread agreement over which emotions constitute blame. Rather, there is a shared commitment to the basic view that to blame is to respond to another's actions with a negative emotion.

Turning to the epistemic side of this view, epistemic blame is understood as the expression of reactive emotions such as indignation or resentment, directed at the epistemic agent. This is prompted by the belief that the agent has (culpably) violated an epistemic norm (Boult 2021:5). As Boult notes, few authors have directly defended this theory, but it has presupposed and dominated a wide range of discussions in epistemology (Menges 2017; Strawson, 1962; Wallace 1994, 2013; Wolf 2011). For example, we can speak of resenting people for their bad attitudes, hasty beliefs or a failure to believe what they should (McHugh 2012; Nottelmann 2007). This type of blame specifically targets the epistemic failing such as a faulty belief and involves holding others responsible for such harms (Rettler 2017).

Can this account apply to instances of epistemic vice? We can recall that Tanesini (2021) raised a direct criticism over the appropriateness of reactive responses such as resentment for our epistemic failings. Aligning such responsibility-responses with accountability, Tanesini remarks that we 'tend not to resent, punish, or get angry at people for their poor aesthetic judgement…or resent or punish people for their poor epistemic evaluations' (Tanesini 2021:173). Given Tanesini's criticism, an emotional account of blame does not seem well-suited to epistemic blame, nor vice.

To explore this intuition of whether we exhibit emotional responses to epistemic failings, consider this example:

Arrogant Andrew

Andrew is intellectually arrogant. He constantly interrupts and talks over his colleagues, believing he knows better on topics he has limited knowledge or experience of. He also spends most of his time boasting about his latest achievements and successes, showing no interest in what his colleagues have been up to. Andrew exhibits intellectual arrogance as defined by Tanesini (2016).

Tanesini details how intellectual arrogance can generate a variety of epistemic harms (ibid.:72). Specifically, it can foster epistemic vices on others, such as the vices of servility and timidity. We can imagine that Andrew's female colleagues who have dealt with his arrogance for many years, begin to form such vices. For example, his constant silencing of others causes one colleague, Katie to be too afraid to present her ideas in meetings for fear of being berated by Andrew or belittled. Another, Rachel, who once corrected Andrew for his inappropriate behaviour, is now too worn down to deal with him. She no longer speaks up and nods along to Andrew's correction of others.

We can say that Katie is intellectually timid. Andrew's actions amount to locutionary silencing, understood as instances where one is literally prevented from speaking (Tanesini 2016:76-77). This is a form of intimidation, which in turn fosters intellectual timidity in its targets. Timidity is a resignation to being treated with less than due respect. It can present as a fear of expressing one's opinions, a tendency to withdraw to avoid attention, and a reluctance to contribute to conversations (ibid.:88).

We can also consider that Rachel is intellectually servile as a result of Andrew's arrogance. She no longer trusts her own judgement and excessively defers to the views held by others, in this instance Andrew. Again, this can be a result of illocutionary silencing. By biting one's tongue, over time individuals defer to the opinions of others, becoming servile (Tanesini 2021:89).

In response to Andrew's behaviours, anger, contempt and resentment seem justified. These responses are directed at the epistemic behaviours that Andrew displays, such as his locutionary silencing of Katie and Rachel, and the vices of timidity and servility that he formed on others. These emotions are therefore directed at his vice, his intellectual arrogance.

We can anticipate that Tanesini, and others, may argue that our blame being directed towards Andrew is that of moral blame, and not epistemic. Tanesini argues for such a line of thought in a different example, discussing how a teacher's anger at her misbehaving students is because they show a lack of regard for her teaching efforts, which is a moral response (2021:173). This argument, that epistemic blame simply collapses into moral blame, has been proposed by others too, most prominently by Trent Dougherty (2012). I refer to this position as 'epistemic blame scepticism' and have argued that this type of reductionist reasoning employed by epistemic blame and normative sceptics, creates a total collapse of the normative realms (Meehan 2019). As arguments throughout this chapter will also demonstrate, we have good reason to believe that epistemic and moral blame can come apart.

With this objection aside then, it seems fair that we sometimes blame vice-bearers by expressing emotions. This means that emotion-based accounts of blame may seem apt for epistemic vice, especially if the intuition is met in the case of arrogant Andrew.

However, despite its promising nature, I argue that another form of epistemic blame is more suitable for vice.

4.2.4 The Agency-Cultivation View

The final, prominent account of epistemic blame is referred to as the agency-cultivation view. It is based on 'forward-looking' understandings of moral responsibility that aim to prevent certain behaviours in order to achieve certain ends. For example, Manuel Vargas (2013) argues that blame reactions often serve to discourage certain kinds of behaviour. Blame can therefore function as a kind of external motivator to behave in certain ways. Even further, this motivation gradually becomes internalized by the agent in the form of responsiveness to moral reasons.

Extending Vargas' argument, Adam Piovarchy (2021) develops an account of epistemic blame that encourages positive epistemic behaviour. According to this view, blame consists of a judgement that someone is blameworthy, which in turn makes them a suitable target for specific interpersonal responses (such as resentment or indignation). This creates a disposition to engage in blame reactions, which can include 'verbal condemnation, calls for censure or shame, avoidance, emotional distance…' (Boult 2021:7).

Modelled on this, epistemic blame plays a similar functional role, acting as a 'vector for epistemic agency cultivation' (ibid.). Epistemic blame is a judgement connected to a set of negative interpersonal reactions and functions to discourage bad epistemic behaviours. In turn, this can cultivate epistemic agency, understood as a kind of responsiveness to epistemic reasons (Piovarchy 2021:801).

The central feature of this account of blame is that the function of epistemic blame is to discourage bad epistemic behaviours and bring about epistemic goods. As Boult notes, the function of this form of epistemic blame 'is to bring people into the realm of epistemic agency, and to reinforce their capacities as epistemic agents' (Boult 2021:8). Blame then, is intended as a positive and ameliorative practice. It is because of this that I consider it the most suited form of blame for epistemic vice. Let us examine my reasons for considering this to be true.

4.2.5 Ameliorative Approaches to Blame

The positive function of blame can be fleshed out in a multitude of ways. For example, Paulina Sliwa (*forthcoming*) argues that responsibility is about acknowledging one's wrongdoings, understood as a normative footprint. Wrongdoing changes the normative landscape in three prominent ways:

I. *Wrongdoing creates reparatory duties/rights: the duty to apologize, to make amends, to explain one's motives, to acknowledge the harm done, to compensate.*

II. *Wrongdoing changes feeling rights/duties: the right to feel upset, angry, disappointed with us, the duty to feel anguish, guilt, remorse, and shame.*

III. Wrongdoing modifies existing relationship rights/duties: the right to trust, help, support, good will.

(Sliwa *forthcoming:*14)

Applied to epistemic normativity, we can see the same can be true when epistemic norms are violated. Blaming someone epistemically creates an awareness of the duties or epistemic rights one may have violated, that the wrongdoer must acknowledge and compensate for. It also can create new rights in the wrongdoing. Andrew's colleagues have a right to feel angry or disappointed in his arrogance. Finally, it creates modifications to existing rights and duties, with the wronged perhaps revising their trust in the wrongdoer.[59]

Fricker (2016:167) also defines the role of blame in a similar sense, labelling it as 'communicative blame'. Presented as a paradigm account of blame, it consists of responding to a wrong and letting the person know that they are at fault (ibid.:171-172). It also intends to help the wrongdoer acknowledge their wrong, but this motive may not always be a present motive in the psychology of the blamer (ibid:173). Blame also does not always need to be a verbal act, just as communication need not always be verbal. Blaming someone in this sense then can involve informing the wrongdoer of their wrong, distancing oneself or even silence (ibid.). Additionally, this blame may be attached to an action or omission, or the wrongdoer's motives, attitudes, dispositions, or beliefs.[60] Given the epistemic nature of what we can blame, it seems clear that Fricker's account is also suited to the epistemic realm.

Fricker understands the purpose of blame as being to 'inspire remorse in the wrongdoer' and to 'prompt a change for the better in the behaviour (inner and outer) of the wrongdoer.' (ibid.).[61] Specifically, the illocutionary act of blame is to make the wrongdoer feel sorry for what they have done. This differs from merely making them feel bad for selfish reasons, as it entails an acknowledgement of the significance of what they have done or failed to do. Fricker also notes that this amounts to blame even if the blamer does not knowingly intend to invite remorse (ibid).

[59] For more information on distinctively epistemic rights, see Watson (2021).
[60] Fricker observes that we can blame people for the doxastic aspects of their racism just as we blame them for the motivation behind it (2016:171).
[61] We may worry that if the wronged does not recognise the reasons for being blamed, the blame's value and ameliorative purpose is diminished (Tanesini 2021:190). Fricker responds to this type of concern, arguing that blame can still serve to make the wrongdoer understand the wrongness of their act (2016:176). We can also value blame for its positive effect on the wronged party, such as preventing further harm or adjusting duties towards the wrongdoer.

In this sense Fricker considers communicative blame to resist the 'bad reputation' that blame is often given.[62] Blame is often considered to be a projection of guilt or shame, moralistic and high-minded, or inspired by vengeance (Fricker 2016:168; Owens 2012:25). However, communicative blame, when employed appropriately, can be constructive and serve a positive aim. In this sense, Fricker states that the role of blame, as in the purpose it serves, need not be negative, and the various reactions that constitute blame also need not involve strong feelings such as resentment, indignation, or punishment.

The perspectives constituted by the agency-cultivation view, alongside the accounts presented by both Sliwa (*forthcoming)* and Fricker (2016), classify blame responses as having a common objective. Blame is categorised (broadly) as a reaction to some (epistemic or moral) norm violation or wrong, which itself aims to discourage bad conduct and encourage good conduct (where this conduct is epistemic, moral or both). In this sense, blame results in some positive outcome, by either producing or contributing to moral and epistemic goods. Focusing specifically on the epistemic goods, these can be defined as cultivating epistemic agency, recognising the rights of the wrongdoer, aligning understanding and reason, or inspiring remorse or redress in the wrongdoer. The bad epistemic behaviours which blame can discourage depend on what is being blamed, whether that be bad beliefs or epistemic vices.[63]

Through this broad aim, I categorise these accounts as having an 'ameliorative' approach to blame. This is because they seek to either reduce epistemic harms or promote and encourage epistemic goods. Blame that takes an ameliorative form is often aligned with feminist approaches to responsibility and often includes either (or both) communicative and functionalist accounts of blame (Michelle Ciurria 2021:8). One of the main proponents of an intersectional feminist account of blame is Ciurria, who argues that blame should be at least capable of satisfying intersectional feminist aims which can include: marking someone as a norm violator or perpetrator, seeking uptake from the respondent or expressing a negative attitude (2021:9). The purpose of blame can therefore be ameliorative, for it can 'instruct, inform, protest and challenge' which in turn can promote intersectional feminist aims by disseminating anti-feminist values and ideals (ibid.:11). Blame itself can also take a variety of

[62] Sher (2006) and Scanlon (2008) also defend blame's bad reputation.

[63] I also believe that self-blame is possible under this account, where a vice-bearer takes responsibility for their vices and the harms that flow from them. This would be a further way to understand the ameliorative effects of blame. However, this is more difficult to accommodate for communicative accounts of blame specifically. I also do not consider this the most effective form of blame for vice, given their stealthy nature (Cassam, 2015, 2019a; Tanesini 2021). If vices are hidden from the vice-bearer, this gives us more reason to opt for a form of blame that relies on others acknowledging our wrongdoings.

forms, including strong emotions such as rage if it is aiming to seek recognition, respect, and change. Ciurria also defends this form of responsibility against eliminativist accounts that seek to abolish responsibility altogether (ibid.:41).[64]

Ciurria presents the example of blaming someone for slut-shaming women through functionalist and communicative account of blame. By viewing the purpose of blame as to communicate a norm violation and to seek uptake from the blamed party, a response such as uttering 'I don't slut-shame women', can serve as a form of blame that can reprimand the slut-shamer, show support for the victim, or to seek recognition from a witness. In this sense, Ciurria argues that '…blame can realize and promote intersectional feminist aims: by transmitting normative information which speaks to its wider function.' (ibid.:11).

In drawing attention to Ciurria's feminist account of responsibility is to note that the practice of blame can be sensitive to the aetiology of vice, specifically the oppressive systems and structures that contribute to some vices. Nor does it automatically follow that we should withhold blame when vice-bearers have formed their norm via these structures.

It should be clear then, that it is at least possible that blame, specifically epistemic blame in this instance, can have positive results, whether these results are felt by the wronged, the wrongdoer, or both.

One point to briefly consider here is that an ameliorative understanding of blame effectively overcomes concerns raised by both Alfano (2014) and Tanesini (2016, 2021). As discussed in Chapter 2, one objection is that that blaming agents for their epistemic vices can act as a self-filling prophecy, or worse, can result in the formation of further epistemic vices. Because of this, we should avoid blaming agents for their epistemic vices. However, it should be clear that by developing a more sophisticated and ameliorative understanding of blame, blame amounts to more than simply pointing out faults in the agent or labelling their vices. Instead, it can effectively make steps towards the revision and eradication of the vice or generate positives for those harmed as a result of vicious behaviours.[65]

[64] This follows Macnamara's (2015) communicative account of blame. Macnamara argues that blame is analogous to a message and has two core features. The first is that it represents someone as a norm violator and the second is that it aims to make the wrongdoer understand and accept this representation.

[65] I wish to make it clear that I do not believe blame is the *only* ameliorative solution to vice. My aim is to motivate how an ameliorative understanding of blame is compatible with instances of epistemic vice. I also discuss ameliorative solutions to vice in more detail in Chapter 6 of this thesis.

I consider blame that has an ameliorative purpose to be the most suitable form of blame to attribute to epistemic vice.[66] From this approach, it is at least possible that blaming a vice-bearer can result in the prevention of further harms or promote epistemic goods (e.g., it can encourage an understanding of one's bad behaviour, generate rights for the wronged party, or allow the wrongdoer and wronged party to make epistemic adjustments to protect themselves further). Blame can also be expressed in a variety of ways, as long as they are keeping with this ameliorative aim.

Importantly, this ameliorative aim is shared widely among vice-epistemologists, despite not considering blame itself to be one way of achieving this aim (Battaly 2019; Cassam 2019a; Holroyd 2020:141; Kidd 2020; Tanesini 2021).

For example, Cassam argues that one of the reasons to concern ourselves with the study of vices is to find effective ways to reduce the harms they result in (Cassam 2019a:186-187).

To summarise, I have argued that a distinctively epistemic type of blame is applicable to epistemic vices. This blame can also be ameliorative, defined broadly as a reaction to an epistemic norm violation that seeks to reduce epistemic harms and/or promote and encourage epistemic goods. Here, I have gone some way in motivating my first objective, to illustrate that an epistemic and ameliorative form of blame can be suitable for epistemic vices.[67] I will continue to motivate this position by addressing concerns on the appropriateness of assigning blame to epistemic vice.

4.3 The Argument from Lack of Control

As discussed in the previous chapters, many vice epistemologists have argued that blame should not be an appropriate response to vices that are formed outside of the vice-bearer's

[66] I am not committed to any specific account under this umbrella. However, I tend to agree with Ciurria (2021) that the most effective form of blame for vice would likely be functionalist or communicative. In this sense, the resulting bad conduct that is prevented or good conduct that is encouraged will depend on the specific account.

[67] The ameliorative effects of blame can be substantial e.g., Ciurria's (2021:13) view that blame can function to realise, advance or promote feminist intersectional principles. This blame is still valid even if there is no uptake on the blame too.

control (Battaly 2016a, 2018a; Cassam 2019a; Kidd 2016, 2020). I refer to this overall position as the 'argument from lack of control'.

Three main concerns are offered in support of this argument, which I will briefly discuss below having already alluded to them in the first two chapters of this thesis. Through examining these arguments, I will also illustrate how they oppose the responsibilist view that blame is integral to vice, which I refer to as the 'constitutive blame thesis'.

The first argument is offered by Battaly, whose concern over the lack of control over vice acquisition is used to motivate and ground her account of epistemic vice, 'personalism' (Battaly 2016a, 2018a). Battaly's hybrid account of epistemic vice draws from virtue reliabilism and responsibilism. It claims that epistemic virtues and vices are character traits or personal qualities expressed by the agent (the responsibilist claim), however, we need not *necessarily* be responsible for possessing or exercising these qualities. It follows from this that we need not be praiseworthy or blameworthy for our epistemic vices (the reliabilist claim).

Battaly's motivation for the latter claim is that individuals often lack control over their acquisition of vice, meaning it is unfair and inappropriate to blame or praise agents for their vices (Battaly 2016a:108; 2018a:120). She provides support for this view by citing examples where agents lack control over their environment and actions that result in the development of vices. One such example is the case of Robert Harris, who after years of suffering from abuse and abandonment, murdered two victims in San Diageo in 1978 (2016a:107). Battaly appeals to Gary Watson's (2004) analysis of this case, arguing that even though Harris possessed the vice of cruelty, his upbringing undermines the judgement that he is responsible for possessing the vice.[68]

Battaly's concern mirrors a wider debate surrounding control and responsibility. Within discussions on epistemic responsibility, there is a concern about whether individuals can be blamed for their bad beliefs (Heironymi 2008; Levy 2005; McHugh 2013, 2014). One perspective, known as doxastic voluntarism, asserts that we do have voluntary control over our

[68] As explored in Chapter 3, Battaly revisits her position on blame in response to an objection that she might be excusing vice-bearers for their bad behaviour (2018:124-125). She argues that a form of blame that is not contingent on could be a suitable form of responsibility for vices. However, she emphasises that this blame is not a necessary feature of vice. I will return to this argument momentarily, focusing for now on her view that responsibility requires control.

beliefs (Peels 2017; Weatherson 2008). Conversely, doxastic involuntarism argues that we cannot be blamed for our beliefs as we frequently lack control over them (Alston 1989).

Within doxastic voluntarism, a secondary debate emerges as to whether this voluntary control is direct or indirect. Direct voluntary control refers to actions that occur immediately upon choice, such as choosing to recall your favourite memory (Ginnet 2001; Weathersoon 2008). Indirect voluntary control refers to actions for which an individual lacks direct control, but they can influence these actions by choosing to perform a series of intermediate actions (McHugh 2013). For example, someone unskilled in carpentry may have indirect control over building a wooden table. Although they cannot directly control this action, they can influence it by choosing to learn carpentry skills and gathering materials.

Returning to vices, Battaly contends that 'an individual might have had little or no control over which character traits she came to possess' (Battaly 2018a:115). She also argues that our 'environment or education might have done all the work' when it comes to the formation of our virtues and vices, or we possess them due to matters of luck or adversity (Battaly 2016a:100). This suggests that depending on the vice in question, we can have no control or only indirect control over its formation. Similar to debates on control and responsibility, this lack of control undermines our responsibility for vice, rendering blame to be 'underserving' (ibid.).

Battaly's argument therefore consists of two key claims. The first is that blame is frequently an unsuitable response to epistemic vices due to the lack of control vice-bearers have over their vice formation. The second is that even if blame is warranted, it is not an integral feature of vice.

Presenting a similar scepticism over our blaming practises, Kidd asserts that an agent's frequent lack of control over their vice makes the practice of 'vice-charging' difficult. Vice charging is understood as the practice of implicitly or explicitly holding an agent responsible for their display of vice or the actions which stem from it (Kidd 2016:181-182).

Kidd formulates this concern as one centring around the relationship between agential responsibility and epistemic vices, and argues that we should pay closer attention to how

oppression can damage our epistemic character (Kidd 2020:69).[69] Certain environments can be 'epistemically corrupting', meaning 'one's character comes to be damaged due to one's interactions with persons, conditions, processes, doctrines or structures that facilitate the development and exercise of epistemic vices' (ibid.:71).[70]

Recognizing the role that social oppression can have on our epistemic character prompts a closer examination of how individuals acquire or develop their vices e.g., through systems of social privilege, power hierarchies, or unjust institutions. We can also focus more widely on the psychological bases of epistemic vices as well as the situational factors and broader sociological, cultural and ideological conditions that shape the development and maldevelopment of epistemic agents (ibid.:70).

Whilst Kidd does not discuss matters of control here explicitly, he does detail how epistemically corrupting and oppressive conditions are difficult to resist for most epistemic agents (Kidd 2020:74). Utilizing Medina's (2012:28) concept of an 'epistemic predicament', Kidd argues that how vulnerable one is to resist epistemically corrupting conditions depends on the scope, strength, stability and specificity of the vice (Kidd 2020:74). However, given that the vast number of vices are formed in at least partially corruptive environments, it can be assumed that many vices are developed beyond an individual's control.

Addressing these concerns, Kidd advocates for 'aetiological sensitivity', defined as a commitment to actively attend to the complex, contingent conditions under which the epistemic characters of subjects develop (ibid.:78).[71] This should influence how we respond to epistemic vices, particularly when it comes to our responsibility responses (ibid.:79). Certain forms of blaming may be integral to oppression itself, concealing unjust systems that produce those vices and shortcomings (ibid.). Additionally, concentrating on blame 'occludes other responses to corrupted subjects, like anger, disappointment, regret, and sadness' (ibid.:80). For example, sometimes a more appropriate response to vice may be to express anger or sadness towards the system that brought about the vice in the first instance. [72]

[69] See also Dillon (2012); Medina (2007); and Tanesini (2021:182) for a similar argument.
[70] See Chapter 6 for a further discussion of Kidd's account of epistemic corruption.
[71] Kidd further argues that a form of vice epistemology that is sensitive to vices in this way can be referred to as 'critical character epistemology' (Kidd 2020:76). See also Dillon (2012:100).
[72] This implies that Kidd considers these responses to be distinct from blame.

These concerns suggest that when confronted with epistemic vices we should be suspicious of blame responses, for they are often 'counterproductive, unjust and liable to perpetuate patterns of oppression' (ibid.). Whilst Kidd does not explicitly argue against the responsibilist claim that vice-bearers are inherently responsible for their vices, it follows that if blame and vice can come apart, the responsibilist claim must be rejected.

Finally, a similar argument for how control undermines responsibility for epistemic vice is offered by Cassam (2019a). Cassam's account of vice, obstructivism, states that epistemic vices are character traits, attitudes or thinking styles that agents are blameworthy or otherwise reprehensible for. However, raises concerns over agents' vice acquisition, and particularly, whether blame is the appropriate response to vice or if a weaker form of responsibility, like criticism, is better suited.

Cassam presents two forms of responsibility for vice: acquisitional and revisional responsibility. According to Cassam, acquisitional responsibility is the view that an agent is responsible for their vice if they are responsible for acquiring or developing it (2019a:124). The implication behind this view is that one is responsible and blameworthy for a vice because they acquired it voluntarily. Virtues are something that are acquired by individuals through training and habituation and are not innate (Zagzebski 1996). For example, acquiring a virtue such as open-mindedness takes time and effort, and it is in this sense that one is responsible for being that way.

However, Cassam rejects this version of responsibility for vice, claiming that it does not paint a plausible picture of vice acquisition and is only concerned with the actual or imagined origin of one's vices (Cassam 2019a.:125). Vices do not require time or effort and vice acquisition does not require training. For example, arrogance is not something that people practice or work hard to achieve and is something that can be acquired naturally. Furthermore, Cassam claims that character formation often occurs in the development stages of our childhood when individuals are still immature, meaning it would be unfair to blame agents for vices they had no control over acquiring in these stages (ibid.:128). In this sense, Cassam claims that vice-bearers 'can't properly be blamed for something over which one has no control', thereby positioning his perspective on blame around the concept of control (ibid.:127).

Turning to revision responsibility, Cassam discusses how this requires managerial control (ibid.:124). This concerns the manipulation of objects or beliefs to make them coincide with our thoughts about said object/belief (e.g. changing the layout of a room). Cassam concludes that blame is often contingent on whether managerial control is present for the vice. This is because it relates to whether our character traits are malleable or not. This means that if the trait is not malleable, then people lack revision responsibility for it and are not blameworthy.

However, Cassam claims that the trait can still be categorized as an epistemic vice if the agent is deemed responsible for it in a different sense, namely if their trait can be criticised (ibid.:133). For an agent to possess a criticisable vice, the vice must reflect badly on the agent (Sher 2006). Whether this is the case depends on the nature of the trait (on its harmfulness) and whether it is a deep or superficial trait – one that defines, or helps to define, the kind of thinker/knower one is. For the trait to reflect badly on a person, it is not necessary that they are revision responsible or blameworthy for it, nor that the trait is malleable.

Cassam's arguments provide further support for the claim that vices-bearers are not in control of their vices, which in turn casts doubt over the constitutive blame thesis that blame is integral to vice. We can also see how under Cassam's account, it is criticism, not blame, that is integral to the definition of an epistemic vice. One of the basic requirements for a character trait, attitude or thinking style to be categorised as a vice is that it must be criticisable. Whilst in some cases we can go one step further and blame the agent for their vice, this is not a necessary requirement of vice. This further supports the claim that blame is not integral to vice.

The above arguments offered by Battaly (2016a, 2018a) Cassam (2019a) and Kidd (2016, 2020) have made up the broader view that I have termed 'the argument from lack of control'. In essence, these arguments claim that as vice-bearers often possess little to no control over their vices, blame is rarely an appropriate or fair response. If blame is not a suitable response for vice, it follows that it cannot be a constitutive feature of vice.

4.4 Attributability Responsibility

There are at least three possible responses to the argument from lack of control. Firstly, we can choose to abandon the concept of blame for vice altogether or at least consider it rarely

applicable to epistemic vices (Kidd 2016, 2020). Secondly, we can choose to criticise vice-bearers for the vices they have limited control over but refrain from outright blaming them (Cassam 2019a). Thirdly, we can choose to adopt a form of blame without control conditions, such as attributability responsibility (Battaly 2016a, 2018a; Tanesini 2016, 2021).

To speak to the first response, abandoning the possibility of blame for vice may appear like the most straightforward solution given its seemingly problematic application. However, by drawing from our previous discussion on the ameliorative approach to blame, it is easy to find reasons as to why it is important to blame vice-bearers. In recognizing blame as a valuable practice that aligns with our intuitions for addressing poor epistemic conduct and the ameliorative aims of vice epistemology, there appear to be ample reasons not to reject it.

We can also briefly address here Kidd's scepticism towards blame (2016, 2020). In addition to concerns about the restricted control we possess over our vices, Kidd argues that blame can serve as a mode of oppression or prevent other responses such as anger, disappointment, and regret from being assigned (Kidd 2020:79).

To speak to both of these concerns, I have argued that any blame assigned to epistemic vices should be ameliorative in spirit, meaning a form of blame that perpetuates oppression would not be a relevant response to vice. For example, a form of blame that only serves to silence or keep people fearful would not be the type of blame fitting for an epistemic vice. It would not be constructive or improve the vice-bearer's character or bring about positive change. As also discussed, there is a wide range of blame responses under this umbrella that we can adopt, including anger, sadness, and disappointment.

We do not, therefore, have compelling reasons to abandon the concept of blame for epistemic vices based on these objections.

I will return to the second response, to assign criticism, and not blame to vices, towards the end of this chapter. For now, I will focus on the third response. Here I will motivate the argument that a form of blame without control, attributability responsibility, is compatible with the nature of epistemic vice. In turn, I will object to the argument from lack of control.

To motivate this position, let us briefly discuss the debate on control and moral responsibility (Arpaly 2003; Fischer and Ravizza 1998; Scanlon 1998, 2008; Sher 2006, 2009; Smith 2005, 2013). The discussion on whether control is a requirement for moral blame stems from debates on determinism, where control entails the freedom to choose differently (Slote 1982:24). The 'voluntarist' argues that individuals cannot be blamed for what was not under their control (Levy 2005; Rosen 2004). Conversely, the 'non-voluntarist' position, claims that blame does not require control (Hieronymi 2008; Scanlon 1998; Sher 2006, 2009; Smith 2005; Talbert 2008).

A perspective that is commonly aligned with the non-voluntarist stance is attributability responsibility, which requires relatively little in the way of control. According to this view, assigning responsibility does not hinge on the degree of control that an individual has over their 'attributes', understood as their actions, traits, attitudes, or mental stances.[73] Instead, it depends on whether these actions are reflective of the wrongdoer's authentic or deep 'self', their beliefs, attitudes, values, or commitments (Sher 2006:57; Shoemaker 2015:38; Watson 2004: 270).

Evaluating someone based on their action or attitude therefore goes beyond superficially assessing the specific attribute. We should assess whether it is properly attributable to them and whether it reflects their 'judgement-sensitive attitudes' (Scanlon 1998:180), 'evaluative judgements' (Smith 2005:251) or 'moral personality' (Hieronymi 2008:362). When these judgements are harmful, inappropriate, or otherwise objectionable, then we can be deemed appropriately blameworthy for them.[74]

Attributionism would therefore not excuse individuals from blame if their characters were formed under challenging conditions (Scanlon 1998:278-285) if what they are being blamed for was out of their control (Hieronymi 2008; Smith 2005: 267-270) or the individual could not recognize the moral implications of their behaviour (Talbert 2012).[75]

[73] I employ the term 'attributes' as an umbrella term encompassing traits, beliefs, actions, mental states etc., which are subject to evaluation (Hieronymi 2008; Levy 2005; Smith 2005).
[74] There is some debate as to whether the attributionist account of blame amounts to a robust form of blame (Levy 2005; Wallace 1996). Following attributionist accounts, I hold that to consider someone responsible is to consider them deserving of praise or blame (Hieronymi 2008:358; Talbert 2008, 2019). This blame can take a variety of forms (e.g., anger, punishment, recognition of a norm violation) and serve different purposes (e.g. emotional release or communicating a wrong-doing). Battaly also appears to consider attributability responsibility a form of blame. Discussing the example of Harris and the Hitler Jugend, she argues that 'if attributability responsibility is viable, they will be blameworthy for their vices nonetheless' (2016a:114).
[75] Tanesini (2021:178) argues that Smith's account of responsibility is more closely aligned with answerability. I disagree with this claim, agreeing with Talbert's (2019) view that Smith's view is an attributionist one.

Control is not completely irrelevant to attributionist accounts, however. It can play a part in determining certain 'excusing conditions' for when actions cannot be attributed to agents, meaning they cannot be deemed responsible. These are conditions where a person's behaviour does not flow from their deep-self, or they are not acting fully as an agent. This can include young children, instances of brainwashing, accidents, or coercion (Zheng 2016:65). The key claim, however, is that according to attributability responsibility, control is not an essential requirement for responsibility. This allows individuals to be responsible if they do have voluntary control, but they also contend that individuals can be responsible even if they do not.

We are therefore only responsible for consistent, 'full-blooded' attributes which are truly 'ours' (Smith 2005:237; Zheng 2016:64). They are 'ours' and consistent because they reflect our deepest values and judgements, our true 'personality' (Hieronymi 2008:362).

Attributability responsibility is distinguishable from accountability responsibility, a non-voluntarist position that has stricter control conditions (Watson 2004:273). Broadly, on this view, individuals are held accountable for an action only if they had a reasonable opportunity (directly or indirectly) to avoid violating the standards for what they are being held responsible for (Watson 2004:276). For example, a worker who is aware of the ethical guidelines at work is accountable for violating them given that they had a reasonable opportunity to avoid breaching them.

Having outlined the distinction between attributability and accountability responsibility, we can now examine how attributability responsibility is an appropriate way to hold individuals responsible for their epistemic vices.

Consider an individual, Jason, who is epistemically insouciant. He displays a lack of concern for the truth or other epistemic goods and is indifferent to whether his beliefs are grounded in reality (Cassam 2018:2).

Leaving aside the responsibility condition momentarily, Jason's behaviour constitutes an epistemic vice. His epistemic insouciance is consistent as opposed to fleeting, it belongs to him in the sense it is his personal character trait, and it makes him a bad thinker - whether that be due to the bad motive it expresses, the bad consequences it results in, or in some other way that it impairs him as an epistemic agent (Kidd et al. 2020:1-2).

Under an attributability responsibility framework, we are responsible for attributes that are 'deep', consistent, and express our real values and judgements. When these conditions are met, we are responsible for the behaviour and an appraisal is warranted, which can take the form of praise or blame depending on whether the behaviour is objectionable or commendable.[76]

Jason's insouciance seems to be properly attributable to him, given that it is consistent and expresses his deep self. It also expresses an objectionable value or judgement, such as a casualness or indifference to epistemic goods (Cassam 2018:5). We can also assume that there does not seem to be any 'excusing conditions' that excuse his behaviour, e.g., he is not being controlled by an evil demon or is too young to be considered an agent in the sense appropriate for blame.

As Jason's insouciance can therefore be properly attributed to him, and because it stems from an objectionable value or judgement, he can be considered blameworthy.

We can see that this blame is not contingent on the extent of control that Jason possessed over his insouciance, such as the conditions it was first formed under. Instead, it depends on whether it can be attributed to his authentic self through his objectionable values and judgements.

Attributability responsibility can therefore explain how we can blame vices formed outside of the vice-bearer's control. But what about concerns that we should not do so because it is unfair? (Battaly 2016a, 2019a; Cassam 2019a; Kidd 2016, 2020).

Let us consider an example where a vice-bearer had no control over the formation of their vice and lacked reasonable opportunities to reflect on or change their vice. I consider these types of scenarios to be the most compelling instances where blame would be deemed unfair. One such example is Smith's (2005:267) example of Abigail, who was raised by a deeply racist family and community. Abigail develops evaluative tendencies and attitudes that align with prevailing views from her upbringing. Even in adulthood, her attitudes persist and continue to reflect the bad judgements formed in her childhood. We can contrast Abigail with Bert, who was raised in a loving and tolerant home and community. However, later in his life, he adopts racist and

[76] Most accounts of attributability responsibility are neutral on what makes these evaluative judgements or values 'objectionable'. Tanesini (2021:175) considers the mere attributability of a bad trait to be sufficient for negative appraisal, in the same sense that something reflecting badly on an individual is enough to warrant criticism (Cassam 2019a:133-135). However, Battaly (2019) presents challenges to this interpretation, as discussed in the next section.

intolerant values. Abigail's racist beliefs were ingrained in her long before she could reflect on these evaluative judgements, whereas Bert's racist beliefs stem from his own considered and mature endorsement.

By most intuitions, Bert seems blameworthy for his racist beliefs and attitudes. However, even though Abigail's attitudes and beliefs are bad, voluntarists regard it as unfair to blame her for them given that she had no control over the conditions that led to the vice's formation, nor reasonable opportunities to reflect on the wrongness of her behaviour. Abigail may become blameworthy at a later stage if such opportunities arise, but at this moment, she is not blameworthy. Blame would therefore be an unfair response.

Smith (2005) responds to this voluntarist argument. She argues that to consider Abigail's belief or attitude as attributable to her and consequently 'a legitimate basis for moral appraisal', we do not need to consider she is responsible for 'becoming the sort of person who holds such an attitude' (Smith 2005:267-8). What we should be evaluating is Abigail's beliefs per se, and whether it reflects an objectionable judgement.

This implies that by excusing blame on the grounds that Abigail had limited control over the formation of her vice, we are conflating questions of attributability with whether she was responsible for *becoming* closed-minded. However, when we blame someone, we are responding to the badness in the content of their attitude or value, e.g. the wrongness of regarding one race as superior to the other. We are *not* responding to the origin of their action.

As Smith observes, if Abigail responded to accusations of blame by stating "I am not responsible for my attitude—I was just raised this way", we would not be convinced to withdraw our blame. This is because 'citing the origin of one's attitude is irrelevant when what is in question is its justification' (Smith 2005:268).

Focusing on the act in question and not the origin need not be unfair. We can still be sensitive to Abigail's circumstances and even deem Bert to be more blameworthy for his racism by adjusting the degree of responsibility or the type of blame that we assign to them both (ibid.). Here we can return to earlier arguments on the ameliorative role of blame and the variety of our blame responses. Blame need not involve strong reactions of resentment, punishment or indignation and can serve a practical role (Smith 2013:32). For example, we may blame Abigail

by communicating to her the badness of her beliefs and the harm that they cause. Alternatively, we may distance ourselves from Bert in order to prevent being harmed by his racist beliefs. As Smith observes, a sensitivity to Abigail's circumstances '…is quite compatible with claiming that both Abigail and Bert are fully responsible for their attitudes and for the judgements they reflect' (ibid.).

These arguments demonstrate that attributability responsibility, combined with an ameliorative perspective of blame, allow for blame to be a fair response to epistemic vices, even those formed or maintained in environments beyond one's control. This responds to the array of arguments opposing the suitability of assigning blame to epistemic vices due to a lack of control or perceived unfairness (Battaly 2016a, 2019; Cassam 2019a; Kidd 2016, 2020).

4.5 Battaly's Responsibility Problem

As discussed previously, despite arguing for the claim that a lack of control diminishes our responsibility for vice, Battaly reconsiders her position on blame in response to an objection that she might be excusing vice-bearers for their bad behaviour (Battaly 2018a:124-125). She proposes that a form of blame is conceivable for personalist vices, but this blame is not inherently linked to the vice. The type of blame Battaly favours aligns with attributability responsibility, asserting that for one to be responsible in this sense, 'the trait must be a full-blown personal quality—it must express the subject's 'real self'; i.e., her evaluative judgements and corresponding motivations' and 'the subject must be generally responsive to reasons and must also have the capacity to recognize the trait as her own and to evaluate it.' (Battaly 2016a:113).

According to this revised definition, personalism holds that agents are responsible for their epistemic vices, even those that they lack control over the formation of, but this is not a necessary requirement for the definition of vice.

However, Battaly (2019) presents a challenge to non-voluntarist accounts of responsibility that attempt to explain how we are responsible for our epistemic vices. Battaly argues that interpreting responsibility in this way creates issues for our categorisation of vices and subsequently, what we can be held responsible for.

If we are too restrictive in determining what can be properly attributed to us, there is a risk of excluding implicit biases from our responsibility practises. Alternatively, a broader view of what can be properly attributed to us risks being overly inclusive and including cognitive defects e.g., blindness, meaning these are subject to responsibility practises.

Essentially, then, the dilemma for vice epistemologists is to formulate a non-voluntarist responsibility analysis that encompasses implicit biases but excludes cognitive defects.

To elaborate on this concern, Battaly considers Cassam's (2019a) claim that vice-bearers are criticisable for traits that stand in a close enough relation to them or are 'deep' in the sense that they help to define the kind of thinker and epistemic agent one is (Cassam 2019a:23). Let us briefly recap Cassam's position.

Firstly, there are 'cognitive defects' for which neither blame nor criticism is appropriate e.g., blindness (ibid.:127-128). These are not categorized as epistemic vices since they are not at least criticisable. Secondly, there are intellectual failings that for one reason or another are not blameworthy but are open to criticism. This is true whether or not the agent exercised control over their trait (ibid.:134). These qualify as epistemic vices. Thirdly, there are intellectual failings that are not just criticisable but are also blameworthy. This blame is based on revisional responsibility, which states that for an agent to be blamed for V they must be in control of it to be able to revise or modify it (ibid.:124-125). These are also considered epistemic vices.[77]

Battaly focuses on Cassam's non-voluntarist claim that epistemic vices are at least criticisable, despite a lack of control over their formation. She focuses on Cassam's rationale that an intellectual failing can be criticised if it stands in a 'close enough relationship' to the agent to reflect badly on them (Sher 2006:57). Failings that stand in a close enough relationship to the agent are also 'deep' qualities rather than superficial ones, meaning they 'define the kind of thinker or knower one is' or 'define the kind of intellectual or epistemic agent he is' (Cassam 2019a:134).

Battaly considers, whether, on Cassam's account, implicit biases such as testimonial injustice are criticisable in the same way (Battaly 2019:28-29) She considers Fricker's (2007:39)

[77] See Chapter 2 for a detailed assessment of Cassam's account.

example of a card-carrying feminist, who, despite her explicit beliefs, implicitly perceives men to be more credible than women due to her prejudiced perception. If her unjust behaviour is deemed reprehensible, it must be because the implicit bias is closely connected to her. This is because even if the prejudiced perception does not align with her explicit beliefs, it still shapes her implicit beliefs and motives, defining and influencing her as an epistemic agent (Battaly 2019:30). This seems to get the right result so far; we should be held responsible for our implicit biases (Battaly 2019:29; Cassam 2019a:169–173).[78]

However, by this same standard of responsibility, cognitive defects such as blindness also appear to be reprehensible. They also count as deep qualities because they also shape an agent's implicit motives and beliefs (Battaly 2019:30). Yet we should not want to accept this conclusion, given it could mean we are responsible for certain cognitive impairments and disabilities.

Battaly considers this objection to apply beyond Cassam's non-voluntary account of criticism, and to any non-voluntarist account that wants to consider implicit biases to be (at least) criticisable but not cognitive defects.

The dilemma therefore lies in the fact that a non-voluntarist approach to responsibility for epistemic vices is either too narrow and risks excluding implicit biases or is too broad and includes cognitive defects as vices for which we are responsible for.

How can we still defend attributability responsibility for vice given this dilemma? One way to escape this concern is to explain how certain vices are non-voluntary responsible in a way that captures implicit biases but *not* cognitive defects. I believe this is possible if we appeal to attributionist accounts that distinguish between morally relevant and irrelevant factors when assigning blame and apply this to vice responsibility too.

Under Smith's (2005) non-voluntary account, we are not responsible for *everything* that is attributable to us. Whether an attribute is attributable to us in a 'relevant sense' or is a 'legitimate basis for moral appraisal' is not just down to whether the trait reflects some deep quality of ourselves, but also whether it reflects an 'evaluative judgement' (ibid.:237).[79] This means that

[78] See also Holroyd (2016, 2017); Schmidt (2022) and Zheng (2016).
[79] See also Smith (2004, 2007, 2008, 2012, 2015).

the quality which is up for assessment should be tied to some value or judgement. If this evaluative judgement itself is '…mistaken, inappropriate, or otherwise objectionable' (ibid.:254), then we are considered responsible for it in a way that warrants negative appraisal.

Continuing the previous example, we can therefore be held responsible for our sexist beliefs if they reflect an evaluative judgement which is objectionable, such as the judgement that some genders are superior to others. However, we will not be considered responsible for a spontaneous sexist thought that enters our mind and does not bear a rational relation to our underlying judgements. This latter thought is not attributable to us in the normative, relevant sense required for responsibility, despite being in principle attributable to us, in so far as it is a thought occurring in my own mind.

Smith also argues that the attribute which is subject to evaluation need not 'need not arise from conscious judgement, choice, or decision' (ibid.:252). This implies that implicit biases can be blameworthy if they reflect evaluative judgements that are deemed problematic. In contrast, a cognitive defect, despite being a 'deep' quality attributable to me, is not dependent on an objectionable evaluative judgement. For example, my impaired vision is not normatively connected to my judgements in the relevant sense required for responsibility.

What Smith's account demonstrates is that we are not blameworthy for an attribute *solely* because it reflects badly on us or defines the kind of thinker we are. The attribute must also 'belong' to us in a normatively relevant sense, by reflecting an objectionable evaluative judgement. This allows individuals to be held responsible for their implicit biases as they can subconsciously reflect objectionable evaluative judgements. Importantly, it also prevents us from being held responsible for our cognitive defects as they are not attributable to us in a normatively relevant sense, for they do not reflect objectionable evaluative judgements.[80]

To summarise, by distinguishing between the relevant and irrelevant factors for blame we avoid Battaly's concern to non-voluntarist accounts of vice responsibility. We get the right result that

[80] This line of argument extends beyond Smith's non-voluntarist account of responsibility, meaning I am not committed to this specific account of attributability responsibility. For example, under Scanlon's (1998:20) view, we are responsible for actions that appropriately express our 'judgement-sensitive attitudes'. This means that we are responsible for aspects of ourselves which are sensitive to our judgements. As Levy (2005) observes, on this view 'it makes no sense to ask me to justify my height, my skin colour or my compulsions, simply because none of these aspects of me are sensitive to my judgements' (Levy 2005:4). However, individuals can be asked to justify their 'fundamental values and sense of what is important and what trivial....my judgements and sensitive attitudes reveal where I stand on questions of value' (ibid.).

we can be blameworthy for our implicit biases, and not our cognitive defects. As we are not blameworthy for the latter, they will also not be considered epistemic vices.

4.6 The Constitutive Blame Thesis

To recap, I have argued for two of my three objectives. One, that an epistemic and ameliorative form of blame should be assigned to epistemic vice, and two, that this is a form of attributability responsibility. We can now turn to my third and final aim, that the form of blame I have motivated so far is integral to the definition of epistemic vice. As we have seen, objections to this claim have been proposed by many vice epistemologists (Battaly 2016a, 2018a; Kidd 2016, 2020; Tanesini 2018, 2021). Cassam was the most favourable towards an inherent responsibility condition on vice but argued that criticism, not blame, was integral (2019a:4).

Having laid most of the groundwork in my previous discussion, I will demonstrate that by adopting an attributability form of responsibility for vice, it necessarily follows that this form of blame is integral.

The first point to acknowledge is that it seems clear that attributability responsibility best suits our understanding of epistemic vices as character traits, attitudes and thinking styles. As we have seen, under an attributability responsibility framework we are responsible for attributes that are 'deep', consistent, and normatively relevant expressions of our values and judgements. When these conditions are met, we are responsible for the behaviour and an appraisal is warranted which can take the form of praise or blame depending on whether the behaviour is objectionable or commendable.

These conditions are conducive to the criteria of what determines a vice. A vice must be consistent as opposed to fleeting, properly 'belong' to the individual and be 'bad' in some sense, whether that be due to the bad motive it expresses, the bad consequences it results in, or in some other way that it impairs the individual as an epistemic agent (Kidd et al. 2020:1-2).

What this suggests is that the qualities that attributability responsibility considers to be blameworthy, just are vices. In an epistemic sense, they are deep qualities of an epistemic agent that are consistent and reflect some objectionable value or judgement. Attributability blame

theorists hold that we are always blameworthy for such qualities. Therefore, if we accept that we are attributability blameworthy for these qualities, as argued, we are blameworthy for our vices.

Attributability responsibility is a form of character-based responsibility, designed to explain the strengths and weaknesses of our moral or epistemic character - our epistemic virtues and vices. For example, Watson (2004) states that attributability responsibility is the type of responsibility we assign to virtues and vices. He claims that 'to blame [morally] is to attribute something to a [moral] fault in the agent...[Such] kinds of blaming and praising judgements...invoke only the attributability conditions...these appraisals concern the agent's excellences and faults—or virtues and vices...' (2004:266). Likewise, Gabriele Taylor details how 'to merely to use the labels "virtue" and "vice" is to indicate candidates for praise and blame' (2006:6). Whilst Watson and Taylor are referring to moral virtues and vices, it is easy to make the same comparison for our epistemic virtues and vices.

Attributability responsibility therefore directly refers to the type of blame for epistemic vices or the type of praise for epistemic virtues. To blame is to attribute some relevant quality to your deep self that is objectionable in some relevant sense (Hieronymi 2008; Scanlon 1998; Smith 2005). These attributions are epistemic vices.[81]

If we hold that vices are responsible in the attributability sense, it necessarily follows that blame is integral to the definition of vice. It is essentially the case that you cannot be attributability blameworthy unless you are epistemically vicious. This is because attributability-blame *just is* blame for vices. Any epistemically bad behaviour that is attributed to us in this relevant sense is therefore blameworthy.[82]

From this, it demonstrates that if Battaly (2016a, 2018a) and Tanesini (2018, 2021) are willing to commit to a form of attributability blame for vice, it must be the case that this blame is integral to the definition of vice. It cannot be the case that attributability responsibility is

[81] We can also be blamed for the epistemic harms that vices result in, which trace back to the vice. Also recall that considering someone responsible is to consider them deserving of praise or blame (Hieronymi 2008:358; Talbert 2019). Blame is therefore understood in a broad sense and can be gradable.

[82] What makes the behaviour bad could be spelt out in either obstructivist or motivationalist readings. To return to an objection I raised against Battaly in the previous chapter, it is perfectly compatible that vices hold both a motivational and responsibility component. The motivational component can help us understand the badness of the vice. The responsibility component can help us respond to the badness.

favoured but does not always apply to vice, whether that be for reasons due to a lack of control (Battaly 2016a, 2018a) or for prudential reasons (Tanesini 2018, 2021).

For Cassam, it appears that the majority of cases that he considered to be criticisable vices, will be on my view, attributability blameworthy. For example, an agent who possesses the vice of closedmindedness but came to acquire their vice through a lack of educational opportunities, is still responsible for their vice in the attributability sense, as it reflects their character. However, the blame response to this trait should be ameliorative, such as serving to discourage future closedmindedness.

Before concluding, we can briefly return to our earlier, second response to the problem of lack of control, namely that we can just choose to criticise agents for their vices and not blame them (Cassam 2019a). I have demonstrated that attributability responsibility is integral to vices and allows us to blame vice-bearers that Cassam previously could not due to strict control conditions. Importantly, this form of blame is also epistemic and ameliorative, two features that were underdeveloped in Cassam's account. I do contend, however, that with some slight modifications, Cassam's account is not too far removed from my view. We both agree that a form of responsibility is integral to the definition of vice which is ameliorative in aim. If Cassam weakens his control conditions for blame, then it may be possible that this responsibility amounts to a form of blame that is epistemic and ameliorative.[83]

4.7 Conclusion

Throughout this chapter, I have motivated three objectives. Beginning with the first objective, I demonstrated how an epistemic and ameliorative form of blame can be suitable for epistemic vices. I outlined epistemic blame as a response to a violation of some epistemic norm that aims to reduce bad epistemic conduct and bring about epistemic goods (Boult 2021, Fricker 2016; Piovarchy 2021; Sliwa *forthcoming*). I also discussed a variety of ways that blame can be

[83] We can refer to Fricker (2020:105) for a similar claim that the vice-bearers deemed criticisable by Cassam are instead blameworthy but just deserve a different level of blame e.g., criticism. Cassam's (2019a:127-128) view also excludes cognitive defects from being considered epistemic vices, as they are not attributes that we are responsible for. This is also true of my view.

directed towards ameliorative aims, drawing from feminist accounts of responsibility and functional and communicative accounts of blame (Ciurria 2021)

Turning to my second objective, I outlined 'the argument from lack of control', which I labelled the most prominent objection to the possibility of blaming vice-bearers for their vices. This argument was supported by Battaly (2016a, 2018a) Cassam (2019a) and Kidd (2016, 2020), who argued that blame responses would be too strong a response to vices that agents had little control over their acquisition of. Cassam (2019a) was more favourable towards the idea of holding vice-bearers responsible for their actions generally but argued that a weaker form of responsibility, namely criticism, was the most fundamental response to vice. This was also based on arguments surrounding control, specifically that vice-bearers need to have managerial control over their vices to be blamed for it, a condition that is rarely compatible with vice-acquisition.

I outlined a response to this argument with reference to attributability responsibility. This is a form of responsibility applicable to bad traits or attitudes that are reflective of the possessor's character or deep self (Shoemaker 2015; Watson 2004). Crucially, one need not have control over these traits in order to be appropriately blamed for them. I argued that this form of responsibility was best suited to our understanding of vices as bad character traits, attitudes or ways of thinking that are part of our epistemic character. I discussed how both Battaly (2016a, 2018a) and Tanesini (2021) also discussed the favourability of this form of blame for vice, but with Tanesini considering it rarely appropriate.

It followed from these arguments, that blame was not a constitutive feature of vice. This was either because criticism was more fundamental (Cassam 2019), was not applicable to cases where agents did not have control over their vice acquisition (Battaly 2016a, 2018a; Kidd 2016, 2020) or for prudential reasons (Tanesini 2018, 2021). I also responded to Battaly's (2019) 'responsibility problem' concerning the suitability of applying non-voluntary accounts of responsibility to epistemic vices.

Turning to my third and final objective, I argued that attributability responsibility is constitutive to epistemic vice as it necessarily follows from its definition (Taylor 2006:6; Watson 2004:266). I demonstrated how attributability responsibility is integral to the definition and nature of epistemic vice, in the fact that attributability-blame *just is* blame for vice.

In presenting and defending these three objectives, I have illuminated the relationship between blame and epistemic vice. Specifically, I have evaluated what type of responsibility and blame is appropriate for vice and how it is distinctively epistemic and plays an ameliorative role. These arguments provide a foundation for future research surrounding the role and nature of responsibility for our epistemic vices.

CHAPTER 5. EPISTEMIC VICES AND EPISTEMIC NUDGING: A SOLUTION?

5.1 Introduction

'Bad' epistemic behaviour is unfortunately commonplace. Take, for example, those who believe in conspiracy theories, trust untrustworthy news sites, or refuse to take seriously the opinion of their epistemic peers. Sometimes these kinds of behaviours are sporadic or 'out of character', however, more concerning are those cases that display deeply embedded character traits, attitudes and thinking styles (Battaly 2016a, 2018a; Cassam 2016, 2019a; Tanesini 2018, 2021). When this is the case, these character traits, attitudes and thinking styles are identified by vice epistemologists as epistemic or intellectual vices. Considering that these vices often block or subvert the acquisition of epistemic goods such as knowledge or truth, it is important for epistemologists to understand how these kinds of traits can be most effectively mitigated. One currently unexplored way in which we might go about doing so is by employing epistemically paternalistic strategies, particularly the strategy of 'nudging' - the practice of altering an agent's decision-making capacities toward a desired outcome (Thaler and Sunstein 2008:6).

By bringing together two underexplored areas of epistemology yet to be discussed in connection to one another, this chapter will examine whether a specific form of epistemic nudging can be employed as a successful practice to combat our epistemic vices. Despite prima facie appeal, I will argue that epistemic nudging often fails in this aim, consequently amounting to a superficial and short-lived way of addressing epistemic vices. Additionally, I argue that the practice of epistemic nudging can often lead to the creation of further vices or epistemic harms such as epistemic injustice or epistemic laziness (Evan Riley 2017, Ian Kidd 2017). I then consider a weaker form of epistemic nudging, offered by Kengo Miyazono (2023). I examine whether this modified understanding of epistemic nudging proves to be more successful in response to my criticisms. However, I argue that a weakened form of epistemic nudging only leaves itself vulnerable to further criticism.

This debate has important ramifications for the literature on both vice epistemology and epistemic paternalism. If epistemic nudging can assist in the mitigation of epistemic vices, this advances the debate for epistemic nudging and provides a solution to the problem of epistemic vices. However, if my argument is correct and epistemic nudging is not only unsuccessful at mitigating epistemic vices but more concerningly leads to the creation of further epistemic vices, then arguably this is a worrisome objection to the field of epistemic nudging.

5.2 Epistemic Nudging Introduced

Epistemic paternalism is the thesis that in some circumstances we may intervene with the inquiry of another (e.g., to promote certain beliefs or decisions) without consulting them on the issue (Ahlstrom-Vij 2013a, 2013b). This interference is often justified on the grounds that it will result in an epistemic good, such as true belief or the acquisition of knowledge (Bullock 2018; Croce 2020, Goldman 1991).[84]

One way in which someone may interfere with another's ability to conduct inquiry is through the practice known as 'nudging.' The standard and most prominent account of nudging is offered by Richard Thaler and Cass Sunstein (2008). By their definition, a nudge 'alters people's behaviour in a predictable way without forbidding any options or significantly changing their economic incentives' (Thaler and Sunstein 2008:6). Thaler and Sunstein's definition of a nudge implicates some need to change people's behaviour into making better choices. This follows the ideology that assumes that human behaviour is not always rational and can therefore profit from a paternalistic point of view, allowing for institutions to influence an agent's behaviour and to steer their choices in directions which will improve their welfare (Kahneman and Tversky 1972). Liberal paternalism is libertarian as it allows individuals to make their choices without restrictions or barriers imposed by those who nudge them (Sunstein and Thaler 2003). At the same time, it's considered paternalistic because the nudgers consciously aim to guide people towards choices that enhance their well-being (ibid.).[85]

[84] See the following for further discussion of epistemic paternalism (Bernal and Axtell 2020; Bullock 2018; Croce 2018; Dunne 2021; Pritchard 2013; Ridder 2013)
[85] For further information on the distinction between nudging and liberal paternalism, see Chock (2020).

Nudging is further defined by Thaler and Sunstein as a method of liberal paternalism (Thaler and Sunstein 2008:5). It is libertarian in so far as the interventions and strategies of epistemic nudging are employed to guide people's decisions without taking away their freedom of choice and the nudges can be easily resisted. Nudging is 'paternalist' in the sense that it steers people toward one choice as opposed to another, where the individual's choices are interfered with for the individual's own epistemic good.

Nudging can be epistemic when it aims to change one's epistemic behaviour, such as changing one's doxastic attitudes, beliefs, or judgements (Adams and Niker 2021; Grundmann 2021:213; Miyazono 2023:2).[86] For example, we can make people believe certain propositions by rendering those propositions particularly salient or framing them in especially persuasive ways. When an epistemic nudge changes the behaviour for the good, whether for a common epistemic good or to serve the nudgee's epistemic interests, then it is also paternalistic.

Common types of epistemic nudging can include disclosing information, using social norms, reminders, warnings, and informing people of the nature and consequences of past choices (Sunstein 2014). One example of epistemic nudging may be the use of the educational tool of 'lying-to-children' which is viewed as an interference to schoolchildren's inquiry. This educational tool, according to Stewart and Cohen (1997) involves teaching false or incomplete theories to students in order to facilitate a better understanding of the more complex theories. For example, a student might first be taught that Newtonian mechanics provides a complete account of the laws of motion, in order to make it easier for them to learn quantum mechanics.[87]

5.3 Epistemic Vices Introduced

The next section of this chapter will discuss the nature of epistemic vices and lay the groundwork for the exploration of the relationship between epistemic nudges and epistemic vices.

Epistemic vices are consistent, personal qualities that make us bad thinkers 'insofar as they prevent us from acquiring and sharing knowledge, express bad motives and desires, or interfere

[86] Grundmann focuses specifically on *intentional* epistemic nudging, meaning the nudge is unavoidable (2021:211)
[87] This example is borrowed from Bullock (2016:437)

with our individual and collective epistemic functioning' (Kidd et al. 2020:1). One of the central aims of vice epistemology is to guide human inquiry. This aim is often referred to as 'regulative epistemology,' which has the overall aim of improving our epistemic conduct. As Alvin Goldman notes, 'If we wish to raise our intellectual performance, it behoves us to identify those traits which are most in need of improvement' (Goldman 1978:511). The traits that need improvement are those which are of interest to vice epistemologists, and it is in this sense that the study of these character traits attitudes and thinking styles is imperative to the study of human inquiry, and by extension, to epistemology.

While there has yet to be any literature exploring the connection between epistemic vices and epistemic paternalism or nudging, the shared concern for making humans better off epistemically by overcoming certain kinds of characteristic weaknesses in thinking, provides a clear starter for how the two domains overlap. In particular, it seems plausible that if epistemic nudging is considered a successful strategy for making positive shifts to our epistemic behaviour, then epistemic nudging could be an extremely useful tool with regard to the mitigation of certain epistemic vices.

Whether this aim is achievable will be the focus of this chapter. This does not rule out the possibility that epistemic nudging may have other ameliorative effects on epistemic vices, as I will allude to in section 5.5. However, I will be assessing it against this aim based on epistemic nudging's role as a behavioural or attitude modifier, which generates positive epistemic goods (Grundmann 2021; Levy 2019a; Niker 2018).[88]

Of course, if epistemic nudging proves to be successful at mitigating our epistemic vices, this will prove a huge benefit both to the field of epistemic nudging and vice epistemology. The advantage of being able to combat epistemic vices, as already defined by vice epistemologists as serious threats to human inquiry, adds strength to the practice of epistemic nudging and the plausibility of it as an epistemic tool. Additionally, vice epistemologists will have a solution for the mitigation of epistemic vices, which would be of great benefit to epistemic communities. However, if we find that epistemic nudging is unsuccessful in combating

[88] Some may argue that this is too strong an expectation from nudging. However, many other ameliorative solutions to vice also attempt to change the vice-bearer's behaviour by expelling the vice at hand (e.g., Baehr 2011, 2015; Pritchard 2013, 2014).

epistemic vices, and more concerningly, can cause the existence of vices, then this will be a great concern for the practice of epistemic nudging as a useful and acceptable epistemic tool. With this in mind, the remainder of this chapter will explore the connection between epistemic nudging and epistemic vice, focusing on the potential role that epistemic nudging can play with regard to the combating of epistemic vices.

5.3.1 Nudging Epistemic Vices

As epistemic nudging is employed in response to flaws in the human decision-making process, and as epistemic vices are perceived as flaws that we possess in our capacity as inquirers, it is clear to see how prima facie epistemic nudging can appear to successfully combat and mitigate the effects of some of our epistemic vices.[89]

We can discuss the following examples that appear to elucidate this intuition:

Harry, your flatmate, frequently forms false beliefs based on unreliable and untrustworthy news sources from right-wing newspapers and websites. He never gets his news from any other source and is unwilling to do so. Harry's behaviour fits the definition of the epistemic vice of closemindedness in the sense that it is a character trait that obstructs him from gaining knowledge about the true events occurring in the world (Cassam 2019a).[90] In response to Harry's vice of closemindedness, over the course of a month, you decide to nudge him away from forming any more irresponsible beliefs from untrustworthy news sources. Some of the measures you take include offering him a discount for the subscription service for a well-trusted newspaper, warning him about the reliability and trustworthiness of the sources he reads his news from, and leaving neutral, unbiased news programs on TV.[91]

This example satisfies Thaler and Sunstein's definition of nudging. While we are aiming to incline the target toward a subsequent action or outcome, we still leave the nudgee's previously

[89] I will not discuss arguments as to whether epistemic nudging is a successful practice or not overall. The purpose of my chapter is to evaluate, if epistemic nudging is successful can it be used to combat our epistemic vices.
[90] I am aware that the nature of close-mindedness as an epistemic vice is contested (Battaly 2018b). Also, see Chapter 2 of this thesis for a discussion on the systematic clause for vices and the distinction between high and low-fidelity vices.
[91] This example displays the three nudging techniques of default rules, warning and increasing ease and convenience (Sunstein 2014).

salient options on the table (Harry is still able to access whatever news sources he may choose to) and Harry is able to resist our efforts to change his vice if he chooses to do so. It is also important to note here that as epistemic vices are systematic, one-off cases of closemindedness or merely defective instances of epistemic behaviour do not count as epistemic vices (Cassam 2019a). Thereby for epistemic nudging to successfully combat epistemic vice, it must be able to affect one's epistemic character traits.[92] It appears that by using epistemic nudging, we are ridding Harry's character trait (not just a singular case of closemindedness), in the sense that we are employing multiple nudges over a series of time.[93]

Another example that seemingly demonstrates how epistemic nudging can combat displays of epistemic vice concerns a well-debated example of epistemic paternalism which involves responses to the anti-vaccination movement. Vaccine denialists manifest the epistemic vice of dogmatism (when a subject claims to hold a belief or knowledge, which is not based on evidence or supporting reasons) when they seek to avoid engaging in the giving and taking of reasons about vaccines with their paediatricians. For example, vaccine denialists often go to great lengths to find healthcare providers who will not challenge their beliefs. They replace their children's paediatricians with naturopaths, homeopaths, and chiropractors whose training in alternative therapies often makes them predisposed to reject evidence-based forms of medicine and less willing to challenge parents' preconceptions (Ernst 2001:90–93).

In response to this vice of dogmatism, some states in the United States have employed epistemic nudging to put in place strategies to minimize the risks anti-vaccinators pose and to attempt to change their mind about the risk of vaccinations. For example, the Michigan model, (also referred to as the Inconvenience Model) aims to increase the burden of those who choose to exempt children from vaccinations by ensuring that anyone who applies for it must attend education sessions about vaccines at the local public health department. In comparison to 2014 reports, this measure had successfully lowered exemption rates by 39% state-wide and 60% in the Detroit area (Higgins 2016). By employing a default system but allowing the vaccine denialists the option to opt out if they attend education classes, options are still left open to the vaccine denialists, meaning this program successfully counts as a case of epistemic nudging which successfully overcame the vice of dogmatism.

[92] For a discussion on how nudging may affect our epistemic character, see Alfano et al. (2018) and Alfano (2013). However, literature on this topic is scarce and predominately concerns the relationship between epistemic nudging and epistemic virtue as opposed to vice.
[93] As I discuss later, this is referred to as a systematic epistemic nudge (Riley 2017).

From these two examples, it seems plausible that epistemic nudging can be employed to mitigate our epistemic vices, and interventions such as the Michigan model have already successfully assisted in doing so.

5.4 Epistemic Nudging as Insufficient for the Mitigation of Vice

Despite the perceived success of epistemic nudging with regard to the aforementioned examples of vice mitigation, I argue that epistemic nudging often fails in achieving this aim and regularly amounts to a superficial and temporary elimination of bad epistemic behaviour. More concerningly, I also argue that epistemic nudging risks making us worse off epistemically by leading to the creation, not mitigation of epistemic vices (Kidd 2017; Riley 2017).

In this section, I argue that upon closer examination, epistemic nudging predominantly conceals vicious behaviour without altering the vice itself. This argument is rooted in the notion that epistemic nudging is often unable to acknowledge the depth of epistemic vices, which often consist of deep psychological dispositions and heavily embedded social structures (Cassam 2020; Dillon 2012; Kidd 2018, 2020, 2022; Tanesini 2021).

Some accounts of epistemic vice emphasise the connection between vices and wider social structures that can influence how we form and sustain vice. Tanesini (2021), for instance, develops a 'social' account of epistemic vice, which in part examines the social causes that are 'partly responsible' for the formation of some epistemic vices (Tanesini 2021:8). She also argues that oppression, when internalised, can damage our epistemic character, meaning we should pay close attention to how particular structures and power hierarchies can affect our epistemic conduct (ibid.:9-10),

Kidd (2020, 2022) also argues that we should pay closer attention to the wider structural systems that damage our epistemic characters, particularly the conditions that can be epistemically corruptive. Epistemic corruption occurs when 'one's epistemic character comes to be because one's interaction with persons, conditions, processes, doctrines or structures that facilitate the development and exercise of epistemic vices' (Kidd 2020:71). Paying attention to these potentially damaging influences on epistemic vices is referred to by Kidd as 'critical character epistemology' (ibid.:70). This account is modelled off Robin Dillon's critical

character theory, which understands one's character to be affected by 'domination and subordination' (2012:84).

There are at least three ways in which vice epistemologists can be sensitive to these structural and corruptive influences. Firstly, vice epistemologists should pay attention to the question of how epistemic subjects come to acquire or develop vices. This can involve examining aspects of the social environment i.e., intersectional structured social identities, systems of social privilege and power hierarchies or the psychological bases of epistemic vices. It also entails considering the wider sociological, cultural, and ideological conditions that shape our growth, and at times, underdevelopment as epistemic individuals (Cassam 2020; Medina 2012). Secondly, many vice epistemologists are already operating within feminist philosophical frameworks, implying that the emphasis on oppressive systems should apply easily (Fricker 2007; Medina 2012; Tanesini 2016, 2018). Thirdly and arguably most importantly for the purposes of this chapter, vice epistemologists ought to attend to the deeper features of one's character in order to stand an effective chance in reducing the vice at hand (Kidd 2020:70). Attending to the deeper structures that influence and sustain vices will provide the most fruitful solutions to vice.

This paints a picture of epistemic vices as 'deep' qualities of one's character (Cassam, 2019a:133). A virtue, Zagzebski argues, is a 'deep quality of a person, closely identified with her selfhood' (1996:104). Once a virtue or a vice develops, 'it becomes entrenched in a person's character and becomes a kind of second nature' (ibid.:116).

Additionally, Kidd has argued for a 'deep' reading of epistemic vices, which he defines as follows:

Deep Epistemic Vices (DEV): 'a deep conception of epistemic vice is one whose identity and intelligibility is determined by the set of practices, projects, or contexts within which it is embedded' (Kidd 2018:14).

When we conceptualize an epistemic vice through this lens, we are appealing to the 'deeper features' of epistemic vices in order to fully understand their nature. We are acknowledging that the identity of a vice can only be understood in relation to a deeper, underlying conception of human nature, or by appealing to a 'worldview' of the vice. Comparatively, 'shallow'

explanations of vices will only identify the status of a vice by locating it within the practices in which it is typically manifested or the projects of inquiry it obstructs (ibid.)

By discussing the importance of oppressive, social and structural factors and epistemic character, it becomes clear that the identity of epistemic vices should at least be informed by these structures, in the deep and social sense. Taking these considerations into account should illuminate our ameliorative practices and attempts to reduce or eradicate vicious behaviour,

5.4.1 Shallow Epistemic Nudges

Having detailed the depth of epistemic vices, we can now evaluate the efficacy of epistemic nudging as a robust tool for mitigating epistemic vices. I will argue that epistemic nudging is best equipped to deal with momentary and short-lived displays of bad epistemic behaviour and is therefore often not an effective means to combat epistemic vices.

Whilst epistemic nudging may change an individual's behaviour for the better, the scope of nudging is often narrow and focused on immediate decision-making (Grundmann 2021; Riley 2017; Saghai 2013). Because of this, it does not necessarily engage with the broader, 'deep' dispositions that constitute epistemic vice.

On his account of epistemic nudging, Grundmann (2021) observes that epistemic nudges primarily target 'shallow cognitive processes', understood as automatic, non-rational cognitive mechanisms that are not fully deliberative nor reflective (2021:210-212). Epistemic nudging therefore typically influences the behaviour of the nudgee by relying on automatic, non-rational cognitive mechanisms which do not engage the nudgee's fully reflective critical capacities. For example, we can increase the likelihood of a proposition being believed by mentioning it in the first place or persuade people that a particular action was morally wrong by directing their attention to its horrible details (ibid.:213). As Grundmann observes, in these types of cases, 'doxastic states are influenced not by giving any reasons, nor by enforcing them in any direct way, but rather by triggering our biases (salience and framing effects, order effects, affective bias, status quo bias, anchor effects) in smart ways' (ibid.). In short, the emphasis on influencing behaviour through bypassing reflective critical capacities suggests that nudges may not directly address the deep-seated aspects of epistemic vices. Instead, the focus on triggering

biases and cognitive shortcuts implies that epistemic nudging operates at a surface level, modifying superficial and momentary behaviour without delving into the underlying nature of these entrenched vices.

As epistemic nudging predominantly targets shallow cognitive processes, we can tease apart two concerns with our 'deeper' understanding of epistemic vice and the potential effectiveness of epistemic nudging in mitigating them.

Firstly, concentrating on modifying the behaviour of the epistemic vice itself will only go so far in overcoming it. As Kidd (2020) outlined, we should be modifying the deeper structures and communities that sustain vices, something epistemic nudging is not capable of as a behavioural-based solution. Secondly, when epistemic nudging does attempt to modify the vicious behaviour, its impact may be confided to the surface level of the vice e.g. by changing beliefs or influencing decisions. Despite its role as a behavioural modifier, epistemic nudging therefore seems, at best, sufficient to respond to 'shallow' interpretations of epistemic vice.

To elaborate on these claims, we can draw insights from the literature on dispositions (Johnston 1992) to explain this further. Consider a vase which has the disposition of fragility. To prevent it from shattering the vase is carefully wrapped in bubble wrap. By protecting the vase with bubble wrap we have not eliminated its fragility; instead, we have merely masked or concealed it. While packaged in bubble wrap, the vase may not seem fragile, but the disposition still persists. The vase is still made from glass and prone to shattering if damaged and if the bubble wrap were removed then the disposition would still be present.

Applying this analogy to epistemic nudging, it can be argued that nudges address our epistemic vices in a manner akin to how bubble wrap deals with the fragility of a vase. Nudging can ‘mask’ our vices, leaving the deep nature of them still present. It therefore does not alter the vice in any meaningful way, leaving it unchanged. Consequently, in the absence of ongoing epistemic nudges, the vice can resurface, similar to how the fragility of the vase persists when the bubble wrap is removed.

We can now reconsider the previously discussed example of Harry, our closed-minded flatmate. It initially appeared that through epistemic nudging we could combat Harry’s epistemic vice. However, when we are not epistemically nudging Harry, his vice is still present and will continue to manifest itself. This is because epistemic nudging does nothing to change

the vice itself, but only masks it for the duration that various epistemic nudging techniques are employed. The deepness of the vice of close-mindedness, its real identity, is still present and cannot be mitigated through epistemic nudging.

Returning to Grundmann's (2021) account of epistemic nudging, despite its shallow application, Grundmann maintains it is still able to generate epistemic goods such as justified belief or knowledge (2021:212). However, he also contends that epistemic nudging does not seem to generate the same epistemic goods from a virtue-theoretic perspective (ibid.:214).

He presents the example of John, a charismatic political leader who commits a murder. Despite not confessing to the crime, John's guilt is beyond reasonable doubt, and he is sentenced to prison. However, other party members do not consider John to be guilty despite the overwhelming evidence of his guilt. Alicia, the court's public relations manager attempts to persuade these party members of John's guilt. She decides to do so by presenting John's trial in a biased manner, telling of his unsympathetic nature and suspicious behaviour. Her strategy is successful, and most party members end up believing that John is guilty of his offence.

Grundmann then asks, when one party member, Mia, believes that John did committ the murder, could we say that she knows this? We may have reservations as Mia's belief is not based on the evidence of John's guilt. Crucially, for virtue epistemologists, Grundmann observes that Mia's belief is not virtuously formed but instead results from her biases. Accordingly, 'If knowledge requires either the epistemic agent's competent performance (as virtue theories of knowledge claim)…then Mia lacks knowledge in this case' (ibid.). From a virtue perspective, Alicia's nudging does not seem to be successful as it did not result in knowledge. However, Grundmann argues that from a non-virtue theory perspective we can say Mia does possess knowledge, making epistemic nudging a successful strategy in this instance.[94]

From Grundmann's observation then, the epistemic goods that nudging can result in will not be considered 'genuine' to a virtue theorist. This is because the nudgee did not engage in any critical reflection nor reasoning to achieve this knowledge, given that nudging need only engage one's shallow cognitive processes. This raises concerns for the related question as to

[94] Specifically, Grundmann argues that from a safety account of knowledge, we can say that Mia does possess the knowledge that John committed the murder (2021:214-215).

whether nudging can generate real epistemic virtues and also speaks to the to the fact that nudging seems to only engage an agent's shallow cognitive processes.

To summarise, epistemic nudges are limited in their effectiveness as a solution to mitigating vices in so far as they predominantly focus on modifying superficial and momentary behaviour, rather than addressing the 'deep' nature of epistemic vices as qualities entrenched in an individual's character[95].

5.4.2 Epistemic Injustice and Laziness

In this section, I will argue that even if we grant the possibility that epistemic nudging can modify deep epistemic vices, a new worry emerges that it will often change our epistemic character for the worse by creating new epistemic vices in individuals.

Focusing now on systematic epistemic nudges, one might be optimistic that the continuous use of epistemic nudges, even operating at a superficial level, might be enough to combat epistemic vices. For example, we may think that by continuously nudging our flatmate Harry to form open-minded beliefs, over time we can mitigate his vice without doing anything other than influencing his beliefs.

However, a new concern arises with this perspective, which is that systematically bypassing one's critical faculties can give rise to epistemic injustice and epistemic laziness and potentially violate one's intellectual autonomy. Let us turn to discuss this concern.

Riley (2017) argues that nudging is problematic on ethical and epistemic grounds. His criticism is with the 'beneficent nudge programme', which consists of systematic nudges that are 'deployed as a general purpose tool for good' (Riley 2017:598).[96] Riley's main criticism of this programme is that continuous nudging, despite being well-meaning, prevents epistemic agents from developing the capacity to 'reason critically, energetically and otherwise well' (ibid.:604).

[95] A question may arise here as to whether we can just nudge the structures or systems that constitute the root causes of deep epistemic vices. This would be an incorrect application of nudging, however, which is concerned with modifying an individual's epistemic capacities. Much stronger strategies would need to be developed to make changes to corruptive systems (Kidd 2020).

[96] The 'good' or value of nudging can be elucidated in a variety of ways depending on the aim of the nudge, e.g., from helping people make better decisions to mitigating bad behaviour.

Riley argues that most forms of nudging do not engage the nudgee's fully reflective critical capacities, in so far as they focus on activating shallow cognitive processes (ibid.:600). He also argues that bypassing our epistemic capacities in this way amounts to a form of epistemic injustice, specifically known as 'reflective incapacitation injustice' (ibid.). Our epistemic capacities are ones which are properly exercised in cases of knowledge, such as reasoning soundly or trusting a genuinely reliable peer. Riley claims that having the fully developed capacities associated with reasoning critically and energetically is necessary for being a robustly and epistemically developed person (ibid.:604). Accordingly, denying or neglecting to provide people the support, opportunities, or means necessary to develop those capacities, or making it relatively more difficult to develop and exercise those capacities, is unjust (ibid.:605). When our capacity to think critically and reasonably is therefore denied, according to Riley we have a clear case of an epistemic injustice—a reflective incapacitational injustice.

Put simply, Riley's central claim is that being subjected to systematic nudges deprives you of an opportunity to develop the capacity to reason well, in a way that amounts to a form of epistemic injustice. We can consider one of Riley's examples here to highlight this claim.

Riley discusses a beneficial nudge programme which pairs undergraduate students with academic advisors (ibid.:611). The aim of this programme is to assist undergraduate students in planning for their future, particularly in choosing the right course that aligns with their interests and career plans. The students enrolled in this programme have yet to make a timely decision on what course to take, and risk dropping out of university if they do not decide in time. Based on the student's interests, an advisor can continuously nudge students to take a particular course. This constitutes a beneficial nudge, as it aims to serve the student's interests.

However, Riley observes that these nudges make no serious attempt to invite the student to exercise and develop their own capacities as a critical reasoner. These nudges therefore cost the student a 'precious opportunity for the development and exercise of her capacity for critical reason' (ibid.:612). Denying these capacities amounts to a form of epistemic injustice, as the advisor is not providing the student the opportunity to actively foster their capacities for reasoning well.[97]

[97] See Sunstein's (2014) examples of 'Flies', 'Less Drinking' and 'Save More,' which all merit the same criticism.

To summarise Riley's position, systematic epistemic nudges in so far as they often bypass our critical capacities, amount to a form of epistemic injustice. Specifically, denying an individual's opportunity to 'reason critically, energetically and otherwise well' amounts to a form of reflective incapacitational injustice.

Building on Riley's argument, we can take this objection to epistemic nudging one step further and argue that by hindering our epistemic capacities, epistemic nudging can create epistemic character vices, specifically the vice of epistemic laziness. Arguably this vice is created even in cases where epistemic nudging does not just prevent agents from creating certain epistemic capacities as Riley argues, but also in cases where agents already possess those capacities but fail to exercise them as a result of epistemic nudging.

Epistemic laziness is defined 'as a culpable failure to acquire or exercise the epistemic capacities required for enquiry' (Kidd 2017:15). Furthermore, Kidd defines epistemic laziness as a 'capital' epistemic vice, meaning they are the source or 'head' of many further vices. Specifically, 'laziness lies at the root of a whole range of vices characterized by failures to do epistemic work—think of vices like inaccuracy or rigidity, both of which are, ultimately, failures to do the work needed to ensure accuracy or revision of one's beliefs' (ibid.). Capital vices, such as laziness are also corrupting as they increase one's vulnerability to other vices.

One example of how epistemic laziness manifests itself, as outlined by Kidd, is when someone may not care enough about the status of their beliefs in order to put in the epistemic work—they become epistemically lazy. Arguably other instances where the vice of epistemic laziness can manifest is as a result of epistemic nudging. As epistemic laziness occurs when we fail to exercise our epistemic capacities, it follows that any tool that blocks or hinders us from doing so, such as epistemic nudging, can lead to the creation of such vices. Nudging someone towards an accurate belief in a way that does not require critical reflection can result in the nudgee forming the vice of epistemic laziness. This concern is worsened when we consider systematic nudges, which are only more likely to sustain the vice for their continual bypassing of epistemic capacities. Relying on systematic nudges can cause certain epistemic capacities to atrophy, ultimately giving rise to the vice of laziness. Such as muscles are lost over time if they are not exercised, epistemic capacities that remain underutilized due to the reliance on epistemic nudging will also diminish, leading to the creation of epistemic laziness.

When it comes to deploying systematic nudges as a reaction to epistemic vices, there is not just the issue that denying the agent's epistemic capacities amounts to a form of epistemic injustice. We also now have the concern that by preventing these capacities from being exercised, we may even worsen one's epistemic character. This would no longer constitute a beneficial nudge, making it no longer a justifiable solution to mitigate epistemic vices.

5.4.3 Nudging and Autonomy

In so far as systematic epistemic nudging hinders our critical capacities, it can result in epistemic injustice and epistemic laziness. A related concern is that in hindering our critical capacities, systematic nudging may also infringe on our intellectual autonomy (Riley 2017:606).

Whether nudging respects autonomy is widely debated and depends on how autonomy is defined (Battaly 2021; Dworkin 2020; Hausman and Welch 2010; Levy 2019a; Ryan 2016, 2018). For example, if intellectual autonomy is essentially defined as the capacity to reason critically, then it seems clear from Riley's argument that if nudging bypasses this it also bypasses one's intellectual autonomy (Riley 2017:606). This is particularly true if nudging is systematic. However, if autonomy is defined thinly as 'freedom of choice' then nudging may not infringe on autonomy (Sunstein 2014:127).

Likewise, if intellectual autonomy is defined as the capacity to reason critically, then violating it through nudging would be to deny an epistemic virtue (Zagzebski 1996, 2013). If autonomy is respected by epistemic nudging, then the previous concerns on how it can harm our epistemic character still applies. I do not place my stake on either side of the debate, as the debate to whether autonomy is respected by nudging is complex and beyond the scope of this chapter. However, one relevant concern may be that even if systematic epistemic nudging violates autonomy, there may be further reasons to value it, nonetheless.

It may be argued that the potential benefits of nudging can outweigh concerns about autonomy violations. This opens the possibility that all things considered, systematic nudging could be justifiable, as we may value the greater good that it results in more than the autonomy itself. To illustrate, we can consider the 'toxic release inventories' case presented by Thaler and Sunstein (2008:192-193). In 1986, the US government required companies to regularly

disclose information on their toxic waste, making the information available on a public database. This inventory enabled the media and environmental groups to produce an environmental blacklist of companies that released large quantities of waste, with adverse consequences such as stock-price devaluation. This nudge was effective because it significantly increased the costs of polluting, a positive outcome, which arguably outweighed the perceived lack of autonomy over the information that the firms possessed.

This example demonstrates how a systematic nudging approach, despite potential autonomy concerns, could be deemed justifiable when the benefits contribute to a greater good, as seen in the reduction of harmful environmental practices.

We can return to systematic epistemic nudging and apply the same reasoning. The argument goes that even if systematic epistemic nudging hinders our capacities in a way that violates our intellectual autonomy, certain epistemic benefits may outweigh this concern. One such benefit could be the mitigation of certain epistemic vices. For example, if systematic nudging reduced an individual's dogmatism, the perceived benefits of this may outweigh concerns about the lack of autonomy.

However, this argument loses its appeal when considered alongside my previous claim that hindering our epistemic capacities can cultivate vices such as epistemic laziness. If we define intellectual autonomy as the capacity to reason critically, then we are not just weighing the value of autonomy against some other epistemic benefit, but also the damaging consequences that this violation can have on one's character. For example, if systematic nudging violated autonomy, but reduced an individual's dogmatism, there may be an argument to make that the vice should be mitigated. However, now factoring in the threat of epistemic laziness, the trade-off for dogmatism does not seem as enticing. The problem that should be considered, is not just that epistemic nudging may violate our intellectual autonomy, but that in doing so it can create further epistemic vices.

Whether or not bypassing one's epistemic capacities amounts to a violation of intellectual autonomy, the threat of epistemic laziness still remains. The main issue at hand then is that by systematically failing to engage our critical capacities, nudging can make us epistemically worse off, in so far as it can lead to the formation of new epistemic vices or otherwise harm us

in our capacity as a knower. Whether nudging of this kind leads to a violation of intellectual autonomy is up for debate and depends on how we choose to define it.

5.5 Weak Epistemic Paternalism

To summarise, we've outlined two significant challenges to the feasibility of employing epistemic nudges as a viable approach for the effective mitigation of epistemic vice.

The first challenge revolved around the limitation that epistemic nudge's influence on our epistemic vices tends to be superficial, disregarding the inherent complexity of vices as deep, psychological tendencies. Epistemic nudges may therefore predominately change the surface level of undesirable behaviour, but not the vice itself, merely 'masking' it.

The second challenge arises even if we grant that epistemic nudges can have a deeper impact on our vices. It raises concerns that epistemic nudging may have the unintended consequence of epistemic laziness, defined as 'a culpable failure to acquire or exercise the epistemic capacities required for enquiry' (Kidd 2017:16). By excessively relying on nudges to guide our epistemic choices, we hinder our epistemic capacities and potentially diminish our intellectual autonomy in this process. This can result in epistemic laziness.

To address both of these concerns effectively, it is crucial to find a balance within the framework of epistemic nudging. This could entail an account of epistemic nudging that possesses the depth necessary to influence and mitigate epistemic vices, whilst simultaneously respecting the autonomy of an individual's epistemic choices to prevent the risk of intellectual laziness. In this context, I will now assess an account of epistemic nudging, referred to as 'weak epistemic paternalism' (Miyazono 2023:1).

In defining epistemic nudging, Miyazono argues that it can be justified if two conditions are met. Firstly, nudged beliefs are more likely to be true than non-nudged beliefs (referred to as the 'veridicality condition'). Secondly, nudged beliefs are not more likely to be irrational than non-nudged beliefs (termed the 'not-more-irrationality condition'). Miyazono's account also responds to the concerns around interference and autonomy that I introduced previously, which plagues many accounts of epistemic nudging.

I will assess whether this weakened form of epistemic nudging can overcome the objections I have presented, particularly my second concern that epistemic nudging may result in the creation of further vices, such as the vice of epistemic laziness.

5.5.1 Justifying Epistemic Nudging

Miyazono makes a crucial distinction between two different levels of epistemic paternalism. The first is 'weak epistemic paternalism' or epistemic liberal paternalism (ELP). The second is strong epistemic paternalism, also referred to as 'epistemic access paternalism' (EAP). Miyazono discusses an example of a public announcement about vaccinations to highlight the differences between the two approaches (2023:6).

Vaccination Example:

Option 1:

In an EAP approach to promoting vaccination, the emphasis is placed on achieving a certain epistemic outcome which is widespread vaccine acceptance and compliance. To attain this outcome, EAP might recommend restricting people's access to some information about the safety of vaccinations. This could include withholding details on the rare but tragic side-effects associated with vaccinations. The rationale behind this approach is that by withholding information, we prevent the agent from being biased or unduly influenced by information that could potentially defer them from being vaccinated (ibid.).

Option 2:

In contrast, ELP takes a more open and informative approach when promoting vaccinations. ELP works by adopting a positive framing strategy to encourage vaccination uptake. For example, ELP might highlight statistics on how many lives can be saved by the vaccine or emphasise the consensus among medical professionals on its safety and benefits. On this model, ELP respects individual autonomy by refraining from banning or blocking inquiries or

information, unlike EAP. ELP's strategies constitute forms of epistemic nudging as they are self-conscious attempts to move people in epistemic-goods-promoting directions (ibid.:6-7).[98]

Miyazono also distinguishes between epistemic 'incentives' and epistemic nudges. Epistemic incentives, as defined by Miyazono, are mechanisms or strategies aimed at adjusting what is referred to as the 'epistemic choice set' (ibid.:7). This choice set comprises a collection of opinions (or the propositions expressing those opinions) along with the relevant information available to epistemic agents. In the context of the vaccination example, the withholding of specific information about vaccinations (option 1) is an example of an epistemic incentive. By limiting access to certain information, it modifies the epistemic choice set by altering the evidence that individuals have at their disposal when forming beliefs about vaccine safety.[99]

Conversely, epistemic nudges are designed to influence the 'epistemic choice architecture' (ibid.:6) which is composed of various contextual factors that can impact decision-making but are often irrelevant to the core epistemic issues at hand. In the vaccination example, framing information positively (option 2) is an epistemic nudge. It does not directly change the evidence or opinions available to the individuals (the epistemic choice set), but instead, it alters the way that information is presented or the environmental context in which decisions are made. Therefore, by framing information in a certain way, epistemic nudges aim to guide individual's judgements by without fundamentally altering the evidence they possess.

With these distinctions made, Miyazono proposes the following criteria for justifiable epistemic nudging:

The jointly sufficient conditions for an epistemic nudge, N, targeting a nudge, X, to be justifiable are:

(1) Veridicality Condition (VC): X is more likely to form a true belief when X is nudged by N than when X is not.

[98] ELP can take different forms depending on different interpretations of 'epistemic goods'. For example, 'veritism' asserts that truth is the ultimate epistemic good, while other aspects like understanding are only instrumentally valuable. Non-veritism claims that other epistemic goods like understanding can be ultimately good (Pritchard 2013).
[99] Miyazono surveys further examples of epistemic incentives, including Rini's (2017) idea of placing a pop-up tag on dubious social media posts and Levy's (2019b) idea of non-platforming offensive views.

(2) Not-More-Irrationality Condition (NMIC): It is not the case that X is more likely to form an irrational belief when X is nudged by N than when X is not.

(ibid.:11)

Having laid out the relevant distinctions and criteria, we can understand ELP as a weak form of paternalism, recommending weak epistemic nudges as opposed to incentives.

With Miyazono's ELP in hand, we can now turn to revisit our two primary concerns with epistemic nudging as a potential strategy to mitigate vice. First, I argued that epistemic nudge's influence on our epistemic vices tends to be superficial, disregarding the inherent complexity of vices as deep, psychological tendencies. The second objection was that epistemic nudging may cultivate further vices such as epistemic laziness by bypassing one's epistemic capacities and potential intellectual autonomy.

To speak briefly to the first concern, ELP by its nature, relies on weaker, subtle nudges and informational strategies to influence epistemic choices. It is relatively clear then, that this definition of epistemic nudging will fall short when dealing with individuals who possess deeply ingrained epistemic vices. It is therefore still vulnerable to the first objection, that epistemic nudging offers a rather shallow approach to mitigating vices. At best then, this form of nudging is the same as discussed previously, meaning the same objection applies. At worst, it is an even weaker form of epistemic nudging, which will be only met with further criticisms of this sort.

However, the second concern which focuses on the implications of hindering the nudgee's epistemic capacities, may seem less significant on the ELP model. Miyazono argues that one of the advantages of ELP is its minimal interference with the autonomy of individuals engaged in decision-making processes (ibid.:8-9). ELP is 'freedom preserving' because it does not block choices or prevent inquiry. Instead, it primarily involves presenting information in ways that are more likely to bring about a better reaction.[100]

[100] What is 'better' is contextual to the aim of ELP e.g., to encourage an individual to receive a vaccine or to make a healthier food choice.

This perspective may potentially address the concern that epistemic nudging could lead to epistemic vices like epistemic laziness, by violating the vice bearer's epistemic capacities as it has weaker 'interference' conditions for a nudge. As we've seen, with ELP, no inquiries are banned or blocked as certain decisions are just encouraged or 'framed'. Decisions are still made freely and there is no interference with the epistemic agent's freedom of inquiry. The concern then, that laziness could be formed if there is limited decision-making or limited reflection or deliberation, may be avoided. We can also observe that on this account of nudging, intellectual agents still have opportunities to 'exercise' their epistemic capacities of inquiry, and they do not become dependent on the nudges. This seems to further suggest that a weak form of epistemic nudging does not face the challenge that it results in intellectual laziness or epistemic injustice.[101]

However, by seeming to respect our epistemic capacities and autonomy, ELP is committed to a new concern that seems equally as threatening to our intellectual character. This is referred to as the 'irrationality problem', a worrying consequence of ELP that Miyazono admits. His solution to this concern is to admit that epistemic nudging (as now defined via the ELP model) is only applicable in some instances. I will assess whether these instances are cases where individuals exhibit epistemically vicious behaviour.

5.5.2 The Irrationality Problem

Miyazono considers the irrationality problem to be the most concerning threat to ELP. In brief, the concern is that beliefs formed via epistemic nudging are 'epistemically defective' (ibid.:9). This is because the beliefs are formed based on how information is framed or presented, which is irrational.

Miyazono returns to the vaccination example to illustrate this concern. Suppose that an epistemic agent forms a belief about the safety of vaccines, through the influence of epistemic nudging. In accordance with ELP, this belief is formed through the nudge of framing information (option 2). For example, I may form my belief that a vaccine is safe due to the

[101] Of course, whether or not Miyazono's ELP respects autonomy is dependent on how autonomy is defined. At least based on the interpretation discussed by Riley (2017), ELP does not seem to hinder our epistemic capacities. If this turns out that ELP does violate our autonomy too, this only adds to my wider claim that ELP is still unsuitable as a strategy to mitigate epistemic vices.

information that '90% of medical scientists think that vaccines are safe' (ibid.:3). I would not arrive at this conclusion if this same information was framed in a negative way, such as 'Only 10% of medical scientists don not think that vaccines are safe'.

However, the irrationality problem is that forming my belief in such a way is irrational. This irrationality stems from the fact that the belief is influenced by the framing of the information upon which I formed my belief, which is an 'irrelevant contextual factor' (ibid.:9). My belief that vaccines are safe should not depend on how the information was presented to me, nor should it be subject to change because of this i.e., if the information was framed negatively.

Comparatively, other forms of epistemic paternalism do not seem to encounter this problem. For instance, if an agent were to form a belief about the safety of vaccinations where information was restricted from them (option 1), they would be making a rational belief (ibid.:15). This is because the belief is based on the evidence available to the agent at the time, which is a relevant factor. For example, while the agent may not have access to information about every possible side-effect, they can still form a rational belief based on evidence they do possess. Alternatively, they can choose to rationally suspend their judgement if they believe that the available evidence is insufficient.

It seems then, that the consequence of accepting a weakened ELP approach is that it potentially leads to irrational beliefs, a problem that does not arise with EAP or other forms of epistemic nudging.[102]

After surveying different responses to the above concern, Miyazono contends that ELP (and resultingly epistemic nudging) may be justifiable in some situations, where certain conditions are met (ibid.:11). I will examine these cases and see if they are the type that may be deployed to mitigate epistemic vices.

We can now return to the justifiable conditions of epistemic nudging. If these conditions are met, the irrationality problem is avoided, meaning weak epistemic nudging can be deployed.

[102] I believe the EAP decision could also be irrational. Forming a belief based on limited evidence, particularly if that evidence is one-sided as the nudge has intended, may be conclusive of epistemic irrationality or other forms of vice.

The jointly sufficient conditions for an epistemic nudge, N, targeting a nudgee, X, to be justifiable are:

(1) Veridicality Condition (VC): X is more likely to form a true belief when X is nudged by N than when X is not.

(2) Not-More-Irrationality Condition (NMIC): It is not the case that X is more likely to form an irrational belief when X is nudged by N than when X is not.

(ibid.:11)

Continuing with Miyazono's example, VC can be met when an agent, referred to as Ken, is more likely to form a true belief about the safety of vaccinations when they are nudged via option 2, than when they are not (ibid.:12). Ken forms the belief that P when they are nudged, and the belief that Q when they are not nudged, and P is more likely to be true than Q.

If nudging is not effective at all, then Ken's nudged belief that P might just be identical to his non-nudged belief that Q. Alternatively, even if nudging is effective ($P \neq Q$), it is likely that the nudged belief that P is not more likely to be true than the non-nudged belief that Q.[103] NMIC is met when it is *not* the case that Ken is more likely to form an irrational belief about the safety of vaccinations when they are nudged versus when they are not nudged.

It is also not the case that Ken's nudged belief that P is more likely to be irrational than his non-nudged belief that Q. Ken's nudged belief that P is irrational is because it is influenced by the framing of option 2, but his non-nudged belief that Q might also be influenced by the framing of some type. This is because of the non-nudged belief that Q results from *some* information, which is framed in *some* way. In that case, the nudged belief that P and the non-nudged belief that Q are equally influenced by the framing effect, meaning is not the case that the former is more irrational than the latter.

Both of these conditions need to be met for a justifiable epistemic nudge to be deployed. On its own, VC is too weak to justify epistemic nudges. If an epistemic nudge modifies the

[103] Miyazono notes that empirical testing would be needed to prove this possible (2023:3).

contextual factors that influence a decision, but can easily be avoided, this satisfies VC but not NMIC. NMIC is not met as the nudged beliefs are more likely to be irrational than their non-nudged counterparts due to the ease at which the relevant contextual factors can be avoided. Whilst this nudge does have an epistemic advantage (it is more likely to be true than the non-nudged belief) it also has an epistemic disadvantage. This is because the nudged belief is more likely to be irrational than the non-nudged belief (because the relevant contextual factor is easily avoidable).

Again, NMIC by itself is also too weak to justify epistemic nudges. Even if a nudge satisfied NMIC by not making the nudged belief more irrational than the non-nudged belief, it lacks justification as an epistemic paternalist intervention. This is because the nudge, while avoiding an epistemic disadvantage (increased irrationality), also fails to provide an epistemic disadvantage, meaning it does not make the nudged belief more likely to be true than not. For a nudge to be justifiably considered an epistemic paternalist intervention, it should not only avoid making beliefs more irrational but also contribute to the likelihood of some truth (ibid.:13-14).

We then reach Miyazono's claim, that VC and NMIC jointly justify epistemic nudges. A nudge must meet both conditions – it must have an epistemic advantage (the nudged belief should be more likely to be true than the non-nudged belief) and it must not have an epistemic disadvantage – (the nudged belief should not be more likely to be irrational than non-nudged belief).

Considering Miyazono's modification, we arrive at his revised claim that epistemic nudging is justifiable in cases in which both VC and NMIC are satisfied. What does this mean then for instances of epistemic nudging and displays of vice?

We now get the resulting claim that epistemic nudging is only justifiable for displays of vicious behaviour if the nudge will 1) result in a true or accurate belief when X is nudged by N than when X is not and 2) that X is not more likely to form an irrational belief when X is nudged by N than when X is not.

By widening the second claim to not just include irrational beliefs but wider displays of bad epistemic behaviour, we can make the following adjustment. A nudge is justifiable when VC

is met, but also when X is not more likely to display bad epistemic behaviour (such as an epistemic vice) when X is nudged by N than when X is not. If the nudge were to bring about a negative epistemic consequence such as intellectual laziness, which would have not occurred if the nudge was not implemented, then we get the claim that the nudge is not justifiable.

We reach the worry though, that this model of epistemic nudging is now *too* weak to be a particularly useful or impactful tool in many cases, especially concerning its role as a useful tool for vice combat**.** As we have seen, weak nudging should not interfere with the epistemic agent's freedom of inquiry as to do so would be a violation of autonomy. As I argued, it cannot result in other epistemic negative behaviours that could be perceived as displays of vice (closed-mindedness, arrogance, laziness etc.). As Miyazono explained, it also cannot lead to irrational beliefs, so must not consist of purely framing information in a positive light, for example.

I argue that what this leaves is an account of epistemic nudging that is justifiable in very few situations. The worry then, is that in weakening epistemic nudging, Miyazono makes it *too* weak, where it now seems justifiable in only rare cases.

Let's try and apply this form of nudging to our earlier examples and test the results. In the first example, you are trying to nudge your closed-minded flatmate, Harry, away from untrustworthy and unreliable news sources. Some of the measures you take include offering him a discount for the subscription service for a well-trusted newspaper, warning him about the reliability and trustworthiness of the sources he reads his news from, and leaving neutral, unbiased news programs on the TV. These appear to all be instances of what Miyazono referred to as framing techniques, which we have seen lead to irrational beliefs.

In this example, we can grant NMIC can be met; we can say Harry would go on to form irrational beliefs if the nudge was not implemented. We can also grant that Harry's autonomy is not violated if the weaker nudge is applied. However, what we cannot grant with the same confidence is VC, that Harry is more likely to form a true belief when he is nudged by N than when he is not. This is because if Miyazono is correct, framing techniques will lead to an irrational, not true belief. At best we can say that Harry will form a bad belief (either irrational or closed-minded) in both instances of being nudged or not nudged. We also cannot grant that

Harry would not continue acting closed-mindedly if the nudge was not implemented, based on our first criticism that nudges are too superficial to mitigate vices.
In the second example, in response to this vice of dogmatism, we discussed The Michigan model. This aimed to increase the burden of those who chose to exempt children from vaccinations by ensuring that anyone who applies for it must attend education sessions about vaccines at the local public health department. On the weak nudging model, this type of practice seems more aligned with an epistemic incentive than an epistemic nudge. This is because this model modifies the epistemic choice set by altering the evidence that individuals have at their disposal when forming beliefs about not vaccinating their children. It does this by ensuring *more* information is made available.[104]

Consequently, in both examples this weak form of nudging is not applicable. Of course, I cannot consider every example of vicious behaviour that this weak model may be applied to, and further empirical research is needed to know just how successful these and all forms of epistemic nudging can be. However, what we can observe for now is that there is little reason to favour a weak form of epistemic nudging given its limited justification. Even by modifying the conditions of nudging so as to not bring about worse epistemic behaviour, we get the result that weak epistemic nudging is now *too* weak and has limited application.

To summarise, a weaker form of epistemic nudging proposed by Miyazono seems to have the added benefit that it does not infringe on the nudgee's autonomy, meaning it respects the nudgee's epistemic capacities. This is because a weak form of epistemic nudging ensures no inquiries are banned or blocked as certain decisions are just encouraged or 'framed'. Practises that take these measures are better known as 'epistemic incentives' and not nudges. However, the weakened model was left open to the irrationality problem, the objection that weak epistemic nudges that 'frame' information in a positive way to encourage certain decisions, result in irrational beliefs. They are irrational in so far as the belief formed is influenced by the framing of information, which constitutes an 'irrelevant contextual factor' (ibid.:9) in the decision-making process.

[104] I take this to act as an incentive in the same way Miyazono considers Rini's (2017) suggestion of a 'tag' on concerning social media posts a measure against the distribution of fake news. Here the social media user is presented with more information about the post via the tag, such as the source it came from and a link to the content's debunking.

In response to this concern, Miyazono proposed two conditions for a justifiable nudge. The first is the veridicality condition: the nudge must be more likely to lead to a true belief than if the nudge was not deployed. The second is the not-more-irrationality condition: the nudged cannot be more likely to form an irrational belief when nudged versus when they are not nudged. It seemed that the second condition could prevent nudging from contributing to vicious behaviour by widening the claim that epistemic nudging could be not used unless it did not result in bad epistemic behaviour (i.e. vices). However, I argued that by weakening epistemic nudging in such a way, it is rarely applicable. According to arguments made by myself and Miyazono, weak epistemic nudging cannot be implemented in situations where 1) it violates an individual's intellectual autonomy, 2) it results in negative epistemic behaviours that could be perceived as displays of vice, and 3) it is an instance of 'framing' as this leads to irrational beliefs. The few instances that this weak epistemic nudging could be applied, combined with my first objection that epistemic nudging is too superficial an approach to eradicate vices, leads us to the conclusion that epistemic nudging is not a successful practice to combat our epistemic vices.

5.6 Conclusion

To conclude, this chapter has discussed both epistemic nudging and epistemic vices and the potential role that epistemic nudging may have in the mitigation of epistemic vices. Despite initial signs of plausibility concerning the role that epistemic nudging may have in shaping our epistemic character, I objected to this possibility. I argued that epistemic nudging could only mitigate a shallow interpretation of epistemic vices at best and not the accurate understanding of epistemic vices as deep psychological dispositions (Cassam 2019a, Grundmann 2021; Kidd 2020). This shallow reading of epistemic vice meant that epistemic nudging only masked, not mitigated, the existence of epistemic vices.

Furthermore, and more concerningly, I demonstrated how epistemic nudging can result in epistemic injustice and lead to the vice of epistemic laziness by hindering our epistemic capacities (Kidd 2017, Riley 2017). I also discussed how this might amount to a violation of our intellectual autonomy, challenging the assertion that the benefits of attaining epistemic goods outweigh the restriction on one's autonomy. I argued that this position may not hold if the restriction also leads to the proliferation of epistemic vices.

Finally, I evaluated a weaker form of epistemic nudging offered by Miyazono (2023). This account seemed to overcome concerns with autonomy violations and intellectual laziness as the decision-making process of nudgee is respected. However, a new concern arose, which was that this weak epistemic nudging led to the irrationality problem. This was the concern that weak epistemic nudges are inherently irrational. Miyazono conceded this claim but argued that epistemic nudging was applicable in some cases still, namely cases where the nudge must be more likely to lead to a true belief than if the nudge was not deployed and that the nudged cannot be more likely to form an irrational belief when nudged versus when they are not nudged. I argued that these modifications did not overcome my first concern, that epistemic nudging is often too superficial an approach to overcome displays of vicious behaviour. I also argued that it resulted in too many restrictions, meaning there were minimal cases where weak epistemic nudging could be deployed and successful. While it appears that epistemic nudging may not be successful in combating our epistemic vices, we can arguably be generally optimistic in other practices that hint toward this possibility, such as the various methods employed via critical education which aims to cultivate and promote certain epistemic virtues.[105]

[105] See Baehr (2013); Battaly (2015, 2016b); Croce and Pritchard (2022); Kristjánsson (2007); Porter (2016).

CHAPTER 6. EPISTEMIC CORRUPTION AND ONLINE ENVIRONMENTS

6.1 Introduction

Since 2016 there has been widespread interest in the phenomenon of 'fake news', both in the public and academic sphere. This heightened interest is notably attributed to the increasing prevalence of misleading, fabricated or intentionally deceptive information spread through various channels e.g., false narratives spread in a political campaign or misinformation concerning vaccinations (Wardle 2019). A vast amount of literature has been dedicated to this phenomenon in epistemology specifically, with philosophers predominantly focusing on the characterization of the term and the various conceptual issues related to how best to understand it (Coady 2019; de Ridder 2019; Gelfert 2018; Habgood-Coote 2019)

However, despite the array of literature on this topic, there has been significantly less focus on the epistemic harms caused by the phenomena of fake news and how we should understand and ameliorate these harms.[106] I argue that one such example of these harms is those that impact our intellectual character; that is, a result of the influx of fake news in the environment is that certain epistemic vices have been exacerbated and epistemic virtues suppressed. In so far as fake news threatens our intellectual character, I argue it is *epistemically corrupting* – an underexplored but essential form of corruption that is vital to our understanding of the harms and wrongs of fake news (Kidd 2019, 2020, 2021, 2021 et al., 2022).

The plan for this chapter is as follows. First, I will clarify the notion of fake news with reference to the related concept of *information* disorder in the media (Wardle 2019). I will then introduce the notion of epistemic corruption, outlining the various ways that information disorder creates what Kidd refers to as epistemically 'corrupting conditions'. Resulting from these corrupting conditions is the formation of epistemic vice. I will demonstrate how information disorder leads to three distinct epistemic vices – prejudice, conspiracy thinking and epistemic capitulation by following Kidd's (2019, 2020) five modes of corruption - understood as the various (non-exhaustive) ways that a corrupting system can install epistemic vices in its corruptees. Finally, I conclude by examining the various ways information disorder can overcome its corruptive

[106] See Croce and Piazza (2023); Levy (2017) and Rini (2017).

state by assessing individualistic, social, and structural ameliorative solutions. I propose a coordinated approach that encourages the formation of both social and individual virtues (Bland 2022a, 2022b; Sullivan and Alfano 2020)

6.2 Information Disorder

Whilst an almost unlimited access to information may seem as though it would facilitate virtuous inquiry, there are various ways in which our online media environments have tended to damage and harm its consumers, with the increasing awareness of fake news being one such example.[107] However, most of the content we identify as fake news (as both academics and online users) is crucially not often news, nor fake (Allen et al., 2020). For example, content can take the form of rumours, memes, videos, and data, and may often be genuine but just taken out of context. To this end, most of the content we categorize as 'fake news' is better understood more widely as 'information disorder' - which refers to the numerous ways that online information can distort the truth and our trust in the media (Wardle 2019).

Information disorder is an umbrella term used to describe three main types of distortion in the media: disinformation, misinformation and malinformation (Wardle 2019). Disinformation is understood as content that is intentionally false and designed to cause harm.[108] For example, take the claims proposed by Donald Trump and others during Barack Obama's 2008 presidential campaign which denied that Obama held an American passport. This information was intentionally false and designed to cause harm to Obama's presidential campaign, as the claim's falsely asserted Obama was ineligible to be President of the USA as he was not a natural-born citizen.

When disinformation is shared, it often turns into misinformation, which is when false information is spread but no harm is intended (Wardle 2019). Examples of misinformation are often found when reports of terrorist attacks are made on social media, such as the attack on the Champs Elysees in 2017 (UNESCO 2018). Many individuals on social media unwittingly published several rumours about the attack, such as claiming another policeman had been killed. The people sharing this type of content are rarely doing so to cause harm, but rather fail

[107] See the Reuters Institute Digital News report (2020) and Edelman Trust Barometer (2020).
[108] See Simion (2023) for a normative account of disinformation and its challenges.

to adequately inspect the information they are sharing. Finally, malinformation is information that is based on reality and is used to inflict harm on a person, organization, or country. One such example is when Russian agents hacked into emails from the Democratic National Committee and the Hillary Clinton campaign, leaking certain details to the public to damage reputations.[109]

Additionally, within these three types of information disorder, there are seven specific categories that identify the various ways that information disorders manifest. These include fabricated content (e.g., completely false political adverts), manipulated content (cropped photos excluding important information, imposter content (false content that uses well-known news logos), false context (genuine videos reframed in dangerous ways), misleading content (misleading statistics), false connection (clickbait headlines) and satire (parody websites posing as news outlets) (Wardle 2019). These categories identify the more specific ways that information is distorted.

6.3 Epistemic Corruption

The general meaning of corruption is understood as articulating damage or forms of degeneration, and it is usually confined to both moral and political philosophy (Miller 2018). The media, or information disorder more specifically, can be corrupting in a moral and political sense, e.g., discrediting a political candidate in an election to influence public opinion, or psychical harm or death brought about by conspiracy theories.[110]

Alternatively, an underexplored epistemic form of corruption is defined as a process that occurs when '...one's epistemic character comes to be damaged due to one's interaction with persons, conditions, processes, doctrines, or structures that facilitate the development and exercise of epistemic vices' (Kidd 2019:8). Epistemic vices are dispositions, attitudes and ways of thinking that make us bad thinkers, in so far as they prevent us from acquiring and sharing knowledge, manifest bad motives, and desires, or disrupt both individual and collective epistemic

[109] Information disorder is undeniably politically and morally harmful. However, for the purposes of this chapter, I will exclusively focus on the distinct epistemic harms it produces.

[110] One example of political corruption was the creation of a convincing duplicate of the Belgian Newspaper *Le Soir,* which included a false article claiming that the presidential candidate Emmanuel Macron, was being funded by Saudi Arabia (Bakamo 2017). For examples of moral harms caused by information disorder, see Cassam's (2019b) discussion of the deaths caused by anti-vaccination conspiracy theories.

functioning (Kidd et al. 2020). These dispositions (in part) form our epistemic character, which is continuously active and shaped by our surrounding environment. As much goes into shaping our epistemic character (our psychological profiles, our upbringing, social interactions and institutions) and the main purpose of research into epistemic corruption is to identify the relevant sorts of factors that lead to the formation (or increase the presence of) epistemic vices. When an institution or process is found to be corruptive, a 'corruptive criticism' can be directed towards it.

Of course, information disorder will not always lead to the development or exacerbation of epistemic vices, but rather the concern lies in the high risk that it could readily lead to such outcomes.[111] This is an important clarification about the general process of epistemic corruption, which is that it has no success condition. This means that the corruptive process does not need to inflict or promote vices in everyone who is exposed to it. We can understand this by referring to an example 'corruptive criticism' outlined by Duncan Pritchard (2015), which is that the increasing reliance on technology in education enables students to 'offload' cognitive work to external devices.[112] Instead of developing virtues such as attentiveness and insightfulness, a reliance on technology creates conditions conducive to vices such as epistemic laziness.[113] However, not every student who incorporates technology in their learning will develop the vice of epistemic laziness, and it is only required for the corruptive criticism that some do. This highlights the distinction between strong and weak epistemically corrupting processes, with the former relating to ones that entrench multiple vices in many members of its system, and the latter which only entrenches some vices in some of its members. (Kidd 2019, 2020). Similarly, I claim that as long as some consumers of information disorder form vices due to their exposure to it, a corruptive criticism is sufficient.

Relatedly, epistemically corruptive systems do not have to intentionally aim to corrupt or promote epistemic vices. In some cases, the process may be epistemically corruptive even if the system's intended aim is to promote epistemic goods or even epistemic virtues. For example, consider an educational app designed to personalise learning content for students and

[111] See Meyer et al. (2021) for empirical research on epistemic vices and misinformation.
[112] This is not to say that all instances of cognitive offloading will be corruptive. For example, if cognitive offloading is viewed as an expansion of one's mind rather than a mere delegation to technology, it may not be corruptive.
[113] Pritchard (2015) argues that technology can hinder the creation of favourable conditions for developing epistemic virtues. This hindrance can manifest by limiting the opportunities to develop virtues and improve one's epistemic character, or by creating conditions conducive to vices. For further discussion of Pritchard's argument in the context of epistemic corruption, see Kidd (2019:228).

help them understand complex subjects. Despite its good intention to promote learning, the app presents information that aligns with the student's existing knowledge without challenging them, it may inadvertently create an environment that hinders critical thinking.[114] Bringing the discussion back to the media and information disorder, we can see that the aims and intentions of information disorder can be varied. For example, malinformation is spread with a direct intention to cause harm e.g. to undermine trust in an organization (UNESCO 2018) or to harass and discredit journalists and critical reporting (Ireton and Posetti 2013).[115] Alternatively, misinformation is not shared with the intention to cause harm, but through inadequate fact-checking or fearmongering e.g., the various COVID-19 conspiracy theories which are increasingly circulated on social media (Allington et al. 2020:176).

In sum, whether information is direct or indirect, information disorder poses a threat to reliable and effective inquiry through its damage to our intellectual character by promoting various epistemic vices in its members. To this end, it is epistemically corrupting.

Using the five modes of corruption of what constitutes an epistemically corrupting system, I will demonstrate how the media, via information disorder, can be epistemically corrupting to one's character and can lead to a subset of epistemic vices.

We can begin by acknowledging some details on the definition of epistemic corruption itself. Following Kidd's (2019, 2020) distinction, we can interpret the damage or harm that epistemic corruption causes to our character in two ways. Firstly, there is active corruption which is damaging in so far as promotes or rewards the exercise of vice, and deteriorates pre-existing virtues already present in the subject's character. Secondly, there is passive corruption, which is damaging as it fails to effectively facilitate or encourage the exercise of virtue. We can also characterize the difference between the corrupted - the person or thing being corrupted, and the corruptors - the persons or things facilitating the corrupting.[116] Applied to information disorder, who or what is doing the corrupting at a general level is often the institution or individual that is creating or sharing the false information. This could be a media outlet publishing false news

[114] Kidd argues that the strength of the corrupting conditions depends on two factors: first the psychological profile of the agents being subjected to the corrupting conditions and second the structure/norms of the system itself. In this sense, some agents may be personally resilient to the acquisition of vices, or the system itself takes steps to reduce its corrupting tendencies e.g., a media outlet that employs fact-checkers.

[115] This type of harassment is disproportionately experienced by women and is frequently misogynistic in nature (Bartlett J. et al. 2014).

[116] Kidd (2020:71) notes that sometimes the corruptor and corruptee can be the same since some subjects are complicit in their epistemic self-corruption.

stories in order to harm a presidential candidate (disinformation), or a social media user sharing a parody article as if it were real (misinformation). In these given examples, the corruptor is the media outlet and the social media user, and whoever becomes exposed to these articles is the corruptee.[117]

There are two main modes of facilitation – material conditions and motivational conditions, which give way to five (un-exhaustive) modes of epistemic corruption (Kidd 2020:72-73): [118]

(1) Acquisition: a corruptor can facilitate the development of new epistemic vices that were not previously a feature of the subject's epistemic character.
(2) Activation: a corruptor can activate dormant epistemic vices which are usually latent or inactive within the subject's epistemic character.
(3) Propagation: a corruptor can increase the scope of a vice, influencing the degree to which it affects the range of the subject's character.
(4) Stabilization: a corruptor can also increase the stability of a vice, diminishing the likelihood of the vice's being easily disrupted or altered
(5) Intensification: a corruptor can further increase the strength of a vice.

These five modes of epistemic corruption are just some of the ways that an epistemic individual could become epistemically vicious as a result of their interactions with corruptors. In the next section of this chapter, I will use these five modes of corruption to examine how epistemic agents can become epistemically vicious as a result of their interaction with the media, via information disorder.

6.4 Identifying the Epistemic Vices

The next step of the corruption criticism is to identify the vice(s) that the process, person, institution or so forth is promoting or drawing out. I will argue that information disorder is

[117] Other examples of a 'corruptor' can include organizations, policies and individual or collective agents (see Baird and Calvard 2018 on how businesses can create and sustain epistemic vices).
[118] I use 'facilitate' in the same broad sense as Kidd, e.g., online environments can 'promote', 'encourage' or 'provide conditions for' the development and exercise of vice (2020:71).

corrupting in so far as it gives way to three unique vices in accordance with the five modes of epistemic vice-corruption (Kidd 2019, 2020).

6.4.1 Prejudice

The first epistemic vice which is promoted by information disorder is the vice of prejudice. Prejudice is understood as a '…a negatively charged, materially false, stereotype targeting some social group and, derivatively, the individuals that comprise this group' (Begby 2013:90). According to Miranda Fricker, prejudice can be defined in a slightly stronger sense, as '...a judgement made or maintained without proper regard to the evidence' (2007:33). Turning to prejudice's vicious nature, Fricker has extensively discussed the vicious nature of the vice of prejudice with regards to epistemic injustice. One type of epistemic injustice is identified by Fricker as a *testimonial injustice* – when a speaker's testimony is not believed due to the hearer's prejudice of the speaker (e.g., a racial bias). An example of this injustice which Fricker discusses is the trial of Tom Robinson in Harper Lee's novel To Kill a Mockingbird, where Robinson, a black man, is falsely accused of beating and sexually assaulting Mayella Ewell, a white woman (ibid). At trial, he testifies his innocence and presents evidence which proves he could not have committed the alleged acts. However, the white jury has their judgement radically distorted by the effects of racial prejudice, resulting in Robinson's wrongful conviction. As the credibility of Robison's testimony is discredited because of the jury's racial prejudices, Fricker identifies this as a case of testimonial injustice, which is a specific form of epistemic injustice.[119]

Additionally, Quassim Cassam (2019a) discusses the epistemic vice of prejudice in some detail which he defines as an 'attitude vice' (ibid.:87-88). Building on Fricker's definition, Cassam defines the epistemic nature of prejudice in so far as it is an affective posture towards another individual's epistemic credentials and implies a negative attitude (ibid.:88). Attitudes are stances or postures rather than character traits, and Cassam defines prejudice as a posture. Postures can be affective and involuntary, and an epistemic posture is one that is aimed at an epistemic object, such as evidence. Cassam notes that prejudice is an attitude formed and

[119] According to Battaly (2017a), testimonial injustice is vicious because it consistently produces bad epistemic effects e.g., it impedes the transmission of knowledge. Fricker (2007) also discusses the virtue of testimonial justice as an individual remedy to testimonial injustice, which gives us further reason to define testimonial injustice as a vice.

sustained without proper consideration of the merits/demerits of its object. Cases of testimonial injustice are examples of epistemic prejudice as they are cases where one displays an epistemic attitude which is an affective posture towards another person's epistemic credential. Moreover, prejudice is vicious in that it prevents the gaining and sharing of knowledge and is at least reprehensible.[120]

In accordance with the fifth mode of corruption, we can see how epistemic prejudice is intensified by information disorder through considering a case of testimonial injustice. Take the various reports following the latest BLM protests, for example, which falsely report or exaggerate incidents of violence during the protests (Beckett 2020). The exaggerations are frequently displayed in the headlines of articles which follow recent BLM protests, such as The Washington Post's 'A Night of Fire and Fury across America as Protests Intensify' and the New York Times' 'Appeals for Calm as Sprawling Protests Threaten to Spiral Out of Control'. However, these reports contradicted the true peaceful nature of the protests, with evidence demonstrating that more than 93% of BLM protests involved no serious harm to people or damage to property.[121]

Additionally, media outlets have often shared false information which added to the narrative of BLM protests as violent. Examples include viral videos of seemingly 'innocent bystanders' being beaten by protesters, shared by the likes of the Daily Mail and then-president Donald Trump. Other examples include reports that protestors had hijacked a train, set fire to supermarkets, and police officer deaths.

By presenting BLM protests (and the movement) as threatening and dangerous through forms of malinformation, the peaceful message and credibility of the protests are suppressed and delegitimized.[122] Additional research supports the claim that false depictions of BLM are based on racial stereotyping. As Smiley and Fakunle (2017) note, 'misconceptions and prejudices (are) manufactured and disseminated through various channels such as the media, including references to a "*brute*" image of Black males' (ibid.:350). They highlight that the racist

[120] According to different accounts of epistemic vice, prejudice may be epistemically vicious because it expresses bad motives or impedes epistemic functioning in some other way (Kidd et al. 2020:1). Whether it is reprehensible or blameworthy is also debatable (see Chapters 2-4 of this thesis).
[121] Armed Conflict Location and Event Data Project (2020)
[122] This claim is supported by a study that found that negative media coverage of the BLM protests demonized and delegitimized the protestors (Leopold and Bell 2017).

depiction of black people in the media as ‘brutish’ or ‘thuggish’ creates various prejudices and misconceptions (ibid.)

We can understand how this is a case of testimonial injustice, in so far as the media’s prejudice and stereotyping of BLM protestors reduces the credibility of the movement through false depictions of the protestors as dangerous and violent. In turn, this creates further prejudice in the consumers of these news reports, which demonstrates how the vice can be intensified by these false reports.[123]

6.4.2 Conspiratorial Thinking

A second vice which is promoted by information disorder, is the vice of conspiratorial thinking. Whilst initially categorised as a vicious character trait by Cassam (2015) and a thinking vice (2019a), more recently he has defined conspiratorial thinking as an ideology (2019b). Defining an ideology as a set of fundamental ideas and beliefs that shape one’s understanding of the world, Cassam rules out a conspiracy mindset being defined as a character trait, as character traits are not generally understood as ideas or beliefs. Cassam also remarks that a person who subscribes to certain ideologies might be more inclined to endorse ideologically motivated conspiracy theories. In such cases, it is the person’s ideology rather than their epistemic vices which is key to their thinking (2019b:48).[124]

Despite revising his definition of conspiratorial thinking as a personality trait, there is still reason to regard conspiratorial thinking as epistemically vicious, by appealing to Cassam’s categorization of it as a ‘thinking vice’ (2019a:71).[125]

According to Cassam, a thinking vice is an epistemically vicious way of thinking (i.e., a way of thinking that obstructs knowledge) which concerns the qualities of a person's thinking, rather

[123] Additionally, a report by UNESCO (2018) found that many people choose to engage with erroneous information that reinforces their prejudices, in preference to engaging with accurate, credible content that may challenge them to shift their opinions.

[124] For non-vice theoretic perspectives that detail the epistemic harms of conspiracy thinking, see Coady (2012); Harris (2018); Peters (2020); Prooijen and Douglas (2018) and Uscinski (2018). For discussion on what leads one to believe conspiracy theories see Brooks (2023); Napolitano and Reuter (2021) and Shields (2022)

[125] Arguably there are no reasons to assume Cassam’s dismissal of conspiratorial thinking as a personality trait would exclusively lead to the conclusion that it cannot be classified as a thinking vice. Cassam’s (2019b) modification seems primarily directed at his (2015) classification of conspiratorial thinking as a personality trait.

than the qualities of them as a thinker (ibid.:56). For example, there is a distinction between *being* closed-minded and *thinking* closed-mindedly - these two are related but distinct.

Cassam appeals to Baron's (1985) definition of thinking as a 'method of choosing among potential possibilities' (ibid.:90) and states that ways of thinking are different ways of doing this e.g. the different ways of including or excluding possibilities and drawing conclusions. When these activities get in the way of knowledge, they become potential epistemic vices. Cassam presents the example of a gambler who after watching a succession of coin tosses land on heads, thinks that the next toss will be tails because a tails is now due. This is an example of superstitious thinking, thinking that posits a causal link between unconnected events (namely, previous coin tosses and the next toss) and leads the gambler to draw a conclusion that does not follow from his premises.

Turning his attention to the vice of conspiratorial thinking, Cassam maintains that some of the vice's characteristics include attempts to 'tie together seemingly unrelated events and focuses on errant data' (Cassam 2019a:70). Studies have also found that belief in conspiracy theories is associated with superstitious and paranormal beliefs in that they are 'underpinned by similar thinking styles' (Swami et al. 2011).

Conspiratorial thinking also plausibly leads to an obstruction of knowledge and other epistemic goods. Returning to Cassam's general discussion of thinking vices, he claims that thinking vices create systematic errors as they make us more error-prone. Whilst this may not always be the case, he argues that often enough thinking vices are not 'a reliable pathway to true belief' (2019a:67). Another way that thinking vices can obstruct knowledge through conspiratorial thinking is by reducing the epistemic agent's confidence in their beliefs. Discussing the confidence condition for knowledge, Cassam states that for one to know the truth of proposition P one must be reasonably confident that P and possess the right to be confident in P. However, thinking vices can undermine an agent's right to be confident in P, as the confidence in P is not based on evidence, but bad ways of thinking. The agent would only have the right to be confident if their belief was reasonable and formed using a reliable method.

Applying this to the vice of conspiratorial thinking, for an epistemic agent to be confident in the conspiracy theory that the 9/11 attacks were an inside government job, their confidence in this belief must be based on reliable evidence. However, as we have seen, conspiratorial

thinking often involves bad epistemic behaviour, such as drawing false conclusions from unreliable premises.

In order to determine whether thinking vices such as conspiratorial thinking are only conditionally vicious, Cassam distinguishes between 'thinking vices proper' and 'vicious thinking' (ibid.:76). A specific case of conspiratorial thinking may or may not be vicious – as thinking vices can be conditional but thinking vices proper are not. This means that conspiratorial thinking can only be vicious if it displays thinking vices *proper*, such as closed-mindedness or gullibility (ibid.:74). This means that it is not the context or environment which determines whether conspiratorial thinking is vicious or virtuous, but it's whether the case of thinking displays thinking vices proper. To add to this claim, Cassam states that it is more obvious that thinking vices like conspiratorial thinking are not reliable pathways to true belief or that they are systematically conducive to true belief. It is in this sense that thinking vices, and conspiratorial thinking, are epistemically vicious (ibid.:67).

Returning to information disorder, conspiracy theories often begin as a form of disinformation designed to provoke or manipulate people for financial or political reasons (UNESCO 2018). For example, consider the various instances of disinformation spread which aimed to undermine the seriousness of COVID-19.[126] These conspiracy theories ranged from claims made by pro-Kremlin media outlets that the pandemic was a hoax, to the claims made by then the US President Donald Trump in support of the conspiracy theory that the virus originated in a Chinese laboratory. When these theories are spread, they turn into misinformation, as many individuals on social media unwittingly reshare conspiracy theories due to misunderstanding and panic.

Bringing the discussion back to the five modes of epistemic corruption, we can see how conspiracy theories could increase the *scope* of the vice of conspiracy thinking in accordance with mode three. For example, consider an individual who believes the conspiracy theory that the MMR vaccine causes autism in children who are vaccinated with it. The individual then comes across a new conspiracy theory which claims that the COVID-19 vaccination is part of a secret plan by Bill Gates to implant trackable microchips into those who are vaccinated with it. Prior to this conspiracy theory, the individual only possessed a conspiracy mindset about the

[126] See Barua et al. (2020) and Gerbina (2021).

health implications of vaccinations, but now they believe there are a number of privacy concerns surrounding vaccinations too.[127] It seems clear that the scope of the vice, the conspiracy thinking, has been propagated by the presence of information disorder in the form of conspiracy theories.

6.4.3 Epistemic Capitulation

We can now turn our attention to a third epistemic vice which is brought about by information disorder. Unlike the two previously identified vices where agents are unaware that the information they are consuming is disordered, these agents recognize that they are being exposed to false information. Despite initially seeming that these agents' intellectual characters would be unaffected by information disorder, I argue that in fact a further distinct epistemic vice could still be brought about - the vice of epistemic capitulation, defined as the deficiency vice of the virtue of intellectual perseverance (Battaly 2017b). Agents who possess this vice often fail to act when they encounter an obstacle or difficulty e.g., being too quick to quit a research project or not re-take a test.

The character virtue of intellectual perseverance requires a disposition to make good judgements about goals, specifically about which intellectual goals are appropriate for one to pursue, and when (Battaly 2017b). Accordingly, the vice of capitulation concerns an agent's ability to make good judgements about their epistemic goals (ibid.:670). For example, an agent who gives up every time they encounter a difficulty fails to recognise that their goals are still attainable despite the obstacles.[128] Conversely, in the opposing direction, agents who stick with goals that should have been abandoned months ago possess the vice of recalcitrance - the excess vice of intellectual perseverance. In both of these displays of vice, an agent's failure to act is rooted in their failure to exercise good judgement about goals.

The vice of capitulation is also closely related to the vice of epistemic laziness, which is understood as the culpable failure to acquire or exercise the epistemic capacities required for

[127] This claim is supported by research that conspiracy believers who endorse one conspiracy theory tend to endorse others (Cassam 2019b; Wood et al. 2012)
[128] Battaly further defines the virtues of intellectual courage and self-control as types of intellectual perseverance. Other vices such as cowardice, apathy and procrastination may also fall under the deficiency vice of capitulation and the excess vice of recalcitrance (Battaly 2017:695).

enquiry. An agent can exhibit epistemic laziness when they exhibit a ‘failure to acquire or exercise the epistemic capacities required for enquiry'. (Kidd 2017:15). This may be because they do not care enough about the epistemic good e.g., the truth of their beliefs or the value of knowledge. Arguably, whilst viciously capitulated agents may care about the epistemic good to be pursuing it in the first instance, they seemingly do not care as much as their counterparts in possession of the virtue of intellectual perseverance, who value the good enough to continue to act despite the obstacles they encounter.

With reference to the five modes of epistemic vice-corruption, information disorder can be said to enable the *acquisition* of the vice of epistemic capitulation, which means it can create a vice that was once not previously a feature of the subject’s epistemic character.

During certain events or campaigns, information disorder increases, which in turn presents epistemic agents with an overwhelming excess of information that must be laboriously trawled through to find trustworthy sources or accurate reports. Take, for example, the so-called ‘disinfodemic’ of COVID-19 which has led to a worrying increase of COVID-related information disorder. Content ranges from false information about the origins of the virus to false mortality rates and is defined by the World Health Organization as an ‘over-abundance of information – some accurate and some not – that makes it hard for people to find trustworthy sources and reliable guidance when they need it.’ (Novel Coronavirus Situation Report 2020).

On a day-to-day basis, an epistemically virtuous inquirer will exhibit the previously discussed virtues of good inquiry e.g., fact-checking evidence and cross-referencing sources. However, a polluted epistemic environment such as those created by the increase in COVID-19 false information has the potential to lead one to develop vices such as epistemic capitulation. For example, once virtuous epistemic agents who previously assessed the truth of all the relevant information that they came across related to COVID-19 might become overwhelmed by the false data, fake cures and vaccine conspiracies and decide to stop fact-checking and just trust the sources they next encounter.

In this sense, the presence of information disorder and the sheer volume of it prevents an obstacle to epistemic agents, which in turn can directly enable the acquisition of epistemic capitulation which was once not a previous feature of the subject’s epistemic character.

6.5 Explaining the Corrupting condition(s)

To summarize this chapter so far, we have identified three distinct epistemic vices that information disorder promotes - epistemic prejudice, conspiratorial thinking, and epistemic capitulation. We have identified these vices in accordance with the five modes of epistemic corruption – the various ways that vices are brought about by corruptive systems. In the next section of this chapter, we can turn our attention to the types of conditions which bring about these identified vices to further understand how they are formed via the presence of information disorder.

Firstly, conditions can be corrupting if they increase the exercise costs of virtues and make them harder to attain, which in turn makes the path to vice much easier and more attractive (Kidd 2019, 2020). While increasing the exercise costs does not automatically make the individual more vicious, it does hinder their capacity to pursue epistemic virtues. In order to demonstrate how this condition is present regarding information disorder, we must first say something briefly about virtue acquisition. The epistemic virtues I will focus on are related to the virtues of good inquiry and critical thinking, two of which include open-mindedness and attentiveness (Zagzebski 1996). Virtuous open-mindedness requires an agent to be considerate of alternative views, and the ability to change or revise their beliefs (Zagzebski 1996). Virtuously attentive agents are observant and pay close attention to the task at hand, focusing on important details that they process in an adequate way (Baehr 2015).

Arguably, both open-mindedness and attentiveness are made harder to obtain by the amount of epistemic labour one must do to acquire and exercise these virtues when it comes to information disorder. Here we can allude to extensive research on the effect that the media (particularly digital media) has on our ability to assess and acquire knowledge, with various studies pointing to 'information excess' as one of the main factors contributing to the challenges of knowledge acquisition in an online environment. For example, finding and extracting information that is credible is more difficult in a fast-moving informational environment, as decision-making requires reflection which is in turn timely (Dahlgren 2018).[129] Additionally, Dahlgren notes that our 'cognitive certainty' – the assurance and confidence that individuals have in their

[129]Additionally, research has shown that fake news spreads more quickly and reaches more people than accurate news stories (Meyer 2019).

understanding of the world - is also threatened by information excess, as we are more likely to be sceptical of information we're presented with online (2018:22).

Whilst a degree of scepticism and scrutiny is of course beneficial when it comes to information disorder, Dahlgren notes that the extent of the competing versions of knowledge creates an excess of doubt, which generates cynicism. Additionally, the sheer volume of information that needs to be trawled through and fact-checked is epistemically overwhelming (as we have seen with regards to our previous discussion on epistemic capitulation) which in addition fosters an overly sceptical mindset to the information we wish to attain.

In this sense, the overload of information online and the fast-paced nature in which it spreads makes it harder for epistemic agents to exercise their virtues. In particular, the fast-paced environment in which information disorder presents itself, demands a high level of attention from epistemic agents and resistance to scepticism, which makes the virtues of attentiveness and open-mindedness much harder to exercise or attain.

A further way that information disorder exemplifies corrupting conditions is by rebranding vices as virtues. In response to identifying examples of epistemic corruption, one could implement countermeasures to try and disguise or conceal the corruption, and one such attempt is 'rebranding' vices as virtues. For example, arrogance is often 'rebranded' as confidence, and dogmatism is rebranded as tenacity (Dillon 2012).

To see how this is also present in information disorder, we can refer to our earlier discussion of the vice of conspiratorial thinking. Some conspiracy theorists may argue that their conspiratorial thinking is actually a form of open-mindedness or intellectual courage. Conspiracy theorists and consumers of conspiracies could therefore be led to believe that they are the open-minded 'truth seekers' discovering previously covered-up truths about the government which they aim to expose. In this sense, the viciousness of conspiratorial thinking is rebranded as virtuous, which conceals the vicious corruption at play.

Finally, information disorder exemplifies corrupting conditions in so far as it encourages the exercise of vice. To illustrate this, Kidd (2020:76) presents an example of this corrupting condition which we can understand as a form of information disorder. The example considered is that of a doctor who works for a tobacco company that incentivizes acts of dishonesty by

financially rewarding its staff if they publish articles insincerely questioning the connection between smoking and various diseases. More generally, this is an example of 'sponsorship bias', where the outcome of a scientific study is significantly distorted in order to align with the sponsor of the study's financial interests (Reutlinger 2020:15).

This example, and many other cases of sponsorship biases, are often examples of disinformation as the research is deliberate, misleading content, designed to actively disinform the intended audience. By encouraging and facilitating vicious behaviour, information disorder exemplifies a further corrupting condition.

6.6 Conditionality and Corrective Claim(s)

As argued for by many vice-epistemologists, (Battaly 2013, 2016b; Kidd 2020, 2022; Medina 2021:119-123; Tanesini 2021:193-205) solutions or corrective measures which are directed towards vice should be both individualistic (directed at the vicious individual) and structural (directed at the vice promoting institution and the wider practises that allow for vices to develop).

In keeping with the wider aim of vice epistemology to identify effective ways to respond to epistemic vice, we can now turn to discuss the ameliorative functions of our corruption criticism. A corruption criticism can possess both conditionality claims and corrective claims. The former describes the conditions that must be in place for corruptive tendencies to be possible and can include certain '…aims or practices or cultures that enable, incentivise, or in some other way encourage the development and exercise of vices' (Kidd 2019:227). If these features can be removed, then the system is contingently corrupting. Alternatively, when corrupting conditions are intrinsically corrupting, the features are integral to the system – the corruption could only be eradicated if the system was dismantled altogether.

The aim of identifying conditionality conditions is to locate the causes of corruption, which in turn allows for solutions or 'corrective claims' to be made. These are ways of identifying the features that need removing or modifying and the edifying features that should be included or enhanced. Turning our attention back to the media and the various forms of information disorder, we can identify it as conditionally corruptive, as it is information disorder itself that

exemplifies the corruptive conditions, not the entire media system. Following this, we can conclude by discussing some 'corrective claims' which aim to point to possible solutions to dismantle information disorder.

The final section of this essay will therefore outline some corrective strategies which could assist in combatting the vices brought about by information disorder. I categorise these strategies into three broad approaches: individualistic, structural, and social. Individualistic approaches are solutions or strategies aimed at addressing epistemic vices which focus on the individual and the vice itself. Structural approaches, on the other hand, entail solutions or strategies centred on the systems and structures that instigate or perpetuate vices. Finally, social ameliorative solutions focus on the impact that epistemic vices can have on the vice-bearer and social communities.[130]

6.6.1 Individualistic Approaches

Much of the material on the ameliorate aims of vice epistemology is individualistic and virtue-centric. This means it focuses on educating individuals on their epistemic vices and what epistemic virtues are needed to combat such vices. For example, both Jason Baehr (2011, 2015) and Duncan Pritchard (2013, 2014) have identified various intellectual virtues which should be fostered and promoted for effective cognitive inquiry. 'Educating' can take various forms, from learning from explicit instruction, providing opportunities for virtue habituation and exposure to exemplars or role models (Baehr 2013; Battaly 2013; Croce and Pritchard 2022; Kristjánsson 2007; Porter 2016).

An individualistic approach could be helpful to flesh out a corrective aim for the vices associated with information disorder. If information disorder is establishing or promoting intellectual vices, then seemingly one way to combat this is to establish and promote intellectual virtues within individuals to counter or eradicate the vice. Let's consider three potential intellectual virtues which can assist in countering the vices that information disorder instils and promotes.

[130] See Kidd et al. (2020:12-13, 2022:84-85) for a discussion of the distinction between individualistic and structural approaches.

The first such virtue is intellectual perseverance, defined as the continued effort in one's pursuit of intellectual goods despite facing difficulties (King 2014). Exercising this virtue in the corruptive conditions of information disorder would require agents to remain engaged and motivated despite the array of false information they could come across. In this respect, this virtue is particularly important in combatting the previously identified vice of epistemic capitulation. If agents possessed the virtue of perseverance instead, they would continue to exercise virtuous inquiry despite the overload of disordered information which increases their epistemic labour.

The second virtue is open-mindedness, defined as a willingness to consider alternative views, and the ability to revise one's beliefs in light of new evidence (Zagzebski 1996). This virtue closely counters the vice of intellectual prejudice, which we saw consisted of making and maintaining false judgements without proper regard for the evidence. By educating for open-mindedness, agents who are disposed to false information which supports their prejudice and intensifies it, instead learn to update and revise their beliefs, by considering alternative accurate information and viewpoints.

The final virtue to be discussed in this section is intellectual carefulness, which is understood as the trait of being cautious and avoiding intellectual errors. Whilst there is some overlap between this virtue and others, intellectual carefulness seems most apt to counter the vice of conspiratorial thinking which agents display when they focus on errant data, draw false or exaggerated connections between unconnected events and jump to false conclusions (Cassam 2019a:56). When aptly discussing this virtue as a 'virtue of the internet' Richard Heersmink (2018:3) notes that being able to avoid mistakes requires an awareness of situations where common mistakes are made, meaning a basic understanding of logic and critical thinking skills is necessary. When exploring how one can educate for virtues such as intellectual carefulness, Heersmink discusses Heather Battaly's (2016b) suggestion that intellectual virtues can be taught as part of undergraduate courses in logic or critical thinking and proposes this include internet literacy skills.[131] Therefore, by educating for virtues like intellectual carefulness, and

[131] Heersmink (2018) also discusses various other ways that online epistemic virtues can be fostered. Additionally, promoting media literacy skills has also been proposed by UNESCO in their (2021) report 'Disinfodemic: Dissecting responses to COVID-19 disinformation. The report states that audiences are often open to learning how to 'inoculate' themselves against disinformation, which can further provide support from 'inoculation' in the form of promoting these virtues). See also Baehr (2013).

by including skills related to internet literacy, vices such as conspiratorial thinking could be diminished as agents learn how to exercise virtuous inquiry in light of information disorder.[132]

6.6.2 A Structural Approach

Whilst it may seem commonsensical that individual vices should invoke an individualistic response, recent trends in vice epistemology have taken a structural approach to the study and amelioration of epistemic vices (Battaly 2013, 2016b; Kidd 2020; 2022; Levy and Alfano 2020; Tanesini 2021).

What has been acknowledged in these texts and by wider feminist and critical race theorists, is that focusing on an agent-centric virtue theoretical framework ignores the social and systematic causes of oppression, discrimination, and subordination (Dillon 2012; Okin 1996). Additionally, a failure to acknowledge these structural and social systems causes us to lose sight of the power dynamics often at play when individuals are subject to vicious conditions, as placing the responsibility on the individual may serve as justification for blaming the oppressed for their oppression (Tanesini 2021; Tessman 2005).

Arguably, a structural solution is therefore more appealing for the problems stemming from information disorder, which as discussed creates corrupting conditions that can prevent individuals from forming and exercising the necessary intellectual virtues needed to recognise and be unpersuaded by information disorder. Additionally, as the identified problem lies in the corrupting conditions of information disorder, we should at least in part direct our solutions at information disorder and the institutions and environments themselves as opposed to the individuals who are subject to the system. A solution from this approach would not place the burden on the individual to develop the virtues required to resist information disorder, but instead focus on building resistance in the wider epistemic community (Battaly 2016b; Kidd 2020, 2022).

[132] This solution is supported by research that has found that analytic thinking reduces belief in conspiracy theories (Swami et al. 2011)

What follows then, is a need to tackle epistemic vices at a structural, societal level. Ironically, this may mean we end up focusing less on virtues and vices, and more on the environments that sustain vice instead (Croce and Piazza 2021; Gardiner 2022:111; Levy 2022:113). For vices formed or sustained by information disorder, this could include media algorithm adjustments, changing reporting mechanisms or regulatory policies (McIntyre 2018, 2019, 2020; Rini 2017). For example, social media platforms or news websites may change their algorithm to prioritize accurate information, minimise sensationalism or reduce the spread of misinformation (Giansiracusa 2021). These same sites could also implement effective reporting systems that allow users to flag or report harmful content (Rini 2017). Finally, governments may introduce or revise regulations that hold media and tech companies accountable for the spread of false or harmful information (Hartley and Khuong 2020).

As noted, these strategies aim to address the structural aspects that contribute to the propagation of epistemic vices. Whilst still being vice-targeting solutions, implementing them exceeds the capabilities of most epistemic agents unless they are in positions of power (e.g., those in charge of educational systems or social media algorithms). They may be particularly appealing for the problems stemming from information disorder, which as discussed creates corrupting conditions that can prevent individuals from forming and exercising the necessary intellectual virtues needed to recognise and be unpersuaded by information disorder.

6.6.3 A Social Approach

Given the impact that these wider systems and structures can have on our epistemic vices, I agree that our solutions to address vices and their resulting harms should not be wholly individualistic. However, I argue that a medium can be struck between both individual and structural approaches, considered to be the social approach. Social solutions build upon an understanding of epistemic vices as social and involve strategies that consider the communal and interpersonal aspects of addressing epistemic vices (Tanesini 2021:8).

Here, I turn to an area of social virtue epistemology that focuses on other-regarding or ‘outward facing’ virtues to develop this solution (Alfano et al. 2022:8-9; Sullivan and Alfano 2020; Tanesini 2021:9; Von Wright 1963:153). I argue that by developing these virtues alongside our individual ones, we can go some way in offering a corrective claim for vicious behaviour. I

also argue that social solutions can be virtue-centric, despite often being perceived as individualistic solutions.

For clarification, let us emphasise the distinction between social and structural solutions. As we've seen, *structural* ameliorative solutions are practises that are aimed at the wider, vice-inducing or creating institutions or systems (Kidd 2020:70-71, 2022:84-85). Alternatively, social ameliorative solutions involve understanding the social impact that vices and virtues can have on one's wider epistemic community. These solutions go beyond individualistic considerations and extend to the impact that an individual's vice can have on the collective community. However, these solutions still focus directly on virtues and vices.[133]

Arguably all three approaches, individualistic, structural, and social can work in tandem with one another.[134] In the remainder of this chapter, I will focus on ameliorative solutions that take into account both individualistic and social approaches. Focusing on virtue cultivation, I argue that by educating for virtues that are social in nature, one's virtues can have a wider impact on your surrounding community.[135] Working alongside individualistic virtues, this can have the resulting impact of a virtuous agent who is concerned with their own epistemic flourishing and others around them.[136]

6.7 Epistemic Networks

With this in mind, we can now turn to assess an ameliorative solution that takes both individualistic and social approaches into account, offered by Emily Sullivan and Mark Alfano (2020). Sullivan and Alfano (2020) discuss the various ways in which epistemic virtue and vice depend on the larger structure of one's epistemic community. When discussing social epistemic networks, they identify three classes of virtues that they claim only arise in social epistemic networks: monitoring, adjusting, and restructuring. These virtues have a similar structure to traditionalist virtues and involve sub-dispositions of attention, motivation, and cognition. Additionally, the identified virtues are scaffolded by one another – effective adjusting is possible only if one is sufficient in monitoring, for example. Each of these virtues is both 'self-

[133] See Tanesini (2022:140) for whether 'mindshaping' could be one such solution.
[134] As Kidd observes (2022:85), not every vice will warrant a structural ameliorative approach.
[135] I focus on these two approaches as I consider structural approaches to be beyond the realm of my expertise and often, epistemology itself. However, I do conclude by briefly detailing how all three approaches may work together.

regarding' and 'other-regarding'. Self-regarding virtues enhance one's own position and are often to the advantage of their possessors. They are contrasted with other-regarding *virtues*, virtues that are primarily beneficial to other people and the surrounding community (Von Wright 1993). This distinction mirrors a distinction between two types of regulative epistemology: one aimed at reforming one's individual epistemic conduct, and another aimed at the reform of our shared epistemic systems and environment (Kidd 2022).

Briefly, monitoring requires one to monitor and understand the structure of a social network and the track records of those in the network. One can monitor their own network (self-regarding) or others (other-regarding). Crucially, the aims of monitoring must be virtuous, such as increasing the well-being of the epistemic community. In this sense, monitoring can also be vicious such as the monitoring of social media platforms for financial gain. Successful virtuous monitoring allows agents to actively keep track of their epistemic position in various domains and contexts, e.g., recognizing one's position in epistemic echo chambers (Sullivan and Alfano 2020:157-158).

Monitoring one's social network involves adjusting the weight one should give to sources and information spread throughout the network. By doing this successfully, one becomes able to *adjust* their credence to account for imperfections such as cases of mis or disinformation. Again, adjustment can be focused on one's own network (self-regarding) or others e.g., by suggesting that an agent puts more or less trust in various sources located in their social epistemic network (other-regarding). This can be beneficial in that I can benefit others by suggesting what amount of trust to give to various sources, or harmful by recommending trust in an unreliable source (ibid.).

Finally, virtuous *restructuring* concerns social networks which are so flawed that they must be modified e.g., seeking out new sources or no longer trusting sources you once trusted. Similar to the previous discussion on contingent and intrinsic corrupting systems, Sullivan and Alfano note that one can 'rewire' their social networks (self-regarding) or the social networks of others (other-regarding). As with the other virtues, restructuring can be with the aim of epistemic improvement, or vicious in the sense that one restructures their network with the aim of excluding reliable testifiers or sources (ibid.:159).

When addressing the question of how to restructure one's social network, Sullivan and Alfano argue that the solution to false and intentionally misleading news (disinformation) requires one to take a more encompassing view of what it means to do well epistemically (ibid.). Individual solutions which are only concerned with one own's concern for truth means agents neglect the epistemic well-being of their wider community, in turn preventing them from acquiring other-regarding virtues. The problem of fake news and other related phenomena is a collective action problem, thus single-minded approaches which focus on improving individual beliefs will therefore be unsuccessful in addressing the wider problem[137].

Sullivan and Alfano discuss a so-called 'single-minded approach' to the problem of fake news and related phenomena when discussing the restricting of one's testimonial network (ibid.:160). They focus on Neil Levy's (2017) claim that agents should cut themselves off completely from sources of fake, misleading, and unreliable news, as even being exposed to false information leaves us vulnerable to acquiring false beliefs. This means we should limit our sources to only those that reliably produce true and accurate information and eliminate those that do not.

However, Sullivan and Alfano argue that whilst this process may appear to be a restructuring virtue, it can manifest as a restructuring vice. They outline three concerns with this so-called 'divide and conquer' (Sullivan and Alfano 2020:160) solution to fake news, which involves cutting oneself off from untrustworthy and unreliable sources.[138]

Sullivan and Alfano first acknowledge that if an agent cuts themselves off from an untrustworthy source, they may become more dependent on the remaining sources that they do trust. Therefore, by making oneself less vulnerable against an untrustworthy source, your overall network becomes less secure as you have made yourself more vulnerable to the remaining sources (ibid.:159-160). This is particularly problematic because the reliable status of a source can frequently change over time. Additionally, being exposed to unreliable sources allows for epistemic growth by providing agents with an opportunity to exercise and develop their epistemic skills e.g., learning how to spot and avoid similar but different bad epistemic behaviour in future instances from bad epistemic examples (Alfano 2013; Sullivan and Alfano

[137] One may argue that if each individual focuses on their personal epistemic enhancement, the collective result of multiple good epistemic agents would inherently lead to an improved epistemic community. How effective this concern would depend on whether Sullivan and Alfano consider epistemic communities to be more than the mere aggregation of individual agents (2020:149).

[138] Sullivan and Alfano (2020) do not deny that there is *never* a source that you should sever ties with, but that there are persuasive reasons to allow for the unreliable source to remain in an epistemic network.

2020). Instead, they argue that agents should focus on safeguarding the overall network security as opposed to an individual's belief security. This means we should abstract away from specific sources and individual beliefs and focus more on the wider structure of the entire network.

Additionally, Sullivan and Alfano argue that limiting any engagement with unreliable sources lessens an agent's potential to develop other-regarding restructuring virtues (Sullivan and Alfano 2020:159). As we have seen, other-regarding restructuring virtues involve being disposed to help others rewire their trust (and distrust) networks so that they are epistemically better off and less vulnerable. However, if an agent is overly concerned with limiting their own exposure to false and unreliable information, they will be unable to advise others on how to better their networks. Instead, in order to help the wider community, agents need to monitor other networks which involves exposing themselves to false and misleading information.

Finally, Sullivan and Alfano argue further that even in cases where an agent has well-intentioned motivations and access to the truth, reducing another's network security by cutting off sources makes others too epistemically dependent on them, even if it's for the good of the wider community. They present the example of Plato’s philosopher-king who prevents the public from accessing art and fiction and tells the public untruths all for the sake of their epistemic well-being. Comparing this behaviour to someone who isolates others from unreliable sources, Sullivan and Alfano argue that this makes people less intellectually autonomous and less able to enjoy epistemic growth (ibid.).

Given the denial of this 'single-minded' solution to fake news, we can ask where does Sullivan and Alfano's solution to information disorder leave us? If we should not aim to isolate ourselves from unreliable sources of information, it seems the alternative is to allow them to remain in our network for the previously discussed social benefits. However, in view of the arguments made in this paper and the threat that information disorder creates to our intellectual characters, there seems to be a potentially worrying concern with this alternative solution. Let us now turn to evaluate it.

We can first survey the position defended by Levy (2017) that cutting yourself off from unreliable sources that create misinformation can prevent epistemic vices from forming. As we have argued throughout this chapter, information disorder gives way to a variety of epistemic vices, meaning those who have not yet developed the previously identified virtues of

monitoring, adjusting, and restructuring are being left vulnerable to develop these vices if fake news remains in one's network.

However, despite this risk, Sullivan and Alfano consider the existence of information disorder an opportunity to exercise and develop virtues (2020:158). Allowing unreliable sources to remain gives agents an opportunity to monitor their networks and exercise the necessary virtues involved in doing so as discussed previously. It also allows for individuals to focus on building other-regarding virtues instead of focusing exclusively on their own exposure to sources. Both of these benefits point to the wider aim of Sullivan and Alfano's (2020) solution, which is not solely focused on instilling individual virtues to combat fake news but to also assist in the formation of other-regarding virtues for the wider epistemic community.

Sullivan and Alfano also present two concerns with the elimination approach to unreliable sources. First is a problem of dependence (agents become too dependent on their remaining sources or too dependent on the person who cuts off the unreliable sources) (ibid.:159). Second is a problem with its single-minded nature (being disposed to fake news is an opportunity to exercise your virtues and focusing on the unreliable nature of a source prevents you from developing other-regarding virtues, particularly through monitoring) (ibid.:160).

However, it is questionable just how concerning both of these problems really are. With reference to the problem of dependence, it's hard to see how keeping knowingly unreliable sources in one's network prevents you from being too dependent on other sources. This is because once a source has been identified as unreliable, even if it is kept in one's network, it still will not be depended on. Additionally, if an agent is still exercising their virtues, then they would also be wary of the reliable status of the remaining sources and its potential to change over time. Finally, with regards to the concern of depending too much on the individual that cuts off your unreliable sources, it seems the dependence is more aptly aimed at depending on the remaining sources, not the individual themselves, which as we have seen is not necessarily problematic.

Additionally, to address the second concern that cutting off unreliable sources is too single-minded an approach, it can be argued that whilst being disposed to fake news allows you to exercise your virtues, this is not an argument for keeping that source in your network once it has been identified as unreliable. Additionally, there is no reason to say that once a source has

been identified as unreliable it cannot be cut off from the wider community and not just for the individual, which would mean the approach is not necessarily single-minded.

By examining the structure of ourselves and others' epistemic networks and by identifying these three other-regarding and self-regarding vices, we can see how Sullivan and Alfano's (2020) framework could be a potential social solution to the vice-inducing corrupting conditions found in many online environments. However, there is clearly an important balance to be struck between both individual and social ameliorative practises e.g., preventing dangerous conditions for individual vices to form versus allowing these conditions to persist in order to develop and exercise other-regarding vices.

How then, can we coordinate the individualistic and social ameliorative approaches to vice, specifically when it comes to our online environments?

One way to do this is to distinguish between 'outside-in' and 'inside-out' ameliorative approaches to vice (Steven Bland 2022a). Simply, outside-in strategies scaffold environments to promote virtuous habits and inside-out strategies cultivate habits of scaffolding benign environments. This differentiation mirrors the distinction between self-regarding and other-regarding virtues, and both strategies emphasise cultivating virtues to overcome epistemically harmful behaviour. However, as Bland understands them, these strategies offer co-dependant social and individual solutions. In other words, both outward-focusing and inward-focusing virtues can work together to promote virtuous habits and environments (ibid.:30).

Bland focuses his attention on cognitive biases, a manifestation of epistemic vice which he believes are the result of internal psychological processes and external, environmental conditions (ibid.:15). Given the internalist and externalist foundations of cognitive bias, Bland argues that it cannot be expected that a solution to debiasing tackles only one of these dimensions.

Finding the balance between these two strategies is complex. We have seen this exemplified in Alfano and Sullivan's (2020) argument between developing social, outward-looking virtues, and preventing individual, inward vices. Arguably what seems to be going wrong in solutions such as these is that they focus exclusively on the individual problem i.e., developing self-

regarding virtues or exclusively focusing on the social problem i.e., developing other-regarding virtues, thereby ignoring how the two strategies intersect.

A fully coordinated approach must perceive ameliorative strategies via an *interactionist* lens. This means that the personal and situational factors that lead to vices are not treated as independent influences, but perceived to interact with one another, meaning epistemic agents can be influenced by their environments yet still exercise some influence over them too (Bland 2022a, 2022b; Kilhlstrom 2013). This means that ameliorative solutions directed solely at the individual or social factors contributing to vice will be incomplete and misleading. We should therefore not view these solutions as isolated from one another but instead combine the two approaches.

What implication does this have on vice ameliorative practices then? And more specifically, on the epistemic vices associated with our online environments as outlined in this chapter?

Putting all of the above together, a coordinated approach to the amelioration of epistemic vices would combine both inward and outward strategies. This may prove particularly effective in an online, epistemic environment which is inherently social (Sullivan and Alfano 2020). Individuals can still be educated for virtues, including ones that primarily benefit their own epistemic character. However, outward-facing virtues should also be educated for.

For example, when individuals uphold outward-virtues such as intellectual perseverance and epistemic integrity, they exhibit a reliable commitment to seeking accurate information, engaging in rigorous fact-checking, and being open to updating their beliefs based on evidence (Kawall 2002; Zagzebski 1996). This personal dedication extends to the content they share online and the conversations they participate in. As a result, they become gatekeepers of information accuracy, acting as filters that prevent the propagation of falsehoods. When these individuals encounter misinformation, their commitment to truth-seeking and evidence-based thinking leads them to be cautious about sharing or endorsing unverified claims. By setting such an example, they foster a healthier information ecosystem, where accuracy takes precedence.

Moreover, practising epistemic resilience could reinforce the value of considering diverse viewpoints and engaging in empathetic discussions. By encouraging individuals to venture

beyond their ideological comfort zones, epistemic resilience cultivates a culture of understanding and empathy. This, in turn, contributes to the bridging of ideological divides and the reduction of polarization.[139]

In this way, epistemic virtues not only shape an individual's own online behaviour but also outwardly influence the behaviour of others. Their impact extends beyond their immediate circle, positively shaping the broader epistemic environment and fostering a culture of reliable information and constructive dialogue.[140]

To conclude with a point I previously alluded to, I do not believe that virtue-centric corrective claims are the only ameliorative solution to the epistemic harms of online environments. Virtue-centric approaches, even social ones, can only go so far. Therefore, they should collaborate with structural solutions to make meaningful progress, such as using warning labels for untrustworthy sources, monitoring online forums, or algorithm changes (Rini 2017). However, I do not believe that virtue-centric solutions should be dismissed entirely, particularly on the grounds that they are too individualistic. Epistemic vices are a complex mesh of individual and social factors, meaning their treatment should be multifaceted too.

6.7 Conclusion

This chapter has argued that information disorder, understood as the correct way to refer to the problem of fake news and other related phenomena, can be epistemically harmful to one's intellectual character in so far as it can promote intellectual vices. Following Kidd's (2019, 2020) definition of epistemic corruption and his framework for a successful corruption criticism (to label something as epistemic corrupting), I have outlined how online media, in so far as it produces information disorder, is epistemically corrupting. I began by identifying the corruptor and the corruptees (the media and its consumers), before identifying three distinct epistemic vices that information disorder leads to - prejudice, conspiracy thinking and

[139] As with most claims concerning the consequences of virtues and vices, these are behaviour predictions. Empirical testing would need to be carried out to know the full impact that character traits have on online environments. For example, some research already suggests that our character traits shape the internet, for worse or better (Meyer and Alfano 2022).
[140] See Battaly (2021) for a discussion of how closedmindedness, understood here as an epistemic virtue, can help epistemic agents engage with polluted media feeds. Battaly argues that by promoting the virtue of closedmindedness we can 'flood the epistemic environment with truths and critical thinking' (2021:312). This can also have a positive knock-on effect for other users of that online environment.

epistemic capitulation. I discussed how these vices are brought about by information disorder in accordance with the five modes of corruption focusing on how vices can be intensified, propagated, and created.

Finally, I concluded by examining the various ways information disorder can overcome its corruptive state by assessing individualistic, social, and structural ameliorative solutions. I argued that virtue-centric solutions are not restricted to the individual, and instead, a coordinated approach that aims to develop individual and social virtues can go some way in addressing the concerns outlined in this chapter.

CHAPTER 7. TRUSTWORTHY INSTITUTIONS: A VIRTUE-THEORETIC ACCOUNT

7.1 Introduction

This chapter makes a novel connection between the literature on trust and vice epistemology to identify institutional virtues of trustworthiness and their corresponding vices. Epistemic vices are dispositions, attitudes and ways of thinking that make us bad thinkers, in so far as they prevent us from acquiring and sharing knowledge, manifest bad motives, and desires, or disrupt both individual and collective epistemic functioning (Kidd et al. 2020). Because of their bad motives or obstruction of epistemic goods, we can be held responsible for these epistemic vices (Battaly 2016a 2018a; Cassam 2016, 2019a; Tanesini 2018, 2021). When asking whether an individual possesses a certain evaluative attribute, one natural way to answer this question is to assess whether they possess the virtues or vices of that quality. The same is also true of institutions, such as governments, educational institutes, corporations, and the media. Accordingly, when asking if an institution possesses a certain attribute such as trustworthiness or untrustworthiness, a natural way to answer this question is to assess whether the institution manifests certain virtues or vices pertaining to trustworthiness. In this sense, just as an individual's character reveals multiple virtues and vices so too can an institution's character (Fricker 2010, 2021).[141]

This branch of vice epistemology has remained relatively underexplored until recently, with most literature to date focusing on the *virtues* of groups (Baird and Calvard 2019; Broncano-Berrocal and Carter 2019). Likewise, the literature on trust and trustworthiness has often focused on the individual e.g., capturing the nature of this attitude, the circumstances under which it may be rational, and what makes an individual trustworthy (Baier 1986; Hawley 2019; Jones 2012; Kelp and Simion 2023; Potter 2002).

However, there is great value to be gained from focusing on the question of what makes an institution trustworthy or untrustworthy, and one that is of timely significance given the

[141] I will be focusing on epistemic institutional virtues and vices, but I am also open to the idea that moral virtues and vices also play a role in determining the trustworthiness of an institution.

consistent downward trend in levels of trust reported across many institutions (Davies et al. 2021; Ipsos 2019; Murtin et al. 2018). For example, the Edelman Trust Barometer (2020) which is based on a survey of trust in institutions, found a downward trend in trust in government, corporate, and media institutions. There is also value to be found in examining the truth-related epistemic harms arising from vicious institutions in the form of corruption and misinformation (Kidd 2019, 2020, 2021, 2021 et al., 2022; Lackey 2020).

This chapter explores the ways that a vocabulary of epistemic virtues and vices can be used in our evaluation of trustworthy institutions and how understanding specific institutional vices can help us determine whether and under what conditions institutions are trustworthy. It is worth noting here that I will be working with an exogenous perspective as to what makes an institution trustworthy or not. This means I will be looking at the properties or characteristics of public institutions from an external point of view e.g., what institutional factors generate a reaction of trust from those outside of the institution, namely members of the public.[142]

As I predict there are many epistemically virtuous and vicious indicators of a trustworthy institution, I will be focusing on the ways that institutions can be deemed trustworthy or untrustworthy in how they communicate with non-experts. This focus is an important and timely one seeing as public trust in expertise, whether that be medical advice, climate research or the media has become a contested subject in recent years (Edelman Trust Barometer 2020). The increasing access to information online offers unlimited sources for the public to inform themselves, but with this comes the difficulty in figuring out which sources are credible and trustworthy. By figuring out what features a trustworthy or untrustworthy institution possesses, we take a step in the right direction to minimize this worry.

The plan for this chapter is as follows. Part one will focus on the question of how institutions can possess virtues and vices over and above the individual-level virtues and vices of the institution's members. Here, I will draw from work on group epistemology, specifically Margaret Gilbert's (1987, 2000, 2002, 2013) plural subject theory and Miranda Fricker's (2021) account of an institutional character or 'ethos'. This allows us to make the distinction between an untrustworthy institution and an untrustworthy individual and demonstrate how it is the institution itself, not the individuals who belong to it, who are vicious. I will then direct

[142] I opt for this perspective because I will be identifying the indicators that can help determine the trustworthiness of an institution. I argue that these indicators can be institutional virtues and vices.

two criticisms towards Fricker's account of institutional vice, taking issue with her interpreted consequential reading of vice and the self-awareness requirement which requires members to be aware of their motive away from epistemic goods. Following this, I present two amendments to Fricker's account in order to overcome these concerns. By dropping this self-awareness requirement and by developing a consequential reading of institutional vices, I argue that both objections can be overcome.

Having explained how institutions can display epistemic virtues and vices, part two of this chapter will focus on how an institution can be trustworthy/untrustworthy *as an institution*, by displaying certain communicative attributes of transparency and honesty. I then explain how these attributes (and their counterpart vices) are virtues and vices of institutions pertaining to trustworthiness. I conclude this chapter by raising and responding to potential objections due to recent work by C. Thi Nguyen (2021) and Stephen John (2018).

7.2 Virtuous and Vicious Institutions

Work in vice epistemology often suffers from an individualist bias. For example, foundational work on the structures and features of vice focus on what is required for an individual to possess a vice e.g., whether the individual must possess a character trait, attitude or thinking style, have epistemically bad motives or be blameworthy for their vice (Battaly 2016a, 2018a, 2019; Cassam 2016, 2019a; Tanesini 2018, 2021a). Analyses of specific vices focus on how the vice is exercised by an individual e.g., how vices of superiority are characteristic of people who occupy positions of privilege (Medina 2013; Tanesini 2021). Finally, ameliorative work provides solutions to individual vice e.g., educating for virtues in individuals or exposure to role models (Baehr 2013; Battaly 2016b; Zagzebski 2017).

Not only is it important to focus on collective virtues and vices for the pressing reasons identified above, but it is also commonplace to talk about collectives, particularly institutions, as possessing particular virtues and vices. For example, we can describe governments as insouciant, research teams as tenacious, and charities as fair-minded. These virtues and vices can also have a precise epistemic dimension to them in so far as they often concern distinctively epistemic goods such as truth, reasoning, or knowledge. A diligent jury, for example, will

consistently and carefully examine evidence, whereas a dogmatic committee will fail to acquire truth through their assertive bias.

In order to ascribe a virtue or vice to an institution, therefore, we need to be able to ascribe it to the institution per se. This will involve appealing to work on group epistemology as a starting point to demonstrate how collectives such as institutions and groups can hold epistemic features such as virtues and vices.

7.2.1 Collective Virtue and Vice

A widespread debate in the epistemology of groups is the argument between summative and non-summative accounts of groups (Bird 2019; Kallestrup 2016; Lackey 2021). Briefly, summativism states that a group phenomenon e.g., a belief or vice, is nothing more than the 'summation' of the beliefs/vices of the group's members (Quinton 1976: 9). Summativist views differ with respect to whether all or some members of the group must possess the relevant property in order for the group to possess it or whether only the 'operative' members of the group do (Lackey 2020).

Conversely, non-summativism makes the general claim that a group-level property cannot always be reduced to a summation of individual-level properties, and instead can be ascribed at the 'collective' level (Gilbert 2000:39). Moreover, on some non-summativist views, the group's possession of the relevant feature may not even partially consist in any group members possessing that feature: it may be possible for a group to possess a belief or vice that none of its members possesses e.g., a jury that delivers a guilty verdict despite all the individual jury members privately believing the person is innocent.

For institutions, a non-summativist view may maintain that the institution of the UK government has the vice of arrogance, yet not all or many of the individual staff members do. Conversely, according to summativism, the ascription of the vice of arrogance would apply to all or many members of the government. Under both readings, the government possesses the vice of arrogance, but under summativism, it is because the vice is exercised in the individuals and for non-summativism, it is because the government itself is the bearer of the vice.

Arguably, a non-summative view will go beyond a mere denial of the positive claim embraced by summativists and explain how it is that an institution can possess virtues and vices over and above the individual members of the institution in question. One of the main reasons in support of a non-summative model for institutional vice is that members of an institution can behave in different ways in the institutional context than they would outside of it. Take, for example, a football supporters club. Individually, each member of the club is (overall) a decent and even virtuous person, but when they come together to watch a match, they become threatening, rude and bad-tempered. If this happens persistently on multiple occasions, the supporters club can be said to hold the vices of anger and insolence. However, the individual members do not hold these vices as they are not exercised in their private lives. Only when they come together as a club and display these behaviours regularly do the vices present themselves, meaning it is the supporters club per se that holds the vices of anger and insolence.[143]

Opting for a non-summative interpretation of institutional virtue and vice also allows us to assign responsibility to these virtues and vices. To identify a vice within an institution would mean recognising the institution itself as the vice-bearer and not (just) the individuals themselves who form the institution (although that sometimes may also be the case). In this sense, a summative model will not suffice for an institutional account of virtue or vice, as the charge would solely be directed to the individual and not the institution.[144]

Whilst work in collective virtue epistemology is relatively modest, several writers have proposed that groups can possess virtues possessed by few or none of its constituent members, thereby siding with non-summativism (Byerly and Byerly 2016, Byerly 2022a; Fricker 2010, 2021; Lahroodi 2007, 2018). These arguments are instances of so-called divergence arguments in favour of non-summativism. Such views aim to establish that phenomena at the group level can diverge from what is happening at the individual level among the group's members. Divergence can occur when group members behave notably differently in the group context than they would outside of it (Fricker 2010). One way this can occur is when the behaviours or traits of a group of individuals are 'cancelled out' at the institutional level. For example, the open-mindedness of individual journalists gets cancelled out when working for a biased news publication. A divergence between a group property and individual property can

[143] See Lahroodi (2018) for a further discussion on this.
[144] By opting for non-summativism interpretation we can hold the untrustworthy institution responsible as well as the untrustworthy individuals comprising it. This is because non-summativism does not deny the existence of simple aggregate groups. I return to this argument in section 7.2.3.

also occur when an individual's membership to certain groups (e.g., a parent, guard, jury member) involves a commitment to certain group standards that override their personal dispositions. For example, a member of parliament may oppose fox hunting but vote to not ban it due to loyalty to their political party.

These are all explanations of how group phenomena, including virtue and vice, can occur at a group level under a non-summative view.

7.2.2 Joint Commitments

At this stage, it seems clear that a non-summativist approach can best explain how an institution itself can have a vice over and above the individual-level vices of its members. However, it is not clear yet what constitutes an institutional virtue or vice if individual members do not hold it. This will be the focus of the next section.

To date, the most developed account of institutional epistemic virtue and vice is offered by Fricker (2010, 2021). Building on the non-summativist joint commitment model, Fricker argues that institutions display virtues and vices in part when the members of the institution jointly commit to an epistemic motive.

The non-summative joint commitment model that Fricker bases her account on is presented by Gilbert (1987, 2000, 2002, 2013). Under Gilbert's view, group belief is the result of members jointly committing to accept a proposition as the group's, even if no member believes it herself. Several individuals constitute a plural subject in virtue of a joint commitment to doing something 'as a body', whether this is holding a collective belief, making collective agreements, or feeling collective emotions (Gilbert 1987:194).

A joint commitment arises when each individual 'openly expresses his or her readiness to be jointly committed with the relevant others' and when all members 'understand themselves to have a special standing in relation to one another' such that 'no individual party […] can rescind it unilaterally' (Gilbert 2002:126). Importantly, this implies the collective commitment is not reducible to the personal commitments of the individuals who are part of it. Additionally, those who hold a joint commitment each have an obligation to one another to 'do whatever is best

with whichever of the possible conforming combinations of actions the others are doing their part in' (Gilbert 2013:402).

Fricker adapts Gilbert's account to offer a 'pluralistic' or collective template for virtue and vice by observing that we can employ the notion of joint commitment with respect to group *motive.* Instead of requiring that group members hold a joint commitment to a belief or trait, members can hold a joint commitment to a motive - a joint commitment to achieving the good end of the motive because it is good. Subsequently, a joint commitment to a virtuous motive is a matter of jointly committing to a virtuous end for the right reason (Fricker 2010:241-242).

As a non-summative view, it follows that group members need not possess the motive as individuals. Rather, in jointly committing to it, they each come to possess it as a group member. Once we add into this group motive a reliability condition, we now have group virtue according to Fricker.

Additionally, Fricker argues that we should relax Gilbert's unanimity requirement on joint commitment as this is at odds with the non-summativist idea that groups can have virtues without most or all of their members possessing them. Instead, Fricker introduces the concept of 'passengers', namely group members who lack good motives/skills as individuals but 'go along with' the commitment as group members (ibid.:254).

7.2.3 Institutional Ethos

An institution's behaviours and motives towards certain commitments constitute what Fricker calls an 'institutional character', or 'ethos', defined as 'the collective analogue of an individual agent's character' (Fricker 2021:90). An ethos allows institutions to hold institutional values (e.g. equality, accountability, integrity) and demonstrates what they stand for. Take, for example, a university department that aims to deliver fair assessment grades. Alongside awarding the correct grade for the essays, it matters that the decision comes from appropriate value commitments. If the university awarded their students a higher grade to gain positive course evaluations, the institution would not be acting from the right kind of values e.g., the value of truth and fairness. Examples like this demonstrate that it is the appropriate ethos, not just the outcome, which is important and integral to an institution.

As well as arguing for the existence of an institutional ethos, Fricker argues that the ethos should be understood from a virtue and vice-theoretic framework (2010:229, 2020:89-90). Specifically, she characterises institutional ethos as consisting of the 'collective motivational dispositions and evaluative attitudes within the institutional body, of which the various good or bad ends orientate the institution's activities' (2020:91). Just as character explains the behaviour of individuals in relation to their motivations, desires, and values, ethos explains the behaviour of an institution in relation to its constituent motivations, goals, and values – those things for which it stands. For example, an ethos of justice is exemplified in a jury by 1) holding certain values (e.g. fairmindedness and equality), contained in 2) certain institutionalised procedures (e.g., trial by jury, right to a defence), which 3) delivers the right sorts of results (e.g., fair sentencing) (Kidd 2021:350).

It is worth noting here that crucial to the definition of a vice is that the vice must occur systematically (Cassam 2019a:4). This allows for differentiation to be made between one-off displays of bad behaviour and acts which stem from one's character. For example, suppose one is not normally or systematically closedminded. In that case, they cannot be said to possess the vice of closedmindedness, meaning the behaviour cannot be attributed to the individual's character. The same can be said of institutions. Institutions can have fleeting lapses of judgement and act 'out of character' which does not speak to the institutions' ethos (Fricker 2020:100). For example, an otherwise trustworthy media publication publishing a non-fact-checked news story would not be said to possess the vice of carelessness as it is not evidence of a systematic decline in their reporting. Consistently displaying these bad motives or values will determine whether the institution possesses a vice as opposed to a fleeting harm or wrong.

Additionally, responsibility is as integral to the definition of institutional virtue and vice as it is to individual virtues and vice. Fricker's account is responsibilist in the sense that a responsibility condition is integral to vice. The vice-bearer (understood as an institution or an individual) only acts viciously when there is a culpable lapse in either the motivation or effects aspect (ibid.:98).

Whilst Fricker does not detail the specific type of responsibility in mind for institutional virtues or vices, one sense in which institutions can be held responsible is in an attributability responsibility sense (Sher 2006:57; Shoemaker 2015:38; Watson 2004:270). Focusing on vices, blame is apt under this form of responsibility as the behaviour or trait is properly

attributable to the agent as it represents part of their character and therefore warrants praise when good and blame when bad. Likewise, with institutions, motives and actions stem from the institution (either the institutional ethos or commitment), therefore reflecting the institution per se, meaning it can be attributed to it and praiseworthy or blameworthy. Other answerability practices may also be required from vicious institutions, such as transparency requirements or whistleblowing (Ceva and Bocchiola 2019).[145]

In sum, we now have the following picture of institutional virtue and vice. Institutions possess an institutional ethos analogous to an individual character. This ethos is sustained by various joint commitments, made by the institution as a 'body' and understood as a commitment to a motive to an epistemic end. When the motive is towards an epistemic good, for the right reason, it is virtuous. Conversely, when it is towards an epistemic bad or away from an epistemic good it is vicious. An institution must display these motives consistently and culpably in order to be charged with an epistemic virtue or vice.

7.3 Objections to Fricker's Account

In this section, I will raise two objections to Fricker's (2020) account of institutional vices and offer two suggestions on how to overcome them. First, I argue that Fricker's consequentialist interpretation of vice fails, raising concerns over the hybrid nature of her account. Secondly, I raise doubts about Fricker's 'self-awareness' requirement for institutional vices, arguing that it at odds with the motivational interpretation of vice that Fricker appeals to (2010:253-254).

7.3.1 A Hybrid Account of Epistemic Vice

Starting with the first objection, we can briefly explain the differences between two competing analyses of the 'badness' of epistemic vices: motivationalism (Tanesini 2018, 2021), and consequentialism (Cassam 2016; 2019a).

[145] Whilst Fricker does not advocate for attributability responsibility explicitly, she does stipulate that blame is the appropriate response to vices (2020:100). Even stronger, she considers blame to be integral to the definition of an institutional vice (Fricker 2020:105). Also, see Chapter 4 for a detailed discussion of attributability responsibility.

On a motivational reading of vice, epistemic vices involve the presence of bad epistemic motivations either away from an epistemic good or towards an epistemic bad (Baehr 2015, 2020; Montmarquet 1993; Tanesini 2018, 2021; Zagzebski 1996). What is integral to vices are therefore the good or bad motives, as they determine whether the trait in question is virtuous (possesses good motives) or vicious (possesses bad motives). On a consequential reading of vice, epistemic vices are dispositions or behaviours that systematically produce epistemically bad effects such as obstructing knowledge (Cassam 2016, 2019a). On this view, epistemic vices are not vices due to their bad vices, but their bad consequences. Vices therefore need not have epistemic motives that account for their badness.[146]

What reading does Fricker's account of institutional vice align with? First, Fricker (2020) holds that the badness of a vice can be explained in reference to its bad motives. By appealing to Gilbert's non-summative joint commitment model, Fricker has argued that institutions possess a joint commitment to a motive, understood as a 'collective motive' which can be virtuous or vicious (Fricker 2010:241). This model realises an institutional ethos – a collective set of commitments to a certain set of values.

For an institution to display a vice, it follows that it must possess a joint commitment/motive towards a bad epistemic end or away from an epistemic good. Recognising the vices of the former description are rare (Baehr 2010; Crerar 2018), Fricker argues that 'any motivational disorder constituting an epistemic vice will instead take the negative form of an inadequate commitment to good epistemic ends.' (Fricker 2020:99). It therefore follows that members must express a wilful, *inadequate* commitment to good epistemic ends in order to display a vice.

However, Fricker also claims that these motives towards or away from epistemic goods are not the only way an institution can display a vice as she rejects an 'exclusively motivational account of vice' (ibid.:100). Fricker also holds that institutions can display virtues and vices in how they achieve (or fail to achieve) good epistemic ends. This aspect of vice is consequential as it concerns the good or bad epistemic ends of an action or behaviour. In this sense, Fricker offers a hybrid model for institutional epistemic vices, drawing from both motivational and consequentialist explanations of epistemic vice.[147]

[146] For a consequential reading of the institutional vice of incredulity see Medina (2021).

[147] Fricker uses the language of reliability versus responsibilism to describe vices as opposed to consequentialism and motivationalism. I opt for the latter readings so as not to lead to confusion over different responsibility claims i.e., not all consequential accounts claim we are responsible for our vice. Battaly's (2016a, 2018a) personalism is also a hybrid account

To speak to this consequential categorisation of vice, Fricker argues that institutions can display vices through 'persistent performative failure, even if the motivational commitments, mediate and ultimate, are all that they should be' (ibid.). Fricker presents the following example to explain this further. A school displays an epistemic vice of sloppy information-sharing by failing to replace an online homework system, meaning no homework is given to the students. Ten years on, after a decade of chances to become more efficient and organised, the teachers have become lazy and fallen into repeated performative failures in the implementation of their policies on information-sharing. However, the school's underlying value commitments to the epistemic good remain the same. Given its consistent and culpable performative failures, Fricker argues the school exhibits the vice of bad information sharing (ibid.:99).

In this respect, Fricker's account presents an 'inner and outer' element to virtues, with two possible ways for vices to form: 'epistemic vices are culpable lapse of epistemic virtue either (i) in its inner aspect of mediate and/or ultimate motivations to good epistemic ends, and/or (ii) in its outer aspect of performance— the achievement of those ends' (ibid.).

Expanding on the inner element of vices, Fricker also makes a distinction between 'ultimate' and 'mediate' ends (ibid.:93). Ultimate ends are the ultimate motivations or values that epistemic virtues are committed to e.g., a cognitive contact with reality (gaining truth or knowledge). Mediate ends are intermediary or instrumental goals that contribute to achieving this end. They serve as a means to an end as opposed to having intrinsic value of their own. Additionally, only the value of the ultimate end confers value on the mediate end. For example, we might have the ultimate aim of gaining true beliefs and the mediate end of fact-checking to achieve this. The only reason fact-checking matters epistemically is because fact-checking promotes knowledge (the ultimate end).

Fricker's account therefore aims to offer a hybrid explanation of epistemic vices through these inner and outer elements. Institutional vices can be interpreted via a motivational *or* a consequentialist lens, depending on whether the failure lies in the 'inner' or 'outer' element of the vice.

between responsibilism and reliabilism. However, Fricker's account is not personalist as she argues that we are always responsible for our epistemic vices (2020:105).

However, I consider the consequential component of Fricker's account to be unsubstantiated, meaning her account should be assessed as primarily motivational. My reasoning behind this claim is that we should also consider the outer element of Fricker's institutional vice to be motivational in so far as it concerns achieving the end of the inner motives. Consider, for example, a media publication motivated by the epistemic good of truth. The publication enacts on this motive by performing in certain ways e.g., running staff training programmes and ensuring rigorous fact-checking. Under Fricker's account, if this trait is a systematic and culpable one, we can label it virtuous with respect to the institution's motive towards truth (the inner element) or in respect to the implementation of this motive (the outer element). Yet either way, both possible elements of the virtue are centred around the publication's truth-seeking motive as the outer performances are concerned with implementing that motive through actions.

In other words, it appears that the 'outer elements' of virtues and vices (performances) can be traced back to the 'inner elements' of virtues and vices (the motives and values). In this sense, the outer performances are implementations of the motive, serving as practical ways to bring the motive into effect. The outer actions are carried out because of the motive towards an epistemic bad or away from an epistemic good. Outer performances are implementations of the motive, i.e., ways to put the motive into practice. Through this reading, performances (understood as the consequentialist element of vice) can be traced back to the inner, motivational element, making the motive the defining feature of the virtue or vice.

What about Fricker's example of the school with the vice of bad information sharing? If we recall, Fricker offers this case as a reason for why she offers a hybrid account of vice, as it demonstrates how institutions can display epistemic vice through persistent performative failures, even if the motivational commitments are all that they should be. In this example, the (outer) performative aspects alone are what determine the institution's vice, meaning they are seen as distinct from the (inner) motivational ones.

We can present two responses here. Firstly, it seems more plausible that a school that is systematically and culpably failing to share important information does *not* respect the relevant epistemic good of knowledge sharing. Whilst occasional lapses in information sharing are compatible with the school being motivated to a good end overall, systematic, and culpable failures, the kind which is required for vice, are not.

More importantly, as we have seen, performative failures are bad because they represent a failure to implement a motive. In this case, the school has failed to implement its ultimate or mediate ends of cognitive contact with reality or valuing knowledge. Another way of putting this is that there is a misalignment between the institution's motive and the implementation (actions) of that motive. But this does not explain why the school is vicious on a consequentialist interpretation. The school's lapse in efficient information-sharing practices is not vicious because it results in bad epistemic effects or obstructs knowledge, regardless of the motive (Cassam 2019a). The school is vicious because it has not properly satisfied its motivations with its actions. It is difficult to see in what sense this is a consequentialist explanation of a vicious failure, and therefore in what sense the actions of the school alone determine its vice with no reference to motivations.

To summarise, Fricker presented a distinction between the outer and inner elements of vices to explain their badness and culpability, mapping onto motivationalism (inner) and consequentialist (outer) interpretations of vice retrospectively. However, through the above arguments, it appears that the outer elements of these vices are also motivational, in so far as they are failures in implementing the outer motive. This therefore undermines the hybrid nature of Fricker's account of institutional vices.

7.3.2 The 'Self Awareness' Requirement

A potential issue that Fricker is subjected to in her appeal to Gilbert's joint commitment model, is the self-awareness and common knowledge condition that requires the institution's members to be aware of their commitment and thus motive (Gilbert 2000). Described by Fricker (2010:247) as the 'self-awareness' requirement, I argue that an awareness of our virtuous or vicious motives seems at odds with the motivational interpretation of vice that Fricker appeals to.

The reason there is a conflict here is due to the common belief that vicious individuals are not often aware of their vices, let alone possess a wilful and knowledgeable motive to be inadequate at achieving an epistemic end (Cassam 2019a; Holroyd 2020; Medina 2013; Tanesini 2021). Vices of this sort are defined as 'stealthy' or self-concealing, making them

invisible to the vice-bearer (Cassam 2019a). For example, closed-mindedness prevents a closed-minded person from coming to realise that their mind is closed, or arrogance might stop an arrogant person from acknowledging that they are arrogant. These individuals would not just be unaware that they are displaying an epistemic vice but also be unaware that they are doing anything epistemically wrong.

With this understanding, consider again the vice of closed-mindedness that we wish to ascribe to an institution. Under Fricker's account, this vice must be displayed either through 1) members willingly possessing a joint commitment to be closed-minded, or 2) members willingly possessing a joint commitment to be inadequate in achieving open-mindedness. By Fricker's own reasoning (2010:253), instances of the first motivation are unlikely, leaving 2) as the plausible contender. However, if the vice is of the type identified above, it is highly unlikely that closed-minded members are aware of their commitment/motivation away from open-mindedness.

To push this point further, we can examine the weakest form of 'willingness' a member of an institution must possess over their joint commitment or motive. In agreement with Gilbert, Fricker argues that the awareness requirement of the commitment to a value or motive can be so weak it is almost passive and default (2010:247, 2021:96). For example, a government official can passively become party to a joint commitment to keep quiet about a political leader's corruption just by failing to dare to be a whistle-blower. In this sense, their willingness to move away from an epistemic good of truthfulness is very weak, but it allows the official to be aware of their bad epistemic motive (away from truthfulness or towards corruption) and still hold onto the vice of 'failing to speak out'. Members of an institution need only know that they possess a joint commitment to a motive, and not whether that motive is virtuous or vicious (2010:247).

However, whilst this weak awareness requirement seems compatible with some vices, such as a failure to speak up about epistemic wrongs, it is still too strong to be compatible with the stealthy and self-concealing vices identified above. For example, it would not be the case that the official was aware of rife closed-mindedness throughout the party, but in failing to speak up about it they too are wilfully exercising this vice. As we have seen, the trait of closed-mindedness itself prevents a closed-minded person from coming to realise that their mind is closed, even in a minimal sense. This demonstrates that even a weak self-awareness

requirement for bad epistemic motives (or motives away from an epistemic good) is at odds with the many stealthy vices.

To summarise this objection, the self-awareness requirement for commitments or motives is at odds with the standard view in vice epistemology: we are often unaware of our vicious motives. Even a weak, passive, awareness requirement contradicts our understanding of stealthy vices. Given these criticisms, I suggest the following amendments to Fricker's account of institutional vice.

7.4 Amendments

Starting with the joint motive concern first, there are various ways Fricker could respond to this objection and ultimately overcome the problem.

Firstly, we can assess Fricker's response to a similar concern raised by Lahroodi (2007) who objects to the self-awareness requirement and its application to collective virtues. Lahroodi (ibid.:292) points out the difficulty in offering a Gilbert-style joint commitment model of collective virtue due to the self-awareness requirement requiring that the subject of a virtue need not be aware of possessing it. They argue that this requirement is at odds with virtues such as open-mindedness, which an individual can be said to have without realising they possess the disposition to consider contrary views.

In response to this concern, Fricker argues that '…group members need not be aware of the virtuous nature of their jointly committed motives or skills' (Fricker 2010:248). By this, Fricker means a group need not be aware of their commitment as *virtuous.* Just as individuals may not conceive of their good motive as virtuous, it follows that collectives need not either. For example, a fair-minded jury might not be aware that their fairmindedness is a virtue, or a diligent research team might not be aware that their diligence is a virtue. Additionally, Fricker holds that a group need not be aware that they are reliably achieving the relevant end of a good motive or skill (ibid.:249). The fair-minded jury will just know that they are doing their best, but not that they are reliability achieving the ends of their motive. The self-awareness requirement can therefore be modified in this way, in which the collective is aware of their joint motive, but not that it is a virtuous one being reliability achieved.

This response does not absolve Fricker of the previous concern, however. Whilst it may be true that an institution need not be aware that its commitment is a virtue or vice, or that it is reliable, they still must be aware of their commitment. However, as we have seen, it is often not the case that vicious individuals or group members are aware that their motive is bad, or even that they possess a motive. Consequently, an awareness requirement on joint motives, even a minimal one, is at odds with current views in vice epistemology, particularly for stealthy or self-concealing vices.

With this response unsuccessful, I suggest that Fricker drops the self-awareness requirement of Gilbert's joint commitment for joint motives, meaning group members need not be aware of their commitments. Excluding this condition does not threaten the joint commitment model itself, allowing institutions to still act as a 'body' via their institutional ethos and display collective attributes. Additionally, with the previous arguments laid out, it seems far more fitting to our understanding of vice that group members are unaware of their motives towards or away from an epistemic good. Just as individuals are not aware of their epistemic motives, it seems only natural to assume the same is true of collectives.

This summarises the first alteration to Fricker's account in response to my previously raised concerns. By dropping the awareness requirement over joint commitments or motives the collective nature of institutions can still be retained and Fricker avoids the contradictions that come with claiming that vicious members need to be aware of their epistemically bad motives.

Moving on to the next amendment to Fricker's account of institutional vice, I argue that to ensure her account can accommodate both motivational and consequentialist explanations of epistemic vice, Fricker must adopt a stronger consequentialist claim.

As a reminder, a consequential account of vice emphasises the epistemic effects of an individual's or collective's behaviour, whether that be good (virtuous) ends or bad (vicious). It is these ends, not motives, which determine whether a virtue or vice has been displayed (Cassam 2016, 2019a). To map this onto institutions, an institution's joint commitment to X would need not be analogous to a joint motive to X. Rather, we should turn our attention to what X is, and whether it produces good or bad epistemic effects. In this sense, when a group

acts as a plural subject via a joint commitment to X, we can interpret X as a group virtue or vice by looking at what epistemic ends it produces and whether they are good or bad.

Under this view, what makes the behaviours or dispositions associated with X virtuous is if they consistently and culpably result in good epistemic effects. Conversely, what makes the behaviours or dispositions vicious is if they consistently and culpably result in bad epistemic effects. Finally, what makes the institution's behaviour vicious is that it results from epistemically bad consequences and is therefore blameworthy. What makes the institution's behaviour virtuous is that it produces good epistemic effects and is therefore praiseworthy. For example, we could determine whether an institution displays a vice such as conspiratorial thinking by assessing whether the member's commitment to their joint conspiracy theory consistently and culpably resulted in bad epistemic ends, such as misinformation.[148]

Taking these two amendments into consideration, we can now present the following, hybrid account of institutional virtue and vice.

With Gilbert's plural subject theory as the foundation, institutions can display virtues and vices via a joint commitment to V which the individual members of the institution need not possess and can often not be aware of. A joint commitment explains how an institution can act as a 'body' and exemplify an institutional ethos which reveals an institution's epistemic motives or behaviours. Under a motivationalist reading, these motives when directed towards an epistemic good for the right reason, constitute an epistemic virtue. A motive to an epistemic bad or away from an epistemic good constitutes an epistemic vice. Under a consequential reading, these behaviours are virtuous if they consistently and culpably result in good epistemic effects. Alternatively, these behaviours are vicious if they consistently and culpably result in bad epistemic effects.

To return to our remaining objections, this consequentialist reading directly responds to the concern that Fricker's account was not able to accommodate both motivational and consequentialist explanations of epistemic vice, as there is now a way to incorporate a consequentialist view for her account.

[148] See Cassam (2020) for a discussion on what makes certain ends characteristic of different vices.

Having made amendments to the consequentialist interpretation of institutional vice and dropped the self-awareness requirement on motivations, Fricker's account can now explain how institutional vices are vicious due to their bad motivations or bad epistemic effects.

With our revised account of institutional vice in hand, we can now move on to the next aim of this chapter which is to evaluate how an institution can be trustworthy, using this institutional virtue and vice framework.

7.5 Transparency and Trust

Like vice epistemology, work on trust also suffers from an individualistic bias, often focusing on the relationship between trustors and trustees or the features of a good or bad trustor or trustee (Baier 1986; Carter 2022; Goldberg 2020; Hawley 2014, 2019). For the same reasons that we should focus on collective virtues and vices, we should also focus on the question of what makes an institution trustworthy.

Having argued that institutions can display virtues and vices as an institution per se, let us now turn to how institutional vices can help us determine whether an institution is trustworthy or not by focusing on the institutional attributes associated with institutional trust. What are the attributes of an institution that can influence trust or distrust in that institution?[149]

One important feature of a trustworthy institution is *transparency*. Through an epistemic lens, transparency is understood as 'a tendency to faithfully share one's perspective on topics of others' inquiries with these others out of a motivation to promote their epistemic goods' (Byerly 2021:105).[150] Trust and transparency go hand-in-hand for institutions, particularly when it comes to gaining trust from non-experts and members of the public.

[149] As I address in the final section of this chapter, these virtues and vices may not always indicate trustworthiness or untrustworthiness but often do, all things considered. Institutional virtues and vices may serve as indicators of trust alongside other social indicators, such as status or authority (Oiggi 2022). Whether an institution is deemed trustworthy may also depend on the trustee too. For example, a trustee's own virtues and vices may influence their propensity to trust (Carter and Meehan 2019).

[150] I will be referring to the intellectual understanding of transparency throughout the remainder of this chapter.

Crucial to transparency is perspective. Whilst difficult to pin down, perspective can include a person's beliefs, intuitions, experiences, evidential standards and so on (Byerly 2023:291). When a person shares their perspective with another on the topic of that other's inquiry, they share their 'take' on that topic which goes beyond just expressing beliefs. The intellectually transparent person can be said to value others' attainment of epistemic goods such as knowledge, understanding, and true belief. Therefore, when sharing your perspective will lead to epistemic goods, the intellectually virtuous person is inclined to do so. When the reverse is true, the intellectually virtuous person is not motivated to share their perspective. For example, telling a friend about the low pass rate of an exam they are about to sit may make them feel defeated, resulting in them failing the test. In this case, the virtuously transparent friend should not share their perspective as it would not result in an epistemic good. Conversely, informing your friend about the success you recently had on this test may make them feel more confident and likely to pass.

It is therefore crucial to virtuous transparency to exercise this virtue in the right moment, at the right time, which is usually when doing so will result in epistemic goods.

When an intellectually transparent person 'faithfully' shares their perspective, Byerly (ibid.) observes two kinds of skills are being exercised. Firstly, the intellectually transparent person is good at figuring out what their own perspective is. For example, they can adeptly differentiate between a false claim and one they do not want to be true, whilst identifying the arguments and evidential standards that shape their views. Secondly, intellectual transparency involves effectively communicating one's perspective to others. This requires a sophisticated vocabulary to articulate certain distinctions, such as distinguishing between a belief and a non-belief or between possessing an argument for a claim's truth versus an argument for a claim's falsity. It also requires skill in facilitating other's understanding and appreciation of one's viewpoint.

Turning our attention to specific institutions, transparency and its opposing vices appear to be important features of a trustworthy or untrustworthy institution and have been found to directly increase degrees of trust in organisations (Rawlins 2008). This attribute is also essential to creating and maintaining trust in educational institutions, scientific institutions and governments (Anhalt-Depies et al. 2019; Kavanagh et al. 2020; Mabillard and Pasquier 2015; Nesset et al. 2021).

One example of an intellectually transparent institution is the government of the Mexican state of Nuevo León (UNESCO 2021). In response to the 2020 coronavirus pandemic, the state's ministry of finance created a microsite to communicate its 2020 budget and financial response. The government also formed part of a COVID-19 microsite that processes, systematizes, publishes, and disseminates information about the pandemic. One of its key aims was to generate reassurance amongst the public during the health crisis, with the platform making visible its strategies, actions and measures that were part of the public health policy adopted by the Federal Government in Mexico.

Conversely, the UK government is a recent example of an institution displaying the vice of deception (Manthorpe Rowland 2021).[151] In March 2021, the Open Government Partnership (OGP) – an international grouping of governments committed to openness and transparency – placed the UK government under review for its failure to deliver its pledge to improve transparency and accountability. One of the main issues that the government was criticised for was its neglect in providing freedom of information requests on time and a series of high-profile controversies on the government's failure to publish coronavirus-related contracts. This was heighted after Ministers were accused of favouring friends and political contacts for coronavirus work, including how they handled bids for personal protective equipment (PPE).[152]

By being transparent about their strategies, actions, and measures on COVID-19, the Mexican government were aware of what information needed to be made public and communicated it effectively in order to advance the public's epistemic status. Alternately, the UK government failed to be transparent to members of the public and instead acted deceivingly by withholding epistemic goods from members of the public.

Not only is transparency an attribute of a trustworthy institution but it can also be categorised as an institutional virtue. Likewise, transparency's counterpart vices e.g., bullshitting, lying and sloppiness can be categorised as institutional vices (Cassam 2019a; Lackey 2020).

[151] It is debatable whether withholding important counts as a form of deception. However, I argue in this case the government lied by omission.
[152] Another example of this is 'document dumping', where an organisation, particularly governments, buries important information amongst less significant information. For example, a government may introduce a controversial new policy during a busy news cycle. See Gardiner (2022) for more on this.

As we have seen, transparency involves a value of others' attainment of epistemic goods such as knowledge, understanding, and true belief. Transparency also involves a tendency to faithfully share your perspective with others to promote these kinds of goods. In this sense, an institution is transparent when they are oriented toward and motivated by promoting others' epistemic goods. For example, an organisation may be institutionally transparent if it openly communicates with customers and provides accurate and truthful information when relevant. This institution exemplifies institutional transparency by aligning its motives with the promotion of epistemic goods such as accurate information about its impact and operations.

On the other side of the spectrum, failures in institutional transparency can be vicious either when the institution is motivated away from the truth or towards deception e.g., our example of the UK government showing a motivation away from the truth (by actively lying). As Byerly (2022b:69) notes, when groups fail to share their perspectives with others or do so poorly, they can cause a variety of epistemic harm e.g., bullshitting, sloppiness, or deception (Cassam 2019a; Lackey 2020).

Likewise, institutional transparency can be an institutional virtue in so far as it consistently produces good epistemic effects. For example, transparent institutions build trust by reducing epistemic harms such as misinformation which can lead to distrust (Kavanagh et al. 2020), preventing corruption (Driscoll 1978; Heise 1985; Rawlins 2008) and reducing wrongdoings (Heald 2006).

Finally, bullshitting, sloppiness or deception are institutionally vicious as they consistently produce bad epistemic effects. An institution is vicious in this sense if it prevents important knowledge from entering the public domain and encourages a lenient attitude towards trust (Cassam 2019a; Lackey 2021; MacKenzie and Bhatt 2019).

To summarize, transparency involves skilfully attending to one's perspective and communicating this perspective to others as to advance others' epistemic goods. Transparency is an attribute of a trustworthy institution, making it an institutional virtue in so far as the institution possesses the right kind of motivations or produces good epistemic effects. Finally, when an institution consistently demonstrates a culpable lapse in this collective and/or in producing bad effects, institutional vices of bullshitting, lying and sloppiness can be displayed.

7.6 Honesty and Trust

Let us now turn to the next component of a trustworthy institution, honesty. Whilst predominantly discussed in a moral vein, honesty has a clear epistemic dimension (King 2021; Wilson 2018; Zagzebski 1996). Defined as a disposition to express the truth (as we see it) through our thought, speech, and behaviour, to avoid intentionally distorting the truth (as we see it), and to do so because we revere the truth and think it is valuable' (King 2021:145). Like transparency, honesty is prima facie tied to trust, important to building credibility in institutions and maintaining a public image and integrity (Pearce and Uridia 2015; Shapin 1995).[153]

Whilst distinct from moral honesty, both forms are concerned with an avoidance to distort the facts. For example, an honest person may not always tell the truth as they can be mistaken, however, what is crucial is that they do not intentionally aim to deceive as they must reliably intend to be truthful.

In this sense, honesty involves a deep motivation to avoid deception. By 'deep' motivation, Alan Wilson (2018:272) argues that we should not accept *any* motivation to avoid deception as sufficient for honesty. Instead, an agent's motivation must have the following features. Firstly, the motivation must be sufficiently persistent, meaning it is not fleeting or sporadic. Secondly, the motivation must be sufficiently strong meaning it can influence the agent's behaviour. Thirdly, the motivation must be robust enough to withstand competing considerations e.g., it outweighs a desire to keep a wallet that you found on a bus. When an agent's motivation has these three features, we can we say that they possess a deep motivation to avoid deception and can be considered 'truly honest' (Wilson 2018:272).

Furthermore, we can distinguish between moral and intellectual honesty by examining the types of motivations at play (King 2021; Wilson 2018). For Wilson, honesty involves a deep motivation to avoid deception, as opposed to a reliable behavioural disposition or a tendency to produce certain outcomes. Drawing from Zagzebski's (1996) account of epistemic virtues, Wilson claims that honesty can be categorised as intellectual when it is grounded in an

[153] For a discussion of the similarities and differences between transparency and honesty see Byerly (2022a).

underlying motivation to achieve cognitive contact with reality (2018:275). For Nathan King (2021:145), what makes honesty intellectual in form, is that it concerns *intellectual* motivations. An intellectually honest person is motivated by a desire to convey the truth and not *because* they care about truth and other epistemic goods. Alternatively, the morally honest person may not care about truth in the same way. Instead, they might be motivated to be truthful for non-epistemic reasons, such as a fear of punishment.

The sparse literature on intellectual honesty and dishonesty focuses exclusively on honesty between individuals and how it factors in the building of a relationship of trust. However, real examples can demonstrate how they are essential features of a trustworthy or untrustworthy institution.

One example of an honest institution is a media outlet providing a correcting statement on a previously published news story that contained factual inaccuracies. In this scenario, we can imagine that the publication is motivated by the desire for truth for the right reasons, i.e., because they care about truth and other epistemic goods, not just because they are worried about their declining audience ratings. Because of this, the publication is motivated by epistemic good for the right kind of reasons.

Conversely, BP, a British oil and gas company, is an example of an institution lacking the intellectual virtue of honesty. Exemplified by *The Deepwater Horizon Oil Spill*, this oil spill is regarded as one of the largest environmental disasters in world history, where in April 2010, an oil rig exploded in the Gulf of Mexico causing devastation and death for the wildlife, residents, and tourist industries (Bryant 2011). More than four million barrels of oil escaped into the Gulf of Mexico during the 87 days BP took to control the well. Following the spill, BP shares plunged by more than 40 percent in the weeks after the disaster, as it became clear the company could not immediately contain the spill (Fisk and Calkins 2016).

In 2012, BP and the United States Department of Justice resolved criminal charges by pleading guilty to 11 counts of manslaughter, two misdemeanours and a felony count for lying to the United States Congress on how much oil had spilt into the Gulf of Mexico following the rig's explosion. Internal documents and emails were found to contradict what BP released to the public and the US government, with the real spillage estimated to be at least 20 times higher

than what they had publicly stated.[154] The Deepwater Horizon disaster is a paradigm case of institutional intellectual dishonesty in so far as BP demonstrated a lack of respect for the truth and intentionally deceived the public and government on the actual figures of the oil spill which resulted in a loss of public trust (Jacques 2015).

Not only is honesty an attribute of a trustworthy institution, but it can also be categorized as an institutional virtue. Likewise, we have strong reason to believe that dishonesty is an institutional vice.

As we have seen, honesty is concerned with an avoidance to distort the facts. As outlined, an intellectually honest person is motivated by a desire to convey and not distort the truth because they care about truth and other epistemic goods. In this sense, an institution is honest when it is oriented toward and motivated by promoting others' epistemic goods and avoiding bad epistemic ends. For example, a news outlet may be institutionally honest if the journalists and editorial team are continuously driven by a sincere commitment to inform the public, adhere to journalistic standards and avoid distortion of facts.

On the other side of the spectrum from intellectual honesty, vicious institutional dishonesty occurs either when the institution is often motivated away from the truth or by a willingness to distort it e.g., a charity embellishing their performance reports with lies in order to secure further funding. This institutional dishonesty stems from a deviation from the epistemic good of providing accurate and true information.

Moving away from a motivationalist view, honesty can also be categorised as an institutional virtue in so far as it consistently produces good epistemic effects. Being honest promotes intellectual goods such as truth and other epistemic features such as trustworthiness, credibility, and integrity (Pearce and Uridia 2015; Shapin 1995). The news outlet can be institutionally honest from this consequentialist perspective if it provides accurate information, contributes to public understanding, and promotes truthful discourse.

Finally, from a consequentialist perspective, an institution is institutionally dishonest if it consistently produces bad epistemic effects. For example, if a charity's embellished

[154] Office of Public Affairs, US department of justice (2012) and Bryant (2011).

performance reports may lead to donors being misinformed about the charity's impact, or to a misallocation of resources (Miller 2017).

To summarise, honesty is a disposition to express the truth and to avoid intentionally distorting the truth through respect for epistemic goods. Honesty is an attribute of a trustworthy institution, making it an institutional virtue, in so far as the institution possesses the right kind of motivations or produces good epistemic effects. Finally, when an institution consistently demonstrates a culpable lapse in this collective and/or in producing bad effects, it can display the institutional vice of dishonesty.

So far, I have argued that the two dispositions of transparency and honesty are associated with institutional trust and can help us determine whether an institution is trustworthy or not. Defining these attributes and epistemic virtues under our hybrid account, I have argued that these virtues are important attributes of a trustworthy institution that can assist in the institutions' commitments to doing what the trustor trusts them to do, whether that be informing citizens on public health matters in the case of the Mexican government or lying about the extent of their damages in the case of the BP oil spill.

7.7 The Dangers of Transparency and Honesty

Having characterised transparency and honesty as institutional virtues and their corresponding traits as institutional vices, we can conclude by addressing two potential problems with this classification. I will first discuss a potential objection raised by Onora O'Neill (2002, 2006) and C. Thi Nguyen (2021) that transparency can lead to epistemic harms such as deception and is in deep opposition with trust. I will discuss two arguments presented in favour of this view, referred to as the epistemic intrusion argument and the intimate reasons argument. The next objection I will consider is offered by Stephen John (2018) who makes the similar claim that honesty can lead to epistemic harms, and in some cases, its counterpart vice of dishonesty may be more epistemically favourable. In turn, this claim also threatens my argument that honesty is a virtuous attribute of a trustworthy institution, as John claims that these communicative virtues do not always lead to an increase in public trust but may instead 'destroy' trust (John 2018:81).

Both objections therefore claim that both transparency and honesty are not always epistemically favourable and can instead undermine trust in an institution when exercised (Nguyen 2021; John 2018).

7.7.1 Transparency as Surveillance

Let us start focusing on the concerns with transparency first. O'Neill (2002, 2006) and Nguyen (2021) are both critical of the positive effects of transparency, specifically when it comes to being transparent towards non-experts. Whilst both authors do not explicitly focus on institutions, we can see how this will be particularly problematic for an institution given that many of them focus centrally on communicating ideas to the public e.g., universities, research centres and media companies.

Nguyen's main criticism is that promoting intellectual transparency and requiring experts to reveal information to the public makes experts act in non-epistemically favourable ways.[155] He presents two arguments for this conclusion. First is the epistemic intrusion argument: the drive to transparency forces experts to explain their reasoning to non-experts (2021:334). However, expert reasons are, by their nature, often inaccessible to non-experts. This means that the demand for transparency can pressure experts to make up false reasons or act only in those ways for which they can offer public justification. Second is the intimate reasons argument: in many cases of practical deliberation, the relevant reasons are intimate to a community and not easily explicable to those who lack a particular shared background (ibid.). The demand for transparency, then, pressures community members to abandon the unique understanding and sensitivity that arises from their particular experiences.

Consider, for example, a coding class that requires all students to disclose their gender and whether their gender is different from the sex they were assigned at birth. The aim of gathering this information is to monitor equality and diversity in the course. A transgender participant, not wanting to reveal their trans status for fear of discrimination, is uncomfortable with this demand for transparency and cannot communicate their reasons as to why without revealing

[155] Whilst Nguyen (2021) does not mention intellectual transparency by name, he focuses on its epistemic or intellectual dimension e.g., its role in communicating or exchanging epistemic goods such as knowledge.

sensitive and intimate information about themselves. Unable to explain their reasons why, the participant decides to drop the class. In this case, again, a pressure for transparency can generate several harms (including epistemic ones), by denying the trans student knowledge that they would have gained on this course. Here a need for transparency overrides a need for sensitivity and the intimate reasons behind certain decisions or behaviours.

Nguyen also argues that transparency amounts to a form of surveillance (2021:333-334). Demanding transparency may root out corruption, however, it also comes at the cost of inhibiting sensitivity and expertise and can amount to intrusive monitoring. This monitoring or surveillance can be justified when the overseen are likely to be corrupt or biased, and the overseers are careful, sensitive, and well-intentioned. However, it can be damaging when the overseen are 'skilled and good-hearted, and the overseers are unthinking, insensitive or inexpert' (ibid.:334). For example, demanding that a charity be transparent about its overhead costs may damage the charity by forcing them to make staff cuts and work from shoe-strong budgets (ibid.:342).

Also drawing attention to the integral relationship between trust and transparency, Nguyen argues that concerns about transparency highlight the essential tension between trust and transparency (2021:332). Drawing from O'Neill's (2002:73) claim that transparency encourages people to be dishonest, Nguyen discusses how in the face of transparency, acting dishonestly can be the best result.

To see this concern in action, consider an example of a research team working on a COVID-19 vaccination. The researchers are asked by the press of a sensationalist media company to reveal the potential side effects of the vaccine. The researchers are aware of the side effects, some of which are lethal, but occur extremely rarely in those who have been vaccinated. The researchers also know that revealing information on the rare, but potentially lethal, side effects will likely lead to a decrease in vaccine uptake. Because of this, they decide to downplay the side effects, intentionally missing off some of the extremely rare but lethal ones. In this case, the pressure for transparency forced the researchers to explain their reasoning to non-experts (the public). Resultingly, the researchers act deceivingly by excluding the potential side effects from their press release.

To summarise, Nguyen's objection raises two distinct concerns for my argument. Firstly, it demonstrates that transparency need not always be a feature of a trustworthy institution. Secondly, it highlights that sometimes acting viciously e.g., dishonestly, can be the more epistemically favourable outcome. On a broader note, requiring transparency in cases such as these means trust and transparency are often profoundly opposed. This is because of cases where transparency requires deception (as with our team of researchers) which in turn reduces trust (Onora O'Neill 2002). Nguyen also discusses how trust and transparency are in opposition because of the reasons cited by O'Neill and because transparency is a valuable indicator in determining when to distrust.

Despite being a useful tool in identifying corruption and bias, Nguyen therefore concludes that transparency 'is best as occasional intervention' (Nguyen 2021:258). At the very least we need to balance transparency with expertise, sensitivity, and awareness.

7.6.2 Dangerous Honesty

A similar concern can also be directed at honesty. Identified as a 'communicative' virtue, John (2018:75) argues that honesty is not always favourable and can be epistemically dangerous. Furthermore, there may be times when dishonesty is to be preferred in order to avoid bad epistemic consequences. John gives the following example to explain his position. Assume that a climate scientist knows that she could report a probability estimate to policymakers and doing so would ensure the policymakers act against climate change. However, she cannot 'own' this prediction, but at best 'offer' it because she is aware that her estimate is subject to significant second-order uncertainty. However, reporting these uncertainties would be more likely to lead to inaction on behalf of the policymakers (ibid.:83).

A proponent of honesty would argue that the scientist should still communicate the less precise estimate because doing otherwise is dishonest. However, John argues that if the scientist knows that reporting the estimate is likely to lead a policymaker to some conclusion which it is in her epistemic interests to believe (e.g., 'climate change will lead to ice sheet collapse') and a more 'honest' estimate is unlikely to lead to such belief, then she may be justified in making the first, less precise estimate.

In this case, dishonestly is preferred, as it leads to a more favourable outcome. Arguably then, in the scientist's role as an informant, she should make the favourable claim, rather than the honest claim.

From these types of instances, John argues that honesty is not always epistemically favourable. We should therefore at least be aware that these virtues may backfire and can be dangerous.

Turning to the relationship between honesty and trust, John presents the further example of 'Climategate', referred to as a high-profile case of 'enforced transparency', to explain this claim (ibid.:81). In 2009, the Climate Research Unit and the University of East Anglia were hacked, leaking emails that climate change deniers took as proof for the claim that climate change is a conspiracy.[156] Climate sceptics claim that the leaked emails showed that the climate scientists at UEA were engaged in non-scientific practices, for example, by confusing correlation and causation, refusing to include specific data sets in their analyses and by refusing to publish papers by particular authors (Bareham 2012).

However, the practices which sceptics described as 'unscientific' were standard and respectable (John 2018:81). For example, inferring causation from sufficient types and kinds of correlations is a justifiable scientific procedure (Papineau 2012) and refusing to publish some kinds of work is part of the 'dogmatism' necessary to promote 'progressive research projects' (Lakatos 1978:89-90).

John argues that in 'laying open the inner-workings of the climate change community' through transparency, openness and honesty did not increase public trust in climate scientists but instead resulted in dangerous, conspiracy-minded beliefs (John 2018:75).

In summary, honesty, like transparency, is not always an epistemically favourable characteristic that results in positive epistemic outcomes. Additionally, acting viciously, such as by being dishonest, may be epistemically advantageous. This undermines the status of honesty and transparency as institutional epistemic virtues and as reliable indicators of an institution's trustworthiness.

[156] Hickman and Randerson (2009).

7.7.2 Responses

Let us conclude by assessing how troublesome these concerns are for the arguments made in this chapter. Turning to the concerns raised with transparency first, one brief response is that many of Nguyen's criticisms are with transparency being demanded inappropriately. For example, demanding that a charity be transparent about its overhead costs which resulted in a poor working environment, is not an instance where transparency should be demanded in the first instance. This is not to say that a charity should not be transparent in some respects, but only those that are relevant and do not come at the cost of a dangerous working environment. Nuances in when transparency should be demanded do not necessarily imply that transparency itself is dangerous, however.

Similarly, it can be argued that 'dangerous' transparency in many of the above examples does not constitute *virtuous* transparency. For example, institutional transparency that results in bad epistemic effects e.g., disinformation or insensitivity would not be the type considered to be intellectually virtuous. Likewise, a form of transparency driven by an obligation to be transparent, despite knowing it will result in epistemic harm, is not motivated by the right kind of thing. A similar response also applies to John's (2018) claim that honesty and openness were harmful in the context of Climategate. Demanding honesty via hacking was inappropriate in this context, and it was arguably the misconstrued folk understanding of science that was at fault for the epistemic damage caused. The issue was not therefore with honesty per se.

Turning to Nguyen's (2021) second concern, that sometimes displays of vice such as dishonestly are more epistemically favourable than transparency, we can recall that virtues and vices must occur systematically (Cassam 2019a; Kidd 2016). This means that whilst institutional dishonesty may be epistemically favourable on some occasions, it is unlikely to do so systematically. For example, unless institutional dishonesty generates good epistemic effects more often than bad effects, it would not be a virtue of said institution.[157]

Again, this response can also be directed at John's (2018) concerns with honesty. Recalling John's example of the climate change scientist who exaggerated her findings, we can question

[157] See Chapter 2 for a discussion of this condition, detailed primarily by Cassam (2019a). Also see Battaly (2018, 2021) for an argument that some vices, specifically closed-mindedness may be a virtue on some occasions and a vice on others.

whether instances of *vicious* dishonesty were being displayed. As mentioned, the trait must occur systematically for it to be defined as such, and in John's example, it seems that their requirement to be dishonest for the greater epistemic good is not a frequent demand. Therefore, it can be true that the scientist is virtuously honest overall, but sometimes acts dishonestly when it will be epistemically favourable.

Appealing to a virtue or vice-theoretic understanding of honesty and dishonesty helps us overcome these consistency worries, given they must be reliable or systematic traits. We are not just concerned with whether an institution is honest or dishonest on some occasions, but whether they are consistently, or systematically so. This means that, on the whole, dishonesty is usually an attribute or indicator of an untrustworthy institution.

From these above responses, it follows that in defining transparency and honesty as virtues and their counterpart traits as vices, we can overcome concerns such as these where transparency and honesty appear to have epistemically damaging consequences, or their counterpart vices are the favourable options. For example, institutional transparency as a mere behaviour or trait may be in tension with trust, but *virtuous* institutional transparency is not. Virtuous institutional requires the behaviour to occur systematically and always aim at an epistemic good or generate a good epistemic end. This explains how one-off or infrequent displays of institutional transparency can sometimes lead to an epistemic bad but not be in tension with overall virtuous traits. The same is also true of vices such as institutional dishonesty. Whilst this may be epistemically favourable in some instances, more often than not when an institution is viciously dishonest, it is epistemically bad.

7.8 Conclusion

Having demonstrated how intellectual honesty and transparency are more likely to be attributes of trustworthy institutions, and their counterpart attributes of untrustworthy ones, we can now summarise the overall argument made in this chapter.

Drawing from Gilbert's plural subject theory and the literature on collective virtue and vice, we have seen that institutions can possess virtues and vices as a 'body' over and above its individual members (Fricker 2010, 2020; Lahroodi 2007:201). This joint commitment explains

how an institution can act as a ‘body’ and exemplify an institutional ethos which reveals an institution's epistemic motives or behaviours. Applying this model to institutions, with some modifications, I argued that by dropping a self-awareness requirement on commitments and by offering a true consequential reading of institutional virtue and vice, it followed that institutional attributes could be virtuous or vicious in relation to 1) their joint commitment to epistemic ends, where the commitment is understood as a motive to an epistemic good, to an epistemic bad, or away from an epistemic good, or 2) in relation to the epistemic end itself, where the attribute can result in good (virtuous) ends or bad (vicious) ends.

Having outlined how institutions can possess epistemic virtues and vices above individual-level vices of their members, I then focused on how institutions can be trustworthy or untrustworthy, by focusing on the attributes of transparency and honesty. These attributes and their counterparts were then defined as virtuous and vices pertaining to trustworthiness, as they mapped onto the previously identified definition of an institutional virtue and vice, either in their motivations or epistemic effects.

Starting with institutional virtues, we have seen that transparency involves being (persistently) motivated in the right kind of way for the epistemic good of truth (Byerly 2022a, 2023). When an institution demonstrates a culpable lapse in this collective motivation and/or in its performance implementation, institutional vices of bullshitting, lying, and sloppiness can be displayed. Under a motivationalist account, transparency can be an institutional virtue in so far as it consistently produces good epistemic effects (Kavanagh et al. 2020). Alternately, bullshitting, sloppiness or deception are vicious as they consistently produce bad epistemic effects, (Cassam 2019a; Lackey 2021; MacKenzie and Bhatt 2019).

Turning next to honesty, institutions can exercise this virtue via a (persistent) motivation to convey and not distort the epistemic good of truth for the right kind of reasons (King 2021). An institutional vice of dishonesty is displayed when an institution demonstrates a culpable lapse in this collective motivation and/or in its performance implementation. Likewise, honesty can be categorised as an institutional virtue in so far as it consistently produces good epistemic effects (Pearce and Uridia 2015; Shapin 1995) Finally, dishonesty is vicious in so far as it can consistently produce bad epistemic effects (Miller 2017).

Understanding what makes an institution trustworthy and untrustworthy is timelier now more than ever. Through a virtue-vice theoretic framework, I aim to have provided one answer to this question.

CHAPTER 8. CONCLUSION

8.1 Review of Critical Points

The overall aim of this thesis has been to investigate the harms of epistemic vices, and our possible responses to mitigate these negative impacts.

I divided this aim into three primary objectives. Let us remind ourselves what each of these objectives were and the key arguments associated with each.

First, under the theme of foundational vice epistemology, I aimed to investigate three prominent analyses of epistemic viciousness: obstructivism, motivationalism and personalism.

Beginning with obstructivism, in Chapter 2, I investigated two key components of Quassim Cassam's account (2016, 2019a), the obstructivist claim and the normative claim. Regarding the former, that character traits, attitudes and thinking styles can systematically obstruct knowledge and other epistemic goods, I argued that the systematic clause conflicted with Cassam's claim that vices can be low-fidelity, meaning they only occur occasionally. I proposed a modification to Cassam's account to deal with this concern, namely that the systematic clause should be concerned with the frequency in which epistemic harms occurred, not the frequency of the vice itself. Turning to the normative claim, that epistemic vices are blameworthy or criticisable, I raised concerns with Cassam's claim that we are not blameworthy for our epistemic vices if we lack control over the acquisition of them. I also raised objections to his preference for revision responsibility, arguing that it is unclear how this form of blame aligned with Cassam's aim for responsibility to serve an ameliorative aim.

Continuing in my analyses of epistemic viciousness, in Chapter 3, I turned to evaluate the motivational claim that epistemic vices involve bad epistemic motivations (Battaly 2016a, 2018a; Tanesini 2018, 2021). I began my investigation into Heather Battaly's account, personalism, exploring both its reliabilist and responsibilist features that define vices as personal qualities that we are not necessarily blameworthy for. I first focused on the scope of personalism and whether it was limited only to epistemic vices formed via indoctrination.

Following this, I focused my attention on Battaly's initial argument that vice-bearers cannot be blamed for vices that they lack control over the acquisition of.

Battaly's scepticism towards the possibility of holding vice-bearers responsible was twofold. Firstly, we can sufficiently explain the 'viciousness' of vices without relying on a responsibility component (2016:111-112, 2018a:122). Secondly, blame is not always a suitable response to vices, indicating that it cannot be an intrinsic feature (ibid.). To her first claim, I argued that responsibility is not intended as an explanation of the viciousness of vice but a response to said viciousness. Before turning to her second claim, I outlined Battaly's refinement of her initial position. Battaly was now open to a form of blame that did not require control, namely attributability responsibility. Crucially, however, this blame does not necessarily apply to all forms of epistemic vice. With this modification in mind, I argued that Battaly's primary reason to reject responsibility for vice, because it is incompatible with strong control conditions, was now resolved with this form of blame. Consequently, it became unclear as to why this form of blame could not be applied to all instances of epistemic vice.

In the second half of this chapter, I turned my attention to Alessandra Tanesini's (2018, 2021) motivationalist account of vice. This view held that epistemic vices consist of self-deceptive epistemic motives towards epistemic bads and away from epistemic goods (2018:350).

Tanesini was more optimistic about the prospects of blaming epistemic vices, claiming that vice-bearers can be held attributability responsible for their vices (2021:171). However, Tanesini opposed the responsibilist view that vices are inherently blameworthy for their vices due to a variety of practical and moral concerns with blaming vice-bearers (ibid.:182-183). Firstly, Tanesini argued that blaming a vice-bearer can amount to merely labelling the vice. In turn, this is counterproductive and can act as a self-fulfilling prophecy. Secondly, the lack of evidence to know if someone truly possesses the blameworthy means we cannot confidently blame others. Thirdly, we often lack the appropriate moral standing to blame vice-bearers due to hypocrisy or privilege.

In my evaluation of Tanesini's account, I raised concerns over the epistemic dimension of the blame that Tanesini proposed, arguing that its epistemic nature was unsubstantiated. I also responded to each of her practical and moral concerns on the possibility of blaming vice-bearers. To the first concern, I argued that an ameliorative account of blame which amounts to

more than mere vice-labelling, bypasses this concern. Secondly, I argued that we do not need to know that the vice-bearer possesses a vice per se, but just that they are displaying bad epistemic behaviour. Thirdly, I argued that Tanesini's argument was justifiable in instances where it would be hypocritical to blame. For the other occasions, an ameliorative account of blame sensitive to the factors that influence our epistemic character can still be applied.

At this point, I spoke to the bad reputation of blame, a recurring theme that runs through most vice epistemologists' dismissal of blame for vice. Blame is often considered to be a projection of guilt or shame, moralistic and high-minded, or inspired by vengeance (Fricker 2016:168; Owens 2012:25). However, some forms of blame, particularly communicative and functionalist blame, when employed appropriately, can be constructive and serve a positive aim.

This argument was explored further in Chapter 4, where I presented an interpretation of blame that is distinctively epistemic and ameliorative. On its epistemic nature, I argued that epistemic blame can be defined as a response to a violation of an epistemic norm, that aims to reduce bad epistemic conduct and bring about epistemic goods (Boult 2021; Fricker 2016; Piovarchy 2021; Sliwa *forthcoming*). I also discussed a variety of ways that blame can be directed towards ameliorative aims, drawing from feminist accounts of responsibility and functional and communicative accounts of blame (Ciurria 2021).

In the second half of this chapter, I responded to the named argument from lack of control that had been raised at various stages in the previous two chapters. In brief, this argument stated that we should not be blamed for epistemic vices that are outside of our control i.e., vices formed as the result of environmental factors which we could not change. I advocated for attributability responsibility here, demonstrating how this is a form of responsibility that does not require control, and coupled with my earlier arguments, can be a fair and productive response. At this stage, I also responded to an objection presented by Battaly (2019) towards attributability responsibility as a non-voluntary form of blame. Battaly's concern was that any non-voluntary account of responsibility for epistemic vice would either be too narrow and not capture implicit biases, or too broad and capture cognitive defects. To this, I argued that if implicit biases reflect some bad value or judgement, even subconsciously, they can be blameworthy. This is because attributability blame attaches to the attribute and the bad value or judgement that the attribute is the result of (Smith 2005:237, 254). I demonstrated how this

also ruled out cognitive defects from being blameworthy, for they do not reflect objectionable judgements.

Having defended an account of attributability blame which is both epistemic and ameliorative, I turned to explain how this form of blame necessary follows as a response to epistemic vices. I argued that the qualities which are worthy of blame under attributability responsibility align precisely with epistemic vices. Therefore, if we accept that we are attributability blameworthy for these qualities, we are blameworthy for our vices.

Moving beyond the foundational theme of the first half of the thesis, I then turned to explore arguments within applied vice epistemology. In Chapter 5, I assessed a potential ameliorative solution to the presence of epistemic vices. This was epistemic nudging, which involves nudging individuals towards epistemically desirable outcomes (Adams and Niker 2021; Grundmann, 2021:213; Miyazono 2023:2).

Despite initial signs of plausibility as an ameliorative solution, I argued that epistemic nudging's effectiveness lies primarily in masking vices rather than addressing their fundamental causes and 'deep' nature. I then argued that more concerningly, epistemic nudging may play a hand in creating further epistemic vices of epistemic injustice and intellectual laziness, as it bypasses and hinders one's epistemic capacities (Kidd 2017; Riley 2017,).

A further worry that arose as a result of this hindrance, was the concern that epistemic nudging violates our intellectual autonomy (Riley 2017). One response to this concern was that the other epistemic goods that come at the expense of this violation outweighed this concern e.g., better decision-making skills. However, based on my prior claim that epistemic nudging also can result in further epistemic vices, this response now seems dramatically weakened. Finally, I turn my attention to 'weak epistemic paternalism'. I argued that this form of nudging may address concerns to do with autonomy and is unlikely to result in further displays of vice. However, this form of nudging was too weak to make any significant changes to our epistemic behaviours.

In Chapter 6, attention then turned to the examination of epistemically corruptive environments, particularly those where information disorder is present. Following Ian Kidd's

(2019, 2020) framework for a successful corruption criticism, I outlined how the media, in so far as it produces information disorder, can be epistemically corrupting.

After applying this criticism, I concluded by examining different ameliorative solutions to these vices and their corruptive environments. I outlined three broad ameliorative solutions that one could take here. Firstly, there are individualistic ameliorative solutions, directed at the vice itself e.g., educating for virtues. Secondly, there are structural ameliorative solutions, directed at the systems and structures that facilitate and create vices e.g., changes to social media algorithms. Thirdly, there are social ameliorative solutions which focus on the impact that epistemic vices can have on the vice-bearer and their social communities e.g., fostering online other-regarding virtues. I argued that virtue-centric ameliorative solutions are not restricted to individual approaches, and instead, a coordinated approach that aims to develop individual and social virtues can go some way in addressing the outlined vices.

This thesis concluded with an exploration of institutional epistemic vices in Chapter 7, conceptualised through the lens of Margaret Gilbert's plural subject theory (1987, 2000, 2002, 2013) and Miranda Fricker's (2010, 2020) argument for an 'institutional ethos' (2020:90).

Applying this model to institutions, I assessed Fricker's account of institutional epistemic virtue and vice and outlined two modifications to her view. Firstly, I added a consequentialist modification to explain how institutional vices are vicious due to their bad motivations or bad epistemic effects, thereby aligning with the intended hybrid nature of her account. Secondly, I dropped the self-awareness requirement of Fricker's account that claimed that the institution's members must be aware of their commitments and motives. I argued that this requirement conflicted with our understanding of epistemic vices as stealthy and containing undetectable vicious motives (Cassam 2015, 2016, 2019a; Tanesini 2021).

This institutional perspective paved the way for an evaluation of trustworthy institutions based on communicative virtues such as transparency and honesty (Byerly 2022a; King 2021; Wilson 2018). I focused on how an institution can be trustworthy/untrustworthy by displaying the communicative attributes of transparency and honesty. I then explain how these attributes are virtues and vices of institutional virtues and vices, that are indicators of trustworthiness and untrustworthiness. I concluded this chapter by responding to the potential objection that transparency and honesty are not always beneficial attributes of an institution, nor indicators

of a trustworthy institution. Worse even, their counterpart vices may sometimes be required (John 2018; Nguyen 2021). In response to this concern, I argued that transparency and honesty as mere behaviours or traits can sometimes be in tension with trust, but virtuous transparency or honesty is not, which has stricter conditions on the systematic nature of traits.

In summary, this thesis has delved into various pivotal themes within vice epistemology. This ranged from foundational inquiries regarding the nature of blame, evaluating ameliorative solutions to counteract epistemic vice, exploring how polluted online environments can foster certain vices, and examining the nature of institutional vices and their connection to trust.

8.2 Further exploration

I invite these three areas for further exploration based on the arguments and analyses presented in this thesis. Given vice epistemology's real-world application to our epistemic practices, the further research invited in these below points may go beyond the scope of philosophy to include collaborative research with educators, policymakers, and other relevant stakeholders.

1. Having understood institutions as epistemically vicious, what other (potentially informal) groups can be said to exemplify epistemic vices?

2. How can we blend the individual, structural and social ameliorative solutions to vice? Practically, what could these solutions look like?

3. Understanding epistemic vices as blameworthy qualities, how can we ensure that individuals follow an ameliorative account of blame when 'charging' others and institutions?

Addressing the first point, the exploration of collective vices within the context of institutions opens up a fascinating area for further exploration, promoting inquiries into the nature of vices at the collective level and how they manifest within groups distinct from institutions. Whilst this thesis delved into how institutions, understood through Gilbert's plural subject theory (1987, 2000, 2002, 2013), can collectively possess virtues and vices, extending this inquiry

into groups provides a rich terrain for further work.[158] This could involve investigating the dynamics of collective vices within groups and understanding how shared values, norms and attitudes contribute to the formation and perpetuation of epistemic vices. In particular, vice epistemologists may explore how whether certain group dynamics such as conformity pressures, groupthink or the influence of charismatic leaders could contribute to the development of collective vices.

Additionally, the study of collective vices in groups could involve an analysis of the role of communication and information sharing within these settings. Examining how epistemic vices are communicated, reinforced, and challenged in group interactions can shed light on the social dynamics that either foster or inhibit epistemic vices and virtues. This exploration might encompass discussion on shared narratives, echo chambers and the dynamics of epistemic trust within groups. We may also ask what ameliorative solutions are best directed at these vices, and whether they differ from strategies aimed at individual vice-bearers.

Turning to the second point, a further area for exploration concerns the different ameliorative approaches to epistemic vice. I distinguished between three such approaches in Chapter 5. First, there are individualistic ameliorative solutions aimed at the individual vice-bearer. Second, there are structural ameliorative solutions that are directed at the systems and structures that instigate vices. And third, there are social ameliorative solutions which focus on the impact that epistemic vices can have on the vice-bearer and social communities. All three approaches invite further investigation, as well as how all three can collaborate to provide a multifaceted solution to epistemic vice. I will briefly survey a few avenues where such solutions may be found:

1. Educational intervention: explore the design and implementation of educational interventions. This could include curricular developments that enhance critical thinking and discussions on the communal impact of vices.

2. Technological solutions: investigate how technological tools and platforms can be leveraged to integrate individual, structural, and social approaches. This could

[158] See Broncano-Berrocal and Carter (2021); Iizuka (2023) Medina (2021) and Miyazono (2023) for some recent contributions to this field.

involve developing algorithms or features within social forums and sites to promote epistemic virtues.

3. Epistemic activism: communities can uproot institutional vices by denouncing and resisting the unfair patterns of epistemic harms. They may also express solidarity with one another to counteract the harmful consequences that may result from individual protest.[159]

Relatedly, a third area for further exploration is the ameliorative nature of blame. Crucially, having defined vices as blameworthy, how can we ensure that individuals prescribe to an ameliorative account of blame? This may involve delving into the psychological, ethical, and educational dimensions of blame.

For example, turning to motivational psychology, we can investigate the motivational factors that influence how individuals resist or embrace blame. We can examine how intrinsic and extrinsic motivations, such as a desire for personal growth or social approval may encourage individuals to engage with ameliorative practises. Additionally, we can explore the cognitive biases at play that may hinder blame in an ameliorative framework. Addressing biases such as defensive reasoning and the self-serving bias is critical for fostering more open and reflective attitudes towards one's vices.

Returning to educational strategies, we can also explore how dialogical pedological approaches may facilitate discussions on blame and responsibility.[160] Creating spaces for open reflection, both in formal educational settings and informal communities, could enhance a vice-bearer's engagement with ameliorative practices. These discussions may also feed into current debates on online responsibility and 'cancel-culture'. For example, what does it really mean to call out an internet troll for their bad behaviour online? Is cancel culture an ameliorative blame practice? And how do we balance blame with issues surrounding epistemic silencing?

By addressing these dimensions to ameliorative blame and our bad epistemic behaviour, researchers, educators, and policymakers can contribute to the development of strategies that

[159] See Medina (2020:120) and Medina and Whitt (2022) for a discussion of how epistemic activism can play a part in the mitigation of epistemic vices.
[160] See Battaly (2022).

not only define epistemic vices as blameworthy but also ensure vice-bearers, whether that be individuals or institutions, actively embrace and adhere to the ameliorative practises of blame in their intellectual pursuits. This points to the wider importance and application of vice epistemology, within and beyond academia.

Bibliography

Adams, M. & Niker, F. 2021. Harnessing the epistemic value of crises for just ends. In: F. Niker & A. Bhattacharya (eds.). *Political Philosophy in a Pandemic: Routes to a More Just Future.* London: Bloomsbury Academic. pp. 219-232.

Ahlstrom-Vij, K. 2013. *Epistemic Paternalism: A Defence*. Basingstoke: Palgrave Macmillan.

Alfano, M., Carter, J. A. & Cheong, M. 2018. Technological Seduction and Self Radicalisation. *Journal of the American Philosophical Association,* 4(3), pp. 298–322.

Alfano, M. 2013. *Character as Moral Fiction*. Cambridge: Cambridge University Press.

Allen J., Howland, B., Mobius, M., Rothschild, D., & Watts, D. J. 2020. Evaluating the fake news problem at the scale of the information ecosystem. *Sci Adv*, 6(14), pp. 1-6.

Anhalt-Depies, C., Stenglein, J. L., Zuckerberg, B., Townsend, P. A. & Rissman, A. R. 2019. Tradeoffs and tools for data quality, privacy, transparency, and trust in citizen science. *Biological Conservation,* 238, pp. 777-780.

Aristotle. 1985. *Nichomachean Ethics.* Translated by T. Irwin, 1985. Indianapolis and Cambridge: Hackett Publishing Company.

Aristotle. 2007. *On Rhetoric: A Theory of Civic Discourse.* Translated by G. A. Kennedy, 2007. 2nd ed. New York and Oxford: Oxford University Press.

Arpaly, N. 2003. *Unprincipled Virtue: An Inquiry Into Moral Agency*. Oxford: Oxford University Press.

Baehr, J. 2010. Epistemic malevolence. *Metaphilosophy,* 41(1-2), pp. 189-213.

Baehr, J. 2011. *The Inquiring Mind: On Intellectual Virtues and Virtue Epistemology*. Oxford: Oxford University Press.

Baehr, J. 2013. Educating for Intellectual Virtues: From Theory to Practice. *Journal of Philosophy and Education,* 47(2), pp. 248-262.

Baehr, J. 2016. *Intellectual Virtues and Education: Essays in Applied Virtue Epistemology*. London: Routledge.

Baehr, J. 2020. The structure of intellectual vices. In: I. J. Kidd, H. Battaly, & Q. Cassam (eds). *Vice Epistemology*. New York: Routledge.

Baier, A. 1986. Trust and Antitrust. *Ethics,* 96(2), pp. 231–260.

Baird, C. & Calvard, T. S. 2019. Epistemic vices in organizations: knowledge, truth, and unethical conduct. *Journal of Business Ethics,* 160, pp. 263–276.

Bakamo. 2017. Patterns of Disinformation in the 2017 French Presidential Election. [pdf] Available at <https://www.bakamosocial.com/wp-content/uploads/2021/03/PatternsofDisinformationinthe2017FrenchPresidentialElection-Report2-Bakamo.pdf> [Accessed April 2023].

Baron, J. 1985. *Rationality and Intelligence*. Cambridge: Cambridge University Press.

Bareham, D. 2020. Hackers breached password security to steal UEA climate change emails. Eastern Daily Press [online]. Available at: https://www.edp24.co.uk/news/crime/hackers-breached-password-security-to-steal-uea-climate-change-emails-527048 [Accessed on 31st August 2022].

Barua, Z., Barua, S., Aktar, S., Kabir, N. & Li, M. 2020. Effects of misinformation on COVID-19 individual responses and recommendations for resilience of disastrous consequences of misinformation. *Progress in Disaster Science* [e-journal], 8. https://doi.org/10.1016/j.pdisas.2020.100119.

Battaly, H. 2013. Detecting Epistemic Vice in Higher Education Policy: Epistemic Insensibility in the Seven Solutions and the REF. *Journal of Philosophy of Education*, 47(2), pp. 263-280.

Battaly, H. 2014. Varieties of Epistemic Vice. In: J. Matheson and R. Vitz (eds.). *The Ethics of Belief.* Oxford: Oxford University Press. Ch.3.

Battaly, H. 2015. Responsibilist Virtues in Reliabilist Classrooms. In: J. Baehr (ed). *Intellectual Virtues and Education: Essays in Applied Virtue Epistemology.* London: Routledge. pp. 163–187.

Battaly, H. 2016a. Epistemic Virtue and Vice: Reliabilism, Responsibilism, and Personalism. In: C. Mi, M. Slote & E. Sosa (eds.). *Moral and Intellectual Virtues in Western and Chinese Philosophy*. New York: Routledge. pp. 99–120.

Battaly, H. 2016b. Developing virtue and rehabilitating vice: Worries about self-cultivation and self-reform. *Journal of Moral Education*, 45(2), pp. 207-222.

Battaly, H. 2017a. Testimonial Injustice, Epistemic Vice, and Vice Epistemology. In: I. J. Kidd, J. Medina & G. Pohlhaus Jr. (eds.). *The Routledge Handbook to Epistemic Injustice*. New York: Routledge. pp. 223–231.

Battaly, H. 2017b. Intellectual Perseverance. *Journal of Moral Philosophy*, 14(6), pp. 669-697.

Battaly, H. 2018a. A Third of Kind Intellectual Virtue: Personalism. In: H Battaly (ed.). *The Routledge Handbook of Virtue Epistemology.* Abingdon: Routledge. Ch.10.

Battaly, H. 2018b. Can Closed-Mindedness Be an Intellectual Virtue? *Royal Institute of Philosophy Supplement,* 84, pp. 23–45.

Battaly, H. 2019. Vice epistemology has a responsibility problem. *Philosophical Issues,* 1, pp. 24-36.

Battaly, H. 2021. Engaging closed-mindedly with your polluted media feed. In: M. Hannon & J. de Ridder (eds.). *The Routledge Handbook of Political Epistemology*. New York: Routledge. pp. 312–324.

Battaly, H. 2022. Measuring and mismeasuring the self. *Inquiry*, 67(2), pp. 738-761.

Beckett, L. 2020. Nearly all Black Lives Matter protests are peaceful despite Trump narrative, report finds. *The Guardian* [online] 5 September. Available at <https://www.theguardian.com/world/2020/sep/05/nearly-all-black-lives-matter-protests-are-peaceful-despite-trump-narrative-report-finds> [Accessed November 2022].

Begby, E. 2013. The Epistemology of Prejudice. *Thought: A Journal of Philosophy,* 2(2), pp. 90-99.

Bird, A. 2019. Group Belief and Knowledge. In: M. Fricker, Graham, P. J., Henderson, D. & N. J. J. L. Pedersen (eds.). *The Routledge Handbook of Social Epistemology.* New York: Routledge. Ch.27.

Bland, S. 2022a. Interactionism, Debiasing, and the Division of Epistemic Labour. In: M. Alfano, C. Klein & J. de Ridder (eds.). *Social Virtue Epistemology*. New York: Routledge. pp. 15-38.

Bland, S. 2022b. An Interactionist Approach to Cognitive Debiasing. *Episteme*, 19(1), pp. 66-88.

Blenkinsopp, J. & Park, H. 2011. The roles of transparency and trust in the relationship between corruption and citizen satisfaction. *International Review of Administrative Article Sciences,* 77(2), pp. 254-274.

Bontcheva, K. & Posetti, J. 2020. DISINFODEMIC: Dissecting responses to COVID-19 disinformation, Policy brief 2 [pdf]. United Nations Educational, Scientific, and Cultural Organization. <https://en.unesco.org/sites/default/files/disinfodemic_dissecting_responses_covid19_disinformation.pdf> [Accessed July 2022].

Boult, C. 2020. There is a distinctively epistemic kind of blame. *Philosophy and Phenomenological Research*, 103, pp. 518-534.

Boult, C. 2021. Epistemic blame. *Philosophy Compass* [e-journal], 16(8). https://doi.org/10.1111/phc3.12762.

Bonjour, L. 2001. Externalist Theories of Empirical Knowledge. In: H. Kornblith (ed.). *Epistemology: Internalism and Externalism.* Oxford: Blackwell. pp. 10–35.

Brown, J. 2017. Blame and wrongdoing. *Episteme*, 14(3), pp. 275-296.

Broncano-Berrocal, F. & Carter, J. A. 2021. *The Epistemology of Group Disagreement*. Abingdon-On-Thames: Routledge.

Brown, J. 2020. Epistemically blameworthy belief. *Philosophical Studies*, 177, pp. 3595–3614.

Bullock, E. C. 2016. Knowing and Not-Knowing for Your Own Good: The Limits of Epistemic Paternalism. *Journal of Applied Philosophy*, 35(2), pp. 433–447.

Burrell, I. 1999. Lord Denning, the Century's Greatest Judge, Dies at 100. *Independent* [online] 6 March. Available at <https://www.independent.co.uk/news/lord-denning-the-century-s-greatest-judge-dies-at-100-1078587.html> [Accessed May 2020].

Byerly, T. R. 2022a. Intellectual Honesty and Intellectual Transparency. *Episteme*, 20(2), pp. 410-428.

Byerly, T. Ryan. 2022b. Group intellectual transparency: a novel case for non-summativism. *Synthese* [e-journal], 200(2): 69. https://doi.org/10.1007/s11229-022-03617-x.

Byerly, T. R. 2023. The Values of Intellectual Transparency. *Social Epistemology*, 37(3), pp. 290–304.

Byerly, T. R. & Byerly, M. 2016. Collective virtue. *Journal of Value Inquiry*, 50(1), pp. 33–50.

Card, C. 1996. *The Unnatural Lottery: Character and Moral Luck.* Philadelphia: Temple University Press.

Carter, J. A. 2022. Trust and Trustworthiness. *Philosophy and Phenomenological Research*, 2, pp. 377-394.

Carter, J. A. & Meehan, D. 2019. Vices of distrust. *Social Epistemology Review and Reply Collective*, 8(10), pp. 25-32.

Cassam, Q. 2010. Judging, Believing and Thinking. *Philosophical Issues*, 20, pp. 80–95.

Cassam, Q. 2015. Stealthy Vices. *Social Epistemology Review and Reply Collective*, 4, pp. 19-25.

Cassam, Q. 2016. Vice Epistemology. *The Monist,* 99(2), pp. 159-180.

Cassam, Q. 2018. Epistemic Insouciance. *Journal of Philosophical Research,* 43, pp. 1-20.

Cassam, Q. 2019a. *Vices of the mind: from the intellectual to the political*. Oxford: Oxford University Press.

Cassam, Q. 2019b. *Conspiracy Theories*. Cambridge: Polity Press.

Cassam, Q. 2020. The Metaphysical Foundations of Vice Epistemology. In: I. J. Kidd, H. Battaly, & Q. Cassam (eds). *Vice Epistemology*. New York: Routledge. Ch.2.

Ceva, E. & Bocchiola, M. 2019. Personal Trust, Public Accountability, and the Justification of Whistleblowing. *The Journal of Political Philosophy,* 27(2), pp. 187-206.

Ceva, E. & Ferretti, M. P. 2021. What political corruption is. In: E. Ceva & M. P. Ferretti (eds.). *Political Corruption: The Internal Enemy of Public Institutions*. Oxford: Oxford Academic. Ch.1.

Ciurria, M. 2020. *An Intersectional Feminist Theory of Moral Responsibility.* Abingdon: Routledge.

Coady, D. 2019. The Trouble With 'Fake News'. *Social Epistemology Review and Reply Collective,* 8(10), pp. 40-52.

Code, L. 1984. Toward a `Responsibilist' Epistemology. *Philosophy and Phenomenological Research*, 45(1), pp. 29–50.

Code, L. 1987. *Epistemic Responsibility.* Rhode Island: Brown University Press.

Crerar, C. 2018. Motivational Approaches to Intellectual Vice. *Australasian Journal of Philosophy,* 96(4), pp. 753-766.

Croce, M. & Piazza, T. 2021. Consuming Fake News: Can We Do Any Better? *Social Epistemology*, 37(2), pp. 1–10.

Croce, M. & Pritchard, D. 2022. Education as The Social Cultivation of Intellectual Virtue. In: M. Alfano, C. Klein & J. de Ridder (eds.). *Social Virtue Epistemology*. London: Routledge. pp. 583-601.

Cronin Fisk, M. & Brubaker Calkins, L. 2016. BP pays $175m to settle claims it hid size of Gulf of Mexico spill. *The Independant* [online] 6 June. Available at: <https://www.independent.co.uk/news/business/news/bp-pays-175m-to-settle-claims-it-hid-size-of-gulf-of-mexico-spill-a7063066.html> [Accessed November 2022].

Dahlgren, P. 2018. Media, Knowledge and Trust: The Deepening Epistemic Crisis of Democracy. *Javnost - The Public: Journal of the European Institute for Communication and Culture*, 25(1-2), pp. 20-27.

Davies, B., Lalot, F., Peitz, L., Heering, M. S., Ozkecei, H., Babaian, J., Davies Hayon, K., Broadwood, J. & Abrams, D. 2021. Changes in political trust in Britain during the COVID-19 pandemic in 2020: integrated public opinion evidence and implications. *Humanit Soc Sci Commun*, 8, p. 166.

de Ridder, J. 2019. So What if 'Fake News' is Fake News? *Social Epistemology Review and Reply Collective,* 8(10), pp. 111-113.

Derakhshan, H. & Wardle, C. 2017. Information Disorder: Definitions. In: *Proceedings of Understanding and Addressing the Disinformation Ecosystem*. Annemberg: University of Pennsylvania. pp. 5-12.

Dillon, R. 2012. Critical Character Theory: Toward a Feminist Perspective on 'Vice' (and 'Virtue'). In: S. L. Crasnow & A. M. Superson (eds.). *Out from the Shadows: Analytical Feminist Contributions to Traditional Philosophy.* New York: Oxford University Press. pp. 83–114.

Doris, J. 2002. *Lack of Character: Personality and Moral Behaviour*. Cambridge: Cambridge University Press.

Dougherty, T. 2012. Reducing Responsibility: An Evidentialist Account of Epistemic Blame. *European Journal of Philosophy,* 20, pp. 534-547.

Driscoll J. W. 1978. Trust and participation in organizational decision making as predictors of satisfaction. *Academy of Management Journal,* 21(1), pp. 44–56.

Driver, J. 2001. *Uneasy Virtue*. New York: Cambridge University Press.

Dunne, G. 2021. Epistemic Paternalism. In: H. LaFollette (ed.). *International Encyclopaedia of Ethics*. MA: Wiley-Blackwell.

Dworkin, G. 2020. Paternalism. In: E. N. Zalta (ed.). *The Stanford Encyclopaedia of Philosophy*. Available at <https://plato.stanford.edu/entries/paternalism/> [Accessed April 2022].

Dyer, C. 1999. Lord Denning, Controversial "People's Judge", Dies Aged 100. *The Guardian* [online] 6 March. Available at <https://www.theguardian.com/uk/1999/mar/06/claredyer1> [Accessed 5th November 2021].

Edelman. 2020. 2020 Edelman Trust Barometer [online] Available at: <https://www.edelman.com/trust/2020-trust-barometer> [Accessed 31st August 2022].

Enoch, D. 2016. II—What's Wrong with Paternalism: Autonomy, Belief, and Action. *Proceedings of the Aristotelian Society,* 116(1), pp. 21–48.

Ernst, E. 2001. Rise in Popularity of Complementary and Alternative Medicine: Reasons and Consequences for Vaccination. *Vaccine,* 20(1), pp. 90–93.

Frankfurt, H. 1971. Freedom of the Will and the Concept of a Person. *The Journal of Philosophy,* 68(1), pp. 5–20.

Freeman, I. 1993. *Lord Denning: A Life*. London: Hutchinson.

Fricker, M. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.

Fricker, M. 2010. Can There Be Institutional Virtues? In: T. Gendler & J. Hawthorne. (eds.). *Oxford Studies in Epistemology*. Oxford: Oxford University Press. Ch.10.

Fricker, M. 2016., What's the Point of Blame? A Paradigm Based Explanation. *Noûs,* 50, pp. 165-183.

Fricker, M. 2020. Institutional epistemic vices: the case of inferential inertia. In: I. J. Kidd, H. Battaly, & Q. Cassam (eds.). *Vice Epistemology*. Abingdon-On-Thames: Taylor & Francis. Ch.5.

Gardiner, G. 2022. Attunement: On the Cognitive Virtues of Attention. In: M. Alfano, C. Klein & J. de Ridder. (eds). *Social Virtue Epistemology.* London: Routledge. pp. 48-72.

Gelfert, A. 2018. Fake News: A Definition. *Informal Logic*, 38(1), pp. 84-117.

Gerbina, T. V. 2021. Science Disinformation: On the Problem of Fake News. *Sci. Tech. Inf. Proc*, 48, pp. 290–298.

Giansiracusa, N. 2021. *How Algorithms Create and Prevent Fake News: Exploring the Impacts of Social Media, Deepfakes, GPT-3, and More.* CA: Apress Berkeley.

Gilbert, M. 1987. Modelling Collective Belief. *Synthese*, 73(1), pp. 185–204.

Gilbert, M. 2000. *Sociality and Responsibility: New Essays in Plural Subject Theory*. Maryland: Rowman and Littlefield.

Gilbert, M. 2002. Belief and Acceptance as Features of Groups. *Protosociology*, 16, pp. 35–69.

Gilbert, M. 2013. *Joint Commitment: How We Make the Social World*. Oxford: Oxford University Press.

Greco, J. 1993. Virtues and Vices of Virtue Epistemology. *Canadian Journal of Philosophy*, 23(3), pp. 413–432.

Greco, J. 2002. Virtues in Epistemology. In: P. Moser (ed.). *The Oxford Handbook of Epistemology.* Oxford: Oxford University Press. pp. 287-312.

Greco, J. 2010. *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity.* Cambridge: Cambridge University Press.

Greenwald, A. G,. Banaji, M. R. & Nosek, B. A. 2015. Statistically Small Effects of the Implicit Association Test Can Have Societally Large Effects. *Journal of Personality and Social Psychology*, 108(4), pp. 553-561.

Grundmann, T. 2021. The Possibility of Epistemic Nudging. *Social Epistemology*, 37(2), pp. 208-218.

Goldberg, S. C. 2020. Trust and Reliance 1. In: J. Simon (ed.). *The Routledge Handbook of Trust and Philosophy.* New York: Routledge. Ch.8.

Goldman, A. I. 1978. Epistemics: The Regulative Theory of Cognition. *The Journal of Philosophy*, 75(10), pp. 509–523.

Habgood-Coote, J. 2019. Stop Talking About Fake News!. *Inquiry: An Interdisciplinary Journal of Philosophy*, 62(9-10), pp. 1033-1065.

Hartley, K. & Vu, M. K. 2020. Fighting fake news in the COVID-19 era: policy insights from an equilibrium model. *Policy Sci*, 53, pp. 735–758.

Haslanger, S. 2015. Social Structure, Narrative and Explanation. *Canadian Journal of Philosophy*, 45, pp. 1–15.

Hausman, D. M. & Welch, B. 2010. Debate: to nudge or not to nudge. *Journal of Political Philosophy*, 18(1), pp. 123–136.

Hawley, K. 2014. Trust, Distrust and Commitment. *Nous*, 48(1), pp. 1-20.

Hawley, K. 2019. *How to Be Trustworthy*. Oxford: Oxford University Press.

Heise J. A. 1985. Toward closing the confidence gap: An alternative approach to communication between public and government. *Public Affairs Quarterly*, 9(2), pp. 196–217.

Heersmink, R. 2018. A Virtue Epistemology of the Internet: Search Engines, Intellectual Virtues and Education. *Social Epistemology*, 32(1), pp. 1-12.

Hickman, L. & Randerson, J. 2009. Climate sceptics claim leaked emails are evidence of collusion among scientists. *The Guardian* [online] 20 November. Available at: <https://www.theguardian.com/environment/2009/nov/20/climate-sceptics-hackers-leaked-emails> [Accessed November 2022].

Hieronymi, P. 2004. The Force and Fairness of Blame. *Philosophical Perspectives*, 18(1), pp. 115–148.

Hieronymi, P. 2006. Controlling Attitudes. *Pacific Philosophical Quarterly*, 87, pp. 45–74

Hieronymi, P. 2008. Responsibility for Believing. *Synthese*, 161(3), pp. 357–373.

Hieronymi, P. 2014. Reflection and Responsibility: Reflection and Responsibility. *Philosophy & Public Affairs*, 42(1), pp. 3–41.

Higgins, L. 2016. More Michigan Parents Willing to Vaccinate Kids. Detroit Free Press. [online] Available at: <https://eu.freep.com/story/news/ education/2016/01/28/immunization-waivers-plummet-40-michigan/79427752/> [Accessed June 16th 2019].

Holroyd, J. 2012. Responsibility for Implicit Bias. *Journal of Social Philosophy*, 43(3), pp. 274-306.

Holroyd, J. & Kelly, D. 2016. Implicit Bias, Character, and Control. In: A. Masala & J. Webber (eds.). *From Personality to Virtue: Essays on the Philosophy of Character*. Oxford: Oxford University Press. Ch.5.

Holroyd, J., Scaife, R. & Stafford, T. 2017. Responsibility for implicit bias. *Philosophy Compass*, 12(3), pp.1-13.

Holroyd, J. 2020. Implicit Bias and Epistemic Vice. In: I. J. Kidd, H. Battaly, & Q. Cassam (eds.). *Vice Epistemology.* New York: Routledge. Ch.7.

Hookway, C. 1994. Cognitive Virtues and Epistemic Evaluations. *International Journal of Philosophical Studies*, 2, pp. 211–27.

Hurka, T. 2001. *Virtue, Vice, and Value*. Oxford: Oxford University Press.

Hutchison, K. 2018. Moral responsibility, respect and social identity. In: K. Hutchison, C. Mackenzie & M. Oshana (eds.), *Social Dimensions of Moral Responsibility.* Oxford: Oxford University Press.

IPSOS. 2019. *Trust - The Truth?* [pdf] Available at: <https://www.ipsos.com/sites/default/files/ct/publication/documents/2019-09/ipsos-thinks-trust-the-truth.pdf> [Accessed 31st August 2022].

Ireton, C. & Posetti, J. 2018. *Journalism, fake news & disinformation: handbook for journalism education and training*. Paris: Unesco.

John, S. 2018. Epistemic trust and the ethics of science communication: against transparency, openness, sincerity and honesty. *Social Epistemology*, 32(2), pp. 75-87.

Johnston, M. 1972. How to Speak of the Colors. *Philosophical Studies*, 68, pp. 221–263.

Joly Chock, V. 2020. The Ethics and Applications of Nudges. *PANDION: The Osprey Journal of Research & Ideas,* 1(2), Article 5.

Jones, K. 2012. Trustworthiness. *Ethics*, 123(1), pp. 61–85.

Kahneman, D. & Tversky, A. 1972. Subjective Probability: A Judgment of Representativeness. *Cognitive Psychology*, 3(3), pp. 430–454.

Kallestrup, J. 2020. Group virtue epistemology. *Synthese*, 197(12), pp. 5233-5251.

Kavanagh, J., Carman, K. G., DeYoreo, M., Chandler, N. & Davis, L. E. 2020. *The Drivers of Institutional Trust and Distrust: Exploring Components of Trustworthiness*. Santa Monica: Rand.

Kawall, J. 2002. Other–Regarding Epistemic Virtues. *Ratio, 15(3)*, pp. 257-275.

Kelp, C. 2019. The knowledge norm of blaming. *Analysis*, 80(2), pp. 256–261.

Kelp, C. & Simion, M. 2023. What is trustworthiness? *Noûs*, 57(3), pp. 667-683.

Kidd, I. J. 2016. Charging Others with Epistemic Vice. *The Monist,* 99(3), pp. 181-197.

Kidd, I. J. 2017. Capital Epistemic Vices. *Social Epistemology Review and Reply Collective,* 6 (8), pp. 11–16.

Kidd, I. J. 2018. Deep Epistemic Vices. *Journal of Philosophical Research,* 43, pp. 43–67.

Kidd, I. J. 2019. Epistemic Corruption and Education. *Episteme,* 16(2), pp. 220-235.

Kidd, I. J. 2020. Epistemic Corruption and Social Oppression. In: I. J. Kidd, H. Battaly, & Q. Cassam (eds.). *Vice Epistemology*. New York: Routledge. pp. 69–86.

Kidd, I. J. 2021. Epistemic Corruption and Political Institutions. In: M. Hannon & J. de Ridder (eds.) *The Routledge Handbook to Political Epistemology*. London: Routledge. pp. 357-358.

Kidd, I. J. 2022. From Vice Epistemology to Critical Character Epistemology. In: M. Alfano, C. Klein & J. de Ridder (eds.). *Social Virtue Epistemology*. New York: Routledge. pp. 84-102.

Kidd, I. J., H. Battaly, & Q. Cassam. 2020. *Vice Epistemology*. New York: Routledge.

Kidd, I. J., Chubb, J. & Forstenzer, J. 2021 Epistemic corruption and the research impact agenda. *Theory and Research in Education*, 19(2), pp. 148-167.

Kihlstrom, J. 2013. The person-situation interaction. In: D. Carlston (ed.). *The Oxford Handbook of Social Cognition*. Oxford: Oxford University Press. pp. 786–806.

King, N. L. 2014. Perseverance as an intellectual virtue. *Synthese,* 191(15), pp. 3501-3523, 3779-3801.

King, N. 2021. *The Excellent Mind: Intellectual Virtues for Everyday Life.* Oxford: Oxford University Press.

Kristjánsson, K. 2015. *Aristotelian Character Education*. London: Routledge.

Lackey, J. 2020. *The Epistemology of Groups*. Oxford: Oxford University Press.

Lackey, J. 2020. Group Belief: Lessons from Lies and Bullshit. *Aristotelian Society Supplementary Volume*, 94(1), pp. 185–208.

Lahroodi, R. 2007. Collective Epistemic Virtues. *Social Epistemology*, 21, pp. 281–97.

Lahroodi, R. 2018. Virtue Epistemology and Collective Epistemology. In: H. Battaly (ed.). *The Routledge Handbook of Virtue Epistemology.* New York: Routledge. Ch.33.

Lakatos, I. 1978. The Methodology of Scientific Research Programmes. *Philosophical Papers, Volume 1*. Cambridge: Cambridge University Press.

Leopold, J. & Bell, M.P. 2017. News media and the racialization of protest: an analysis of Black Lives Matter articles. *Equality, Diversity and Inclusion*, 36(8), pp. 720-735.

Levy, N. 2005. The Good, the Bad, and the Blameworthy. *Journal of Ethics and Social Philosophy*, 1(2), pp. 1–16.

Levy, N. 2017. The Bad News About Fake News, *Social Epistemology Review and Reply Collective*, 6(8), pp. 20-36.

Levy, N. 2019a. Nudge, Nudge, Wink, Wink: Nudging Is Giving Reasons. *Ergo*, 6, pp. 281–302.

Levy, N. 2019b. No-platforming and higher-order evidence, or anti-anti-no-platforming. *Journal of the American Philosophical Association*, 5(4), pp. 487-502.

Levy, N. & Alfano, M. 2020. Knowledge From Vice: Deeply Social Epistemology. *Mind, 129*(515), pp. 887-915.

Levy, N. 2022. Narrowing the Scope of Virtue Epistemology. In: M. Alfano, C. Klein & J. de Ridder (eds.). *Social Virtue Epistemology.* New York: Routledge. Ch.4.

Mabillard, V. & Pasquier, M. 2015. Transparency and Trust in Government: A Two-Way Relationship. *Jahrbuch der Schweizerischen Verwaltungswissenschaften*, 6(1), pp. 23–34.

MacKenzie, A. & Bhatt, I. 2020. Lies, Bullshit and Fake News. *Postdigital Science and Education*, 2, pp. 1-8.

Mackenzie, C. 2018. Moral responsibility and the social dynamics of power and oppression. In: K. Hutchison, C. Mackenzie & M. Oshana (eds.). *Social Dimensions of Moral Responsibility.* Oxford: Oxford University Press.

Macnamara, C. 2013. Reactive Attitudes as Communicative Entities. *Philosophy and Phenomenological Research*, 90(3), pp. 546-569.

Manthorpe, R. 2021. UK government censured for a lack of transparency and accountability. *Sky News* [online] Available at: https://news.sky.com/story/uk-government-censured-for-a-lack-of-transparency-and-accountability-12234248 [Accessed on August 31st 2022].

McHugh, C. 2012. Epistemic deontology and voluntariness. *Erkenntnis*, 77, pp. 65–94.

McIntyre, L. 2018. *Post-Truth*. Cambridge, MA: MIT Press.

McIntyre, L. 2019. *The Scientific Attitude*. Cambridge, MA: MIT Press.

McIntyre, L. 2020. Science Denial, Polarisation, and Arrogance. In: A. Tanesini & M.P. Lynch (eds.). *Polarisation, Arrogance, and Dogmatism*. London: Routledge. pp. 193-211.

Medina, J. 2013. *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and the Social Imagination*. Oxford: Oxford University Press.

Meehan, D. 2019. Is epistemic blame distinct from moral blame? *Logos and Episteme*, 10(2), pp. 183-194.

Meehan, D. 2023. A social account of the vices of self-assessment. *Philosophical Psychology*, 36(5), pp. 1033-1036.

Menges, L. 2017. The emotion account of blame. *Philosophical Studies*, 174, pp. 257–273.

Merritt, M. 2000. Virtue Ethics and Situationist Personality Psychology. *Ethical Theory and Moral Practice*, 3, pp. 365–83.

Meyer, M. 2019. Fake News, Conspiracy, and Intellectual Vice. *Social Epistemology Review and Reply Collective*, 8(10), pp. 9-19.

Meyer, M. & Alfano, M. 2022. Fake news, conspiracy theorizing, and intellectual vice. In: M. Alfano, C. Klein & J. de Ridder (eds.). *Social Virtue Epistemology*. New York: Routledge. pp. 236-259.

Miller, C. 2017. Honesty. In: W. Sinnot-Armstrong & C. B. Miller (eds). *Moral Psychology Volume 5: Virtues and Character.* Cambridge: MIT Press. pp. 237–73.

Miller, S. 2018. Corruption. [online]. In: E. N. Zalta (ed.). *The Stanford Encyclopedia of Philosophy (Winter 2018 Edition)* [online]. Available at <https://plato.stanford.edu/entries/corruption/> [Accessed April 2022].

Miyazono, K. 2023. Epistemic Libertarian Paternalism. *Erkenn*, 1(20), pp.1-20.

Montmarquet, J. A. 1993. *Epistemic Virtue and Doxastic Responsibility.* Maryland: Rowman and Littlefield.

Murtin F., Fleischer, L., Siegerink, V., Aassve, A., Algan, Y., Boarini, R., Gonzalez S., Lonti, Z., Grimalda, G., Schmidt, U., Hortala Vallve, R., Kim, S., Lee, D., Putterman, L. & Smith, C. 2018. Trust and its determinants: Evidence from the Trustlab experiment. *OECD Statistics Working Papers*, 2018(2), pp. 2-74.

Nesset, B., Robb, D. A., Lopes, J. & Hastie, H. 2021. Transparency in HRI: Trust and Decision Making in the Face of Robot Errors. In: Bethal, C., A. Paiva, E. Broadbent, D. Feil-Seifer & D. Safir (eds.). *HRI '21 Companion: Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*. pp. 313-317.

Niker, F. 2018. Policy-led Virtue-Cultivation: Can we nudge citizens towards developing virtues? In: T. Harrison & D. Walker (eds.). *The Theory and Practice of Virtue Education*. London: Routledge. pp. 153-167.

Nguyen, C. T. 2022. Transparency is Surveillance. *Philosophy and Phenomenological Research*, 105(2), pp. 331-361.

Nottelmann, N. 2007. *Blameworthy belief: A Study in epistemic deontologism.* New York: Springer.

Nozick, R. 1974. *Anarchy, State, and Utopia.* New York: Basic Books.

Office of Public Affairs: US Department of Justice. 2012. *BP Exploration and Production Inc. Agrees to Plead Guilty to Felony Manslaughter, Environmental Crimes and Obstruction of Congress Surrounding Deepwater Horizon Incident* [Press release]. 15 November 2012. Available at: <https://www.justice.gov/opa/pr/bp-exploration-and-production-inc-agrees-plead-guilty-felony-manslaughter-environmental#:~:text=(BP)%20has%20agreed%20to%20plead,environmental%20disaster%20in%20U.S.%20history%2C> [Accessed July 2022].

Okin, S. M. 1996. Feminism, Moral Development, and the Virtues. In: R. Crisp (ed.). *How Should One Live?: Essays on the Virtues*. Oxford: Oxford University Press. pp. 211–229.

O'Neill, O. 2002. *A Question of Trust: The BBC Reith Lectures 2002*. Cambridge: Cambridge University Press.

O'Neill, O. 2006. Transparency and the Ethics of Communication. In: C. Hood & D. Heald, (eds.). *Transparency: The Key to Better Governance?* Oxford: Oxford University Press. pp. 75-90.

Oshana, M. 2016. A Feminist Approach to Moral Responsibility. In: K. Timpe, M. Griffith & N. Levy (eds.). *The Routledge Companion to Free Will.* Abingdon: Routledge. Ch.55.

Papineau, D. 2012. *'Correlations and Causes' in Philosophical Devices*. Oxford: Oxford University Press.

Pearce, D. & Uridia, L. 2015. Trust, Belief and Honesty. *EPiC Series in Computer Science,* 36, pp. 215-228.

Pereboom, D. 2014. *Free Will, Agency, and Meaning in Life*. Oxford: Oxford University Press.

Piovarchy, A. 2021. What do We Want from a Theory of Epistemic Blame?. *Australasian Journal of Philosophy,* 99(4), pp. 791-805.

Priest, M. 2021. Epistemic Insensitivity: An Insidious and Consequential Vice. In: I. J. Kidd, H. Battaly, & Q. Cassam (eds.). *Vice Epistemology*. New York: Routledge. Ch.10.

Pritchard, D. 2013. Epistemic Virtue and the Epistemology of Education. *Journal of Philosophy of Education*, 47(2), pp. 236-247.

Pritchard, D. 2013. Epistemic paternalism and epistemic value. *Philosophical Inquiries*, 1(2), pp. 9–37.

Pritchard, D. 2015. Intellectual Virtue, Extended Cognition, and the Epistemology of Education. In: J. Baehr (ed). *Intellectual Virtues and Education: Essays in Applied Virtue Epistemology.* New York: Routledge. pp. 113–28.

Porter, S. L. 2016. A Therapeutic Approach to Intellectual Virtue Formation in the Classroom. In: J. Baehr (ed.), *Intellectual Virtues and Education: Essays in Applied Virtue Epistemology*. London: Routledge. pp. 221–239

Potter, N. N. 2002. *How Can I Be Trusted?: A Virtue Theory of Trustworthiness*. Maryland: Rowman & Littlefield Publishers.

Quinton, A. 1976. Social objects. *Proceedings of the Aristotelian Society,* 76(1), pp. 1-27.

Rawlins B. L. 2008. Measuring the relationship between organizational transparency and employee trust. *Public Relations Journal,* 2(2), pp. 1–21.

Rettler, L. 2017. In defence of doxastic blame. *Synthese*, 195(5), pp. 2205-2226.

Reutlinger, A. 2020. What is epistemically wrong with research affected by sponsorship bias? The evidential account. *European Journal for Philosophy of Science*, 10(15), pp. 1-26.

Rini, R. 2017. Fake News and Partisan Epistemology. *Kennedy Institute of Ethics Journal*, 27(2), pp. 43-64.

Riley, E. 2017. The Beneficent Nudge Program and Epistemic Injustice. Ethical Theory and Moral Practice, 20(3), pp. 597–616.

Ryan, S. 2016. Paternalism: an analysis. *Utilitas*, 28(2), pp. 123.

Ryan, S. 2018. Libertarian paternalism is hard paternalism. *Analysis*, 78(1), pp. 65–73.

Saghai, Y. 2013. Salvaging the concept of the nudge. *J Med Ethics*, 39(8), pp. 487-93.

Scanlon, T. M. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

Scanlon, T. M. 2008. *Moral Dimensions: Permissibility, Meaning, and Blame*. Cambridge, MA: Harvard University Press.

Shapin, S. 1995. Truth, honesty, and the authority of science. In: R. E. Bulger, E. M. Bobby & H. V. Fineberg (eds.). *Society's Choices: Social and Ethical Decision Making in Biomedicine.* Washington: National Academy Press. pp. 388-408.

Sher, G. 2006a. *In Praise of Blame*, New York: Oxford University Press.

Sher, G. 2006b. Out of Control. *Ethics*, 116(2), pp. 285–301.

Sher, G. 2009. *Who Knew? Responsibility Without Awareness*. New York: Oxford University Press.

Shoemaker, D. 2015. *Responsibility from the margins*. Oxford: Oxford University Press.

Sliwa, P. (forthcoming). Taking Responsibility. In: N. Chang & A. Srinivasan (eds.). *Conversations in Philosophy, Law, and Politics*. Oxford: Oxford University Press. Ch.4.

Slote, M. 1982. Selective necessity and the free will problem. *Journal of Philosophy*, 79 (January), pp. 5-24.

Smiley, C. & Fakunle, D. 2016. From "brute" to "thug:" the demonization and criminalization of unarmed Black male victims in America. *Journal Of Human Behavior In The Social Environment*, 26(3-4), pp. 350–366.

Smith, A. M. 2005. Responsibility for Attitudes: Activity and Passivity in Mental Life. *Ethics*, 115(2), pp. 236–271.

Smith, A. M. 2008. Control, Responsibility, and Moral Assessment. *Philosophical Studies*, 138(3), pp. 367–392.

Smith, A. M. 2015. Responsibility as Answerability. *Inquiry*, 58(2), pp. 99–126.

Smith, A. M. 2013. Moral Blame and Moral Protest. In: D. J. Coates & N. A. Tognazzini (eds.). *Blame: Its Nature and Norms*. Oxford: Oxford University Press. Ch.2.

Sosa, D. 2011. Some of the Structure of Experience and Belief. *Philosophical Issues,* 21, pp. 474-484.

Sosa, E. 2007. *A Virtue Epistemology.* Oxford: Oxford University Press.

Sosa, E. 2015. *Judgment and Agency.* Oxford: Oxford University Press.

Stewart I. & Cohen J. 1997. *Figments of reality: The evolution of the curious mind.* Cambridge: Cambridge University Press.

Strawson, P. 1962. Freedom and resentment. In: G. Watson (ed.). *Proceedings of the British Academy,* 48, pp. 187-211.

Sullivan, E. & Alfano, M. 2020. Vectors of Epistemic Insecurity. In: I. J. Kidd, H. Battaly, & Q. Cassam (eds.). *Vice Epistemology*. New York: Routledge. pp. 148-164.

Swami, V., Voracek, M., Tran, U. S., Stieger, S., & Furnham, A. 2018. Analytic thinking reduces belief in conspiracy theories. *Cognition* 133(3), pp. 572-85.

Sunstein, C. R., Hastie, R. & Schkade, D. 2007. What happened on deliberation day? *California Law Review*, 95(3), pp. 915–940.

Sunstein, C. R. 2014. Nudging: A Very Short Guide. *Journal of Consumer Policy*, 37(4): pp. 583–588.

Sunstein, C. R. 2015. Nudges Do Not Undermine Human Agency: A Note. *Journal of Consumer Policy*, 38(3), pp. 207–210.

Talbert, M. 2008. Blame and Responsiveness to Moral Reasons: Are Psychopaths Blameworthy? *Pacific Philosophical Quarterly,* 89, pp. 516-535.

Talbert, M. 2019. Moral Responsibility. In: E. N. Zalta (ed.). *The Stanford Encyclopaedia of Philosophy*. Available at <https://plato.stanford.edu/entries/moral-responsibility/> [Accessed January 2020].

Talbert, M. 2012. Moral Competence, Moral Blame, and Protest. *The Journal of Ethics,* 16(1), pp. 89-109.

Tanesini, A. 2016. Measuring and mismeasuring the self. *Aristotelian Society Supplementary Volume*, 1, pp. 71–92.

Tanesini, A. 2018. Epistemic Vice and Motivation. *Metaphilosophy*, 49(3), pp. 350-367.

Tanesini, A. 2019. Vices of the Mind, by Quassim Cassam [Book Review]. *Mind,* 129(515), pp. 959-964.

Tanesini, A. 2021. *The Mismeasure of the Self: A Study in Vice Epistemology*. Oxford: Oxford University Press.

Taylor, C. 1976. Responsibility for Self. In: A. O. Rorty (ed.). *The Identities of Persons*, Berkeley. CA: University of California Press. pp. 281–99.

Tessman, L. 2005. *Burdened Virtues: Virtue Ethics for Liberatory Struggles*. Oxford: Oxford University Press.

Thaler, R. H. & Sunstein, C. R. 2003. Libertarian Paternalism. *American Economic Review*, 93(2), pp. 175-179**.**

Thaler, R. H. & Sunstein, C. R. 2008. Nudge: Improving Decisions about Health, Wealth, and Happiness. New Haven: Yale University Press.

Tilly, C. 2002. *Stories, Identities, and Political Change.* Lanham: Rowman & Littlefield.

Tuana, N. 2006. The speculum of ignorance: The women's health movement and epistemologies of ignorance. *Hypatia*, 21(3), pp. 1-19.

Unesco. 2022. Promoting proactive transparency during COVID-19 in Mexico. Available at: <https://www.unesco.org/reports/access-to-information/2021/en/case-study-mexico>. [Accessed on: August 31st 2022].

Wallace, R. J. 2013. Dispassionate opprobrium. In: D. J. Coates, & N. A. Tognazzini (eds.). *Blame: Its nature and norms*. Oxford: Oxford University Press.

Watson, G. 1996. Two Faces of Responsibility. *Philosophical Topics*, 24(2), pp. 227-248.

Watson, G. 2004. *Agency and Answerability: Selected Essays.* Oxford: Oxford University Press.

Watson, L. 2021. *The Right to Know: Epistemic Rights and Why We Need Them (1st Edition).* Abingdon: Routledge.

Williamson, T. 2000. *Knowledge and Its Limits.* Oxford, Oxford University Press.

Wolf, S. 1990. *Freedom Within Reason*. New York: Oxford University Press.

Wood, M. J., Douglas, K. M., & Sutton, R. M. 2012. Dead and Alive: Beliefs in Contradictory Conspiracy Theories. *Social Psychological and Personality Science*, 3, pp. 767-73.

Zagzebski, L. 1996. *Virtues of the Mind*. New York: Cambridge University Press.

Zagzebski, L. 1998. The Virtues of God and the Foundations of Ethics. *Faith and Philosophy*, 15(4), pp. 538–553.

Zagzebski, L. 2004. *Divine Motivation Theory*. New York: Cambridge University Press.

Zagzebski, L. 2017. 'What Is Knowledge?'. In: J. Greco & E. Sosa (eds.). *The Blackwell Guide to Epistemology.* New Jersey: Wiley-Blackwell. Ch.3.

Zheng, R. 2016. Attributability, Accountability, and Implicit Bias. In: J. Saul & M. Brownstein (eds.). *Implicit Bias and Philosophy, Volume 2: Moral Responsibility, Structural Injustice, and Ethics*. New York: Oxford University Press. pp. 62-89.