WARWICK
THE UNIVERSITY OF WARWICK

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

http://wrap.warwick.ac.uk/165462

Copyright and reuse:
This thesis is made available online and is protected by original copyright.
Please scroll down to view the document itself.
Please refer to the repository record for this item for information to help you to cite it.
Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

warwick.ac.uk/lib-publications

An aretaic account of responsibility for beliefs

Giulia Luvisotto

A thesis submitted in fulfilment of the requirements
for the degree of *Doctor of Philosophy in Philosophy*

University of Warwick, Department of Philosophy
August 2021

Table of contents

Acknowledgements

I thank Eva Picardi, Roberto Brigati and Barry Stroud for teaching me to doubt, to be patient and to trust my nose. I thank especially Johannes Roessler, who has discussed with me countless versions of this work and taught me that a further, deeper question can always be found.
Without them, I might have not given up philosophy, but I would have certainly taken the habit of settling for answers at a much earlier stage.

I thank my family, who's sincerely proud of me for doing something they don't fully understand.

I thank my friends, the family I was lucky to find and wise to choose, for making the world smaller (and much warmer).

I thank Ahilleas for teaching me that disagreeing is not a bad thing, and for reminding me that philosophy is not all there is to life. To many years of heated arguments, and many more things.

And I dedicate my work to my grandpa, whose teachings and unconditional love I cannot put into words.

Declaration

I declare that this thesis is my own work, that it has not been submitted for a degree at another university, and that it does not contain material which has been used before or which I have already published.

Abstract

This thesis argues that the practices of *attributability* for beliefs constitutes the core of the phenomenon of 'responsibility for beliefs', against a strong tendency in the debate to focus exclusively on the practices of *accountability* for beliefs. The overarching aim of this thesis then is to offer an alternative account to the dominant theory of responsibility for beliefs, the *accountability view*, which is modelled on the practices of accountability for actions and is thus unsuitable to explain the practices of attributability for beliefs. In particular, being control- and norm-based, the accountability view neglects cases in which we are criticisable (and commendable) in responsibility-involving ways for our beliefs even when no breach of (or extraordinary compliance with) norms can be detected. I develop the *aretaic model* as an alternative virtue-centred account. On the aretaic model, someone is responsible for a belief insofar as it expresses their evaluative orientation, which is their sensitivity to both practical and theoretical reasons, oriented by their values. To explain the normativity of aretaic appraisals, I articulate the notion of evaluative orientation in aretaic terms: someone is open to commendation if their belief expresses a virtue, to criticism if it expresses a vice and to either commendation or criticism if it displays a value that is neither a virtue nor a vice. I develop several implications of my view: I resist the reduction of doxastic responsibility onto epistemic responsibility, I reject the distinction between 'moral' and 'epistemic' virtues and vices, and I deny the cogency of the control requirement. Finally, where the accountability view is unable to make sense of the practices of attributability, I suggest that the aretaic model can offer a satisfactory explanation of both the practices of attributability *and* accountability for beliefs.

I. Two accounts of responsibility for beliefs

1. Actions and beliefs

If we had to find an image for the practices involved in deeming each other responsible, it could take the form of a painting by Hieronymus Bosch. Perhaps his *The Garden of Earthly Delights*—large, diverse, full of details you don't notice at first and that fit together only surprisingly (and not always harmoniously). In fact, we talk of 'responsibility' to refer to a wide variety of phenomena. In this thesis, I will not be concerned with all of them;[1] I am interested in the use of 'responsibility' that is applicable only to persons.[2] The challenge is to understand when and why they are responsible.

When we say that persons can be responsible for something, we mean either that they *are* responsible or that they *are open to* certain 'responsibility-involving evaluations'.[3] Of course, the two are tightly related. Yet, according to which one we consider to be explanatorily prior, we'll have a different understanding of the target phenomenon and a different philosophical agenda. The choice mirrors the Euthypro dilemma: if we think that you deem me responsible for something because I am responsible for it, then we'll need to investigate the particular relationship that links that something and myself. If we think that I am responsible for something because I am open to be evaluated in certain ways by you, then we'll need to look into the relationship that holds between you and me, and into the nature of those evaluations. My arguments in what follows do not depend on a commitment to this question. In setting up the discussion, I follow a broadly Strawsonian tradition and consider the fact that someone is open to certain kinds of evaluations to be at least an essential part of what it is to be responsible.

There are several things for which we can be responsible. Of course, we are responsible for our intentional actions, and morally responsible for at least some of them. But in recent years there's been a surge of interest in other, traditionally neglected, cases: we are responsible also for our beliefs, fears, hopes, desires, and we are not always morally responsible for them. Yet, most of the accounts of responsibility for attitudes available in the literature seem to build (more or less implicitly) on intuitions that derive from the case of actions. We think about what it means to be

[1] For instance, I won't consider strict liability, nor any merely causal understanding of responsibility.
[2] I also set aside for now pathological conditions.
[3] I take these evaluations to involve both a cognitive and an emotional element, though I remain neutral as to their precise metaphysical status. However we understand their nature, what's crucial for me is that they can be supported by reasons. For the view that Strawsonian reactive attitudes do *not* include or imply normative judgments, or any propositional thought, see Deigh (2011).

responsible for our intentional actions, and under what conditions ascriptions of responsibility are fit, and then we verify whether those conditions can be met in the case of belief.

Considered together, these intuitions inform what I call *the accountability view* of responsibility. The accountability view does not refer to someone's proposal in particular but rather stands for the backbone of several views, which share some key assumptions. In very general terms, on the accountability view someone is responsible for the intentional actions that are under their voluntary control, and they are open to responsibility-involving moral criticism[4] when, in acting, they violate moral norms that we could expect and demand them not to violate, and with which they could have complied. Conversely, someone is open to responsibility-involving moral commendation for some of the actions over which they exercise control insofar as, in acting, they have complied with the relevant norms exemplarily, or when it was particularly difficult to do so.

To be applicable to the case of responsibility for beliefs, the accountability view is typically modified in two key respects. First, the relevant norms are taken to be epistemic. Second, the control involved is not the voluntary control we exercise directly over our intentional actions; rather, it is usually thought of as voluntary but indirect, or non-voluntary. Along these lines, someone is epistemically responsible for a belief insofar as they exercise control over it, and they are open to responsibility-involving epistemic criticism for a belief insofar as, in believing it, they violate some epistemic norms that we could expect and demand them not to violate, and with which they could have complied. Conversely, they are open to responsibility-involving epistemic commendation for one of their beliefs insofar as, in believing it, they have complied with the relevant epistemic norms exemplarily.

The overarching aim of this thesis is to provide an alternative way of thinking about responsibility for beliefs. Instead of delineating a general understanding of 'responsibility' to then verify how it scores when applied to the case of beliefs, I want to start by looking at cases in which we would consider believers to be responsible for their beliefs, and build an account from there. I call my proposal *the aretaic model.*

In a nutshell, on the aretaic model someone is responsible for a belief insofar as it expresses their evaluative orientation, which I understand as their distinctive sensitivity to both practical and theoretical reasons, oriented by their values. This entails a shift of focus from assessments of beliefs to assessments of persons, and from an emphasis to compliance with norms to an emphasis on the quality of someone's reasons. The aretaic model is not committed to a control requirement.

[4] I will use 'criticism' and 'commendation' instead of 'blame' and 'praise' because they are less philosophically charged. I take them to mean, in general, negative and positive responsibility-involving evaluations.

Moreover, it resists the idea that the only two relevant normative dimensions of assessments are the moral and epistemic ones: believers are evaluated in richer and more varied ways than simply for being 'morally' or 'epistemically' good or bad. To articulate the latter point, I develop the notion of 'evaluative orientation' by means of a virtue-centred framework.

Before proceeding, a terminological clarification is due. The distinction I draw between 'the accountability view' and 'the aretaic model' echoes the distinction between 'accountability' and 'attributability' drawn by Gary Watson in his seminal *Two faces of responsibility*. However, first, Watson is concerned primarily with the case of actions, whilst I focus on responsibility for beliefs. Second, more importantly, in that paper Watson differentiates between two *practices*, *i.e.* of holding accountable and appraising aretaically, whereas I am identifying two *philosophical theories*, or explanatory frameworks. This means that, in arguing against the (monopoly of the) 'accountability view', I try to show the limits of a specific and widespread way of understanding philosophically what it is to be responsible for one's beliefs. But in doing so, I don't deny that there is room for both the practices of accountability and attributability for beliefs. As I understand it, the two practices are distinguishable insofar as they are informed by different concerns: the former are characterised by a certain *interpersonal* element that is not necessary for the latter. When I hold you accountable, I do so because you have somehow wronged or done right by me, and thus I can place certain relevant expectations and demands on you. My project is not to deny that we can be accountable for our beliefs. Rather, it is to show that the practices of *attributability* are central to responsibility for beliefs, even though they've been largely overlooked in the literature, and to argue that we cannot make sense of them by means of a theoretical framework devised to explain the practices of accountability for actions.

This shall suffice by way of preliminary remarks. To ease into the discussion, I suggest we look at some examples of beliefs for which we are responsible.

2. *Some cases*

Here are some examples of 'criticisable beliefs' from the literature:

> We say that a typical well-informed contemporary American ought to believe that the Earth revolves around the Sun and should not believe that the Earth is flat. (Feldman 2000, 667)

> Imagine that someone is insistent that he sees a unicorn galloping toward him and based on the usual trustworthiness of his senses and believes there is a unicorn galloping toward him. If this person later discovers that a powerful hallucinogenic drug had been slipped in his drink and he still believes that a unicorn had galloped toward him, it is appropriate to criticize him. For, normally, if I point out reasons for thinking that your

> normally well-functioning mechanism has gone awry, you should revise your assessment of the resulting upshot. Further, you see yourself as appropriately chastised for being overly confident or hasty in the proper functioning of this mechanism. (McCormick 2011, 175)

> To see this, consider a familiar case of blameworthy belief from the literature, Bonjour's case of Maud. Maud has a reliable clairvoyant power and uses that power to come to believe that the President is in New York. However, she also has strong evidence that the President is not in New York (perhaps, the television is showing apparently live pictures of the President in Washington). Furthermore, Maud has no evidence for the existence of her clairvoyant power, and has plenty of evidence that supports that clairvoyance is impossible. Importantly, Maud's problem is not that she simply cannot appreciate the strength of the contrary evidence. Like the rest of us, she understands that the television's showing apparently live pictures of the President in Washington is excellent evidence that he is not in New York. Furthermore, she understands the scientific evidence which supports that clairvoyance is impossible. Nonetheless, she dismisses the TV broadcast and the evidence against clairvoyance and dogmatically continues to believe that the President is in New York. In so doing, it seems that Maud is blameworthy. Consequently, we may well blame her for her belief. (Brown 2020, 2–3)

> We blame people who believe that whales are fish, that some UFOs are alien spaceships paying regular visits to earth, that a neighbor littered the sidewalk when she did not. Although the blame is attached only when these beliefs are visible or acted on, the blame is for the belief itself. We think it blameworthy for people to hold such beliefs, given the available evidence. (Adler 2002, 64)

This is just a small sample, but it suffices to give an idea of the kind of beliefs that are typically considered to be the paradigmatic case: beliefs that are false, unjustified, incorrect, or inconclusively supported. Even more, they are taken to open their believers to the relevant kind of responsibility-involving criticism *insofar as* they're false or unjustified. In being false and unjustified, they count as instances of violations of norms with which the believers are expected to comply. We "would disapprove of someone" who holds those beliefs because they should not do so and could have avoided it.

Yet, this set of beliefs is quite peculiar, and I think it hardly captures the core of the phenomenon we are after. We rarely find ourselves assessing beliefs held by Flat Earthers and clairvoyants, or about unicorns and whales. It might be that looking at examples like these results from a fear that, if we consider more familiar circumstances, folk psychology will contaminate the investigation. Perhaps, it is simply due to philosophy's preference for bizarre thought experiments

over more mundane situations. Or maybe it shows that we are still quite unclear on *precisely what* phenomenon we are trying to illuminate. In any case, I think that taking these as its paradigmatic instances risks misleading our investigation. All of us deem others (and are themselves deemed) responsible for a great range of beliefs, not all of which are obviously false or unjustified. And if philosophy wants to illuminate *those* practices, we need to consider more familiar circumstances.

Consider these:

PLANT	Returned home from work, I see that one of my beloved plants has a bent stalk. I believe that the damage was caused by my roommate, whilst in fact the wind knocked it over. My roommate picked it up and closed the window.
UNWITTING VOTER	I believe that Mr. X would be a great president because my favourite tabloid says he would be.
SHERLOCK	I am Sherlock Holmes and I truly believe that the murder weapon is a boomerang, when no one else is able to solve the mystery.
SON	I believe that my aggressive son, who has been tried and found guilty, is innocent.
CAPITAL	I believe that Auckland is the capital of New Zealand.
ENVY	I believe that my sister does not deserve to be more popular than I am.
WITTING VOTER	I believe that Mr. X would be a great president because I read his declarations, follow his political moves and am convinced that he fosters the right policies.
NORPOIS	Mr. Norpois believes that it would not be seemly to ask the Comte de Paris a question.

The accountability view and the aretaic model offer two alternative ways of explaining why I am open to responsibility-involving evaluations in these situations. Sometimes their diagnoses are compatible, sometimes they conflict.

The accountability view accommodates quite nicely cases like PLANT, UNWITTING VOTER, SHERLOCK. In PLANT, I am criticisable because I jumped to conclusions rather than carrying out a proper inquiry, even though I could have done better. I haven't been an accurate believer. I formed the belief that my roommate was guilty on inconclusive evidence, hastily, before even asking them what had happened. To count as an accurate believer, I should have looked for other relevant clues (maybe some soil on the floor), wondered why and how my housemate would have damaged my plant, or asked them for an explanation. Similarly, I would be criticisable in UNWITTING VOTER: I

relied on a tabloid article in a loose, inattentive and passive fashion, where I should have instead tried to put myself in the best possible position to judge *ex professo* for whom to vote. I should have realised that such a source does not provide sufficient epistemic reasons. My belief is inconclusively supported. Conversely, in SHERLOCK I am commendable for having been exceptionally good in forming my belief. My belief is an exceptionally good belief to form and hold. It is a belief that most of us would not be able to acquire so quickly.[5] That I hold this belief proves that I am a competent and capable rational being.

Cases like SON and CAPITAL are less straightforward to assess. On the one hand, whatever I conform to in SHERLOCK—be them epistemic norms, epistemic reasons, or standards of rationality–, I violate in SON. In SHERLOCK, I abide by them even though it's not easy to do so, and this brings me closer to a commendable ideal of rationality. In SON, I am open to criticism insofar as my belief is irrational, unsupported, unjustified, false. My belief is false and unjustified also in CAPITAL. Here, I fail either in forming it or in revising it. And it is very likely that I have been given the chance to revise it at least once in my life—there must have been one occasion in which I was looking at the world map and casually inspecting the austral hemisphere, or some such. Yet, on the other hand, it's not obvious that *I am* criticisable in SON and CAPITAL. *Even though* I believe something false or unjustified, I may not be open to the relevant sort of responsibility-involving criticism for my belief. For instance, in SON my false belief might show an unconditional love for my son. I am someone who strongly trusts their family members, who's hopeful and confident. This doesn't mean that I'm justified in believing as I do, but rather that what my belief shows of me affects in considerable ways the kinds of evaluations to which I am open. In CAPITAL, instead, my false belief is simply quite irrelevant. Perhaps I would be considered ignorant in geography; someone might think that I made a mistake and correct me. But it's unclear in what sense I would be open to *blame* or responsibility-involving *criticism* for my belief. If so, then violating 'epistemic norms' may not suffice to be open to responsibility-involving criticism.

In some cases, I violate epistemic standards and I am criticisable. However, I am not criticisable *because* of those violations. For instance, consider ENVY. Here, I believe that my sister does not deserve to be more popular than me. As a matter of fact, she's always kind with everyone, she's very active in the community and volunteers for all sorts of activities, whilst I am fairly grumpy and timid. So my belief is unjustified; she does deserve to be as popular as she is. I have no good epistemic reasons to believe as I do. However, when I am criticised for my belief, I am not criticised

[5] On the accountability view, often it's said that commendation is appropriate only if it were somehow difficult for me to form a true belief. This is because most of our beliefs are true, and we cannot allow for too liberal an account of responsibility-involving commendation. However, notice that, if generalised, this amendment risks relegating commendation for beliefs into quite a narrow corner. And yet, we do praise, admire, value, love, and hold in high esteem others for a great variety of beliefs, not all of which are difficult to form and hold *from an epistemic point of view*.

because it is unjustified. Rather, I am criticised for being someone who's envious and bitter.[6] The fact that I violated some epistemic norms is not what explains the kinds of evaluations to which I am open.

In some cases, it's unclear *whether* I violate epistemic standards. Take WITTING VOTER. Suppose that Mr. X is someone who openly discriminates disabled people, prevents women from obtaining a driver licence, and makes healthcare prohibitively expensive. I did what any good believer should do: I gathered information, read articles and declarations, and pondered on the issue. As a result, I believe he is a great president. An advocate of the accountability view might be led to claim, on that basis, that I am *not* criticisable precisely because I did believe as scrupulously as I possibly could. Of course, this would be an embarrassing verdict, for it contrasts with the common-sense intuition that I am criticisable for holding that belief. At this point, the advocate of the accountability view would deny that I am a good believer: the information I gathered does not provide reasons for that belief at all, since discriminating part of the population simply cannot count in favour of being a good president.[7] Thus, I am not responsive to reasons in the appropriate way, my capacity for belief-formation and retention is faulty. But even so amended, the accountability view seems to miss the crucial point. For if I am criticisable in WITTING VOTER, it is not because my evidence does not in fact support the belief I take it to support. Rather, I am criticisable for the very fact that I take *certain* facts and considerations to be good reasons to believe that Mr. X is a great president; I am criticisable for the fact that I believe that that person is of worth, given the policies he fosters. I am unjust, aggressive, uncompassionate, unpleasant, maybe dumb, maybe dangerous.

Finally, there are some cases in which believers do not violate (nor comply extraordinarily with) any epistemic norm, and they are open to deep forms of criticism (and commendation) nonetheless. This is a passage from Proust's *Madame Swann at Home*:

> Meanwhile, my father, so as to forestall any criticism that we might feel tempted to make of our guest, said to my mother: "I admit that old Norpois was rather 'trite', as

[6] Though it involves also an emotional element, believing certain things seems to be a *necessary* ingredient of envy: it's difficult to see how I could be considered envious if I didn't believe that my sister does not deserve to be more popular than I am, or something along those lines. Furthermore, I can be envious even though I never do anything but *believing* certain things (and feeling certain ways); I need not act upon my beliefs. And being envious is definitely something for which we assess people in deep, responsibility-involving ways. If this is right, then we *need* a way of explaining why I am deemed responsible for being envious, in believing as I do, and it's unclear how the accountability view can make sense of it.

[7] To be more precise, whether the believer in WITTING VOTER counts as violating or not violating epistemic norms depends on one's take on several issues. Those who maintain that responsibility for beliefs requires the exercise of voluntary control over the activities that constitute the processes of belief-formation and retention will hold that the believer in WITTING VOTER cannot be criticised, for they have done all they could to put themselves in the position of assessing the truth of *p*. Those who think someone needs to be appropriately responsive to reasons will maintain that this believer is responsible, provided that they have the capacity to recognise the strength of normative reasons. I will discuss these issues at length in chapter IV.

> you call it, this evening, when he said that it would not have been 'seemly' to ask the Comte de Paris a question, I was afraid you might burst out laughing."
> "Not at all!" answered my mother. "I was delighted to see a man of his standing and age keep that sort of simplicity, which is really a sign of straightforwardness and good breeding."
> "I should think so, indeed! That does not prevent him from being shrewd and discerning; I know him well […]." (Proust 1918, 61)

Even though the whole paragraph is an assessment of Mr. Norpois on the basis of something he says, he's not evaluated for his speech act. He is not appraised because he *said* something, as when a statement is considered polite, imprudent or aggressive. Instead, the narrator's parents are evaluating Mr. Norpois himself, *e.g.* he "was rather 'trite'", in the light of the belief that *underlies* his utterance. What Mr. Norpois is said to demonstrate, *i.e.* simplicity, straightforwardness and good breeding, are not evaluations of his utterance: he is banal or candid not in virtue of voicing his beliefs, but rather because of the very fact that he believes that asking something to the Comte would be inappropriate.

It's quite difficult for the accountability view to make sense of cases like this. For Mr. Norpois does not seem to infringe any norm in believing as he does. For one thing, granted for now that it is possible to violate moral norms by holding a belief, that belief does not violate any moral norm. It is not morally bad to believe what he believes. Nor has Mr. Norpois failed to adhere to his epistemic reasons. Given certain facts about high society, it is perfectly reasonable for Mr. Norpois to believe that the Comte de Paris might be bothered by certain questions. He hasn't done anything he shouldn't have, nor does he believe something he should not believe. And yet, he is assessed in certain negative or positive terms for the very fact that he holds that belief.

Furthermore, it seems irrelevant whether Mr. Norpois can control his belief. There's no reference to his capacities to assess reasons or believe otherwise. In effect, if Mr. Norpois had wanted or deliberately tried to form that belief, and succeeded, some of those evaluations would even be misplaced: for instance, if he self-manipulated into holding that belief, that belief could not be indicative of his simplicity and straightforwardness at all. At most, it would be expressive of his wish to come across as simple and straightforward. Also, the assessments of Mr. Norpois do not involve any sort of demand; Mr. Norpois does not wrong anyone by believing as he does. Those who are judging him to be trite or simple are not asking that he modifies his beliefs, nor do they expect that he does so. Even less, they expect that he does so *because* of their assessments. They evaluate the kind of person Mr. Norpois is as it is manifested by the kind of things he takes to be true.

If it's true that we can be evaluated in deep ways for a broad and varied set of beliefs, the accountability view seems unable to capture the complexity of the target phenomenon. Sometimes, we violate (or comply extraordinarily with) epistemic norms and yet are not obviously open to the relevant criticism (or commendation).[8] Other times, we do violate epistemic norms and are criticisable for our beliefs, and yet those infractions are not what can ground and explain the criticism.[9] And yet other times we are open to responsibility-involving evaluations even though we have not violated or complied extraordinarily with epistemic norms, in any clear sense. If this is right, then violation of or compliance with epistemic norms is neither sufficient nor necessary for deeming someone responsible for their beliefs.

I think that one promising way to make sense of all these cases is to introduce reference to what my attitudes *express* of me, over and above my capacities to comply with norms. Precisely *what* we take someone's beliefs to express is a difficult question, and according to how we address it we'll have different explanations of responsibility for beliefs. As a first approximation, we can say that I am criticisable for my belief insofar as it expresses certain bad things about me, and commendable insofar as it expresses certain good things about me.[10] The aretaic model is built around this intuition.

By emphasising the importance of what someone's belief expresses of them, the aretaic model can make sense of the fact that we wouldn't necessarily criticise me in SON. The fact that I fail to believe that my son carried out such a violent act displays that I've raised the threshold of what counts as sufficient evidence to a considerable extent. Perhaps, it's unreasonably high. And yet, the fact that I'm being unreasonable is not all there is to me as a believer. My belief displays that I am someone who loves their son, who's charitable and determined. Similarly, the aretaic model can make sense of the mild reactions triggered by cases like CAPITAL: I am not criticised in any deep way despite violating epistemic norms by which I could have easily abided because my belief, though false and unsupported, does not express much of me after all. It doesn't trigger any particular form of commendation or criticism because it does not make manifest any of my values. It doesn't say much about me.

On the aretaic model, we can also make sense of cases like ENVY. The problem here is not that I form an unjustified belief but that I am envious and bitter towards my sister. Rather than

[8] The puzzle generated by the fact that not every violation of epistemic norms opens the believer to the relevant kind of criticism is receiving increasing attention in the literature. The challenge is to explain why this is the case. Several proposals have addressed this challenge by qualifying the set of epistemic norms. My suggestion is that, perhaps, the source of the problem is one step further down the line, *i.e.* that the mistake lies in linking responsibility-involving criticism to *violation of norms* to begin with.

[9] In the positive case, we comply extraordinarily with epistemic norms and are commendable for our beliefs, and yet that compliance is not what grounds and explains the commendation.

[10] For sympathetic views, see for instance Adams (1985), Hobart (1934), Sher (2001).

being happy for her happiness, I hold a grudge. I focus on others' lives more than on my own, and I attempt to find an external explanation for my unhappiness rather than being honest with myself. Consider WITTING VOTER. The advocate of the accountability view will either say that I've done what was expected of me as a believer, *e.g.* gathering evidence, considering the issue carefully, reflecting upon it, or that I have violated epistemic norms, given that the reasons I take to be conclusive are not good reasons. But either way, the point is I am not criticised *as a bad believer* as much as *as a bad person*. I am criticisable for *what* I believe because it expresses that I am someone who fosters inequality, doesn't respect other people's rights, underestimates the risks that may follow the implementation of certain policies. Finally, the aretaic model can also illuminate cases like NORPOIS, which all versions of the accountability view have a hard time explaining. In fact, if we agree that Mr. Norpois hasn't violated any norm in believing and that the accountability view offers a norm-centred explanation, then it's unclear how it could ever make sense of why Mr. Norpois is criticisable or commendable for his belief. When we deem him simple-minded, or old-fashioned, or gentle, we do not imply that he must have violated or complied with some norm, *e.g.* not to be old-fashioned. Rather, we imply that his belief displays that he's a certain kind of person, someone who's attentive to others' reactions, who cares about others' judgements, and so forth.

At this point, one might suggest that we employ the aretaic model for cases like NORPOIS and maintain the accountability view to explain the cases that it could easily accommodate, *e.g.* PLANT, UNWITTING VOTER, SHERLOCK. However, it seems to me that the aretaic model can provide a *better* explanation of cases like these. In fact, in PLANT, my housemate would be less concerned with the fact that I jumped to conclusions and more with the fact that I settled for the very explanation that puts them in a bad light, amongst all the possible ones. They would criticise me for failing to give them the benefit of the doubt. The problem (and the very source of criticism) would not be that I performed poorly as a believer, insofar as I formed a belief upon inconclusive evidence, as much as that I am not a good friend. My roommate might deem me unjust, negative, distrustful, unfair. Similarly, you might criticise me in UNWITTING VOTER on the basis that I show disinterest towards political decisions that, you think, are extremely important. Also here, the problem is not that I've formed beliefs inaccurately, but that my inaccurate belief expresses something about me that you might criticise. You might deem me thoughtless, ill-considered, lazy, uninvolved.

To clarify, the advocate of the aretaic model would agree that in cases like PLANT and UNWITTING VOTER I might also have violated certain norms. However, they would stress that I am not criticisable because I violated norms but rather because of what is expressed by the fact that I violated them. That I violate norms, if I do, matters because of what it expresses of me,

rather than in and by itself. So the accountability view and the aretaic model agree that I am criticisable in both PLANT and UNWITTING VOTER but they locate the source of my openness to criticism in different *loci*.

To sum up, I think that we take each other to be honest, kind, funny, bigoted, dull, simple-minded, arrogant on the basis of what we believe, and it's impossible to make sense of these evaluations from within the accountability view. The way in which we are assessed for our beliefs cannot be reduced to how well we form, hold or revise them, nor does it have to do simply with their truth and justification. Before turning to clarifying my proposal, let me flesh out the accountability view in more detail.

3. The accountability view

The gist of the accountability view is well summarised in this passage by Rettler:

> …we sometimes blame people in this same sense for violating the norms that govern doxastic attitudes, such as belief, disbelief, and suspension of judgment. The norms that govern doxastic attitudes are epistemic norms, so we might call this blame epistemic blame. While I remain neutral on what makes a failure distinctively epistemic, examples include having doxastic attitudes that are unjustified, unsupported, unreasonable, inconsistent, incoherent, and so forth. (Rettler 2018, 2210)

I think that we can identify four main elements that are integral to this approach, and we can trace each back to a discussion of responsibility for actions. First, the accountability view is committed to a control requirement, *i.e.* to the idea that someone can be responsible only for what they can somehow control. Conversely, if it's impossible for them to control something, then they cannot be held responsible for it.

I discuss the control requirement in chapter IV. Briefly, I suggest that it originates from a reflection on the case of action, a certain understanding of voluntariness and concerns about fairness. In fact, when we think about the conditions that excuse agents for their behaviour, we typically agree that they are not responsible if they couldn't control what they were doing, *i.e.* if they didn't know they were doing it, or were forced into it, or couldn't do otherwise. Proving that they were not in control of an action or its effects exempts them from the relevant sort of criticism and commendation. And if we hold them responsible anyways, we do so unfairly.

When applied to the case of beliefs, however, the control requirement generates well-known difficulties. The question arises: how can we be responsible for our beliefs, given that we cannot control them? Or rather, what kind of control do we exercise over our beliefs, given that

we seem to be responsible for them and yet we cannot control them voluntarily? Most proposals in the debate can be seen as ways of addressing the latter question, which I call 'the control challenge'. The reactions vary. Some philosophers agree that responsibility requires voluntary control and either accept that we cannot be responsible for our beliefs, or argue that we can voluntarily control them.[11] Others argue for a weaker version of the control requirement: for some, we exercise indirect control over our beliefs, and are thereby responsible for them, in virtue of being directly responsible for the activities that bring them about;[12] for others, we are responsible for our beliefs in virtue of controlling them non-voluntarily.[13]

Notice that, in amending the control requirement to make it suitable to the case of beliefs, the latter two proposals understand 'the ways in which beliefs are formed and retained' in two different ways. Those who maintain that we are responsible for the beliefs over which we exercise indirect control understand 'the ways in which we form and retain beliefs' as the activities that constitute the process of putting oneself in the best possible position from which to assess whether *p*. Amongst these activities there will be gathering evidence, looking for reliable sources of information, asking reliable and knowledgeable people whether *p*. If we accept this weaker reading, then someone will be assessed for a belief in a responsibility-involving way just if they can exercise control over what evidence they can gather and consider. One cannot be open to responsibility-involving criticism for what they believe if they've done everything they could do to get to know the truth.

By contrast, those who allow for a version of non-voluntary control understand 'the ways in which beliefs are formed and retained' in a stronger way: these are not (or not only) the activities of gathering evidence but also the very way in which someone is responsive to reasons, *i.e.* what they take to be conclusive reasons, how they see considerations support other considerations and to what extent. In this light, it's often said that someone is responsible for a belief if and only if they can control the proper functioning of their reasons-responsiveness. In this context, the believer's reasons-responsiveness is interpreted as a "mechanism"[14] or a "capacity" of being moved by the strength of reasons. Thus, I will be open to criticism if, in forming and revising my belief, I fail to exercise (non-voluntary) control over my belief despite being able to do so, *i.e.* if I fail to use my mechanism or capacity properly, or to the best of its possibilities.[15] When we look at how

[11] For instance, Ginet (2001), Steup (2000; 2008; 2011; 2012), Weatherson (2008).
[12] For instance, Meylan (2015; 2017), Nottelmann (2007), Peels (2017).
[13] For instance, Boyle (2009; 2011), Hieronymi (2006; 2008; 2009; 2014), McHugh (2011b; 2012; 2013; 2017).
[14] See for instance the work of McHugh and McCormick, who draw on Fischer and Ravizza (1998).
[15] Within this framework, some believers are shielded from criticism because, even if they used their capacities for reasons-responsiveness at their best, that would not guarantee a truth-conducive belief-formation. These will be either those who have a faulty mechanism or capacity, *e.g.* psychopaths who cannot understand the force of moral reasons or paranoid individuals who cannot understand the force of epistemic reasons, or if they have a properly functioning

someone forms or holds a belief, we do not simply assess how they gathered evidence but also how they made that very judgement itself.

The second ingredient of the accountability view is the idea that adherence to or infraction of the relevant norms is what determines whether someone will be assessed *positively* or *negatively* for some of the things over which they exercise control. Reference to norms is natural when we reflect about responsibility for actions. For one thing, we are all subjected to the same legal code, which is constituted by a series of norms. And even though the legal and the moral codes are of course distinct, the former echoes some of the most central moral intuitions and, in turn, aids a norm-based understanding of them. That someone violated a norm *explains* why we can criticise them in the relevant way. They are open to the relevant sort of criticism *because* they have violated a norm.

Depending on whether you focus on the subject's activities of gathering evidence or take into account also the functioning of their reasons-responsiveness itself, you'll obtain quite different ideas of what an infraction of a norm looks like. On the former view, committing an epistemic breach may consist in failing to put oneself in the best possible position from which to evaluate whether *p* is the case. On the latter, violating epistemic norms will have to do with failures to recognise and weigh properly your reasons. You might take insufficient evidence to be conclusive, make mistakes of evaluation, jump to conclusions, form a belief that is in contradiction with another set of beliefs you hold, and so forth. However we understand the conditions of infractions, they make criticism legitimate.

Importantly, the *nature* of the criticism at issue is determined by the nature of the norm infringed. If I violate moral norms, I am open to moral evaluations (and thereby I am morally responsible); if I violate epistemic norms, I am open to epistemic evaluations (and thereby I am epistemically responsible); and so forth. And typically on the accountability view we are said to be morally responsible for our actions and epistemically responsible for our beliefs.

Third is the idea that we can *expect* and *demand* others to abide by norms, so that, if they don't, we can criticise them in the relevant way. These expectations and demands are different from mere predictions and petitions in that they involve the idea that someone *owes* it to someone (else) to comply with the relevant norms. I expect and demand that a friend does not lie to me in the sense that they owe it to me to be sincere, and I can criticise them in the relevant way if they are not. On

mechanism or capacity that is temporarily out of order in spite of themselves, *e.g.* victims of propaganda, manipulation, hypnosis.

the accountability view, the 'relevant' responsibility-involving kind of criticism is characterised precisely by the fact that it involves a disappointed expectation.

To elaborate on the conceptual connection between the accountability view and the kind of expectations and demands it associates with responsibility, recall that the accountability view originates to explain *the practices of accountability* for actions. We hold someone accountable when they have somehow wronged us (or done right by us) and we expect and demand that they make up for it, by way of apology or compensation. Within the practices of accountability, I do owe it to other people to behave in certain ways and not to behave in certain other ones. However, it is less clear that these same demands and expectations are in place within the practices of attributability—let alone within the practices of attributability for beliefs. So even though the accountability view is right in maintaining a reference to expectations and demands when it explains the practices of accountability for actions, it risks including them also where they do not belong. Once again, the problem is that the accountability view originates with one specific aim and then purports to explain responsibility *tout court*.

Finally, the fourth ingredient of the accountability view is that the evaluations it issues are quite local: it focuses on a certain action or belief and only derivatively on the person who acts and believes. It is the action or the belief that constitutes a violation of a norm (or an instance of compliance with it) and that legitimises a negative or positive evaluation. In holding you accountable for an action or a belief, I judge[16] your action or belief to be somehow faulty (or remarkable) and I criticise (or commend) you on the basis that you are the one in control, or that should have been in control, of it.

The fact that the accountability view originates from a reflection on responsibility for actions has three major problematic implications for the account of responsibility for beliefs that it delivers.

First, it risks imposing unnecessary conditions on responsibility for beliefs. By assuming a conceptual connection between control, norms and responsibility, the accountability view generates questions that might be ill-posed when applied to the case of belief. Amongst these, what kind of control we exercise over our beliefs; what kinds of norms we violate when we are criticized for them; and why we are not always criticisable for committing an epistemic breach. I hope to show that we can illuminate responsibility for beliefs without having to address questions of this sort.

[16] I remain neutral on the difficult question whether 'holding accountable' is a matter of believing, feeling, or reacting in certain ways, and I use 'judge' here as a placeholder.

Second, the accountability view risks distorting our understanding of the very phenomenon we are trying to explain. In order to maintain it as a suitable explanation, the contemporary debate tends to focus on those beliefs that *can* be explained by it; but these are only peripheral instances, if they are genuine instances of responsibility for beliefs at all. The accountability view diverts our attention away from the very practices we wanted to understand.

And third, once presented with more familiar (and I think central) examples of beliefs for which we are responsible, it seems unable to provide a satisfactory explanation. Not only are we criticisable and commendable in the relevant ways also when no infraction of or compliance with epistemic norms occurs, as section 2 suggests. Also, the accountability view seems unable to explain how we can be responsible for *holding* beliefs, irrespectively of how we've formed them. Responsibility for beliefs tends to be reduced onto responsibility for the activities that constitute the processes of belief-formation and retention: the assessments to which I am open in virtue of what I believe depend on how good I was in gathering and weighing evidence, considering reasons, drawing conclusions and so forth. That someone is responsible for believing that *p* means that they are responsible for having formed, retained or revised it appropriately. Along these lines, this framework is unable to clarify why we assess others insofar as they are convinced by the truth of a certain proposition, when it's a reasonable thing to believe. There's no room to assess the *quality* of someone's reasons. But we do criticise and commend people for taking certain particular reasons as conclusive, or for forming certain beliefs rather than others.

Relatedly, besides reducing responsibility for beliefs to responsibility for belief-formation and -revision, the accountability view reduces doxastic responsibility to *epistemic* responsibility. When we are commended or criticised for our beliefs, we are commended and criticised from an epistemic point of view. The recent debate on pragmatic encroachment attempts to broaden this perspective by insisting that we can also be *morally* responsible for our beliefs. However, I want to say, the evaluations to which we are exposed on the basis of what we believe cannot be exhausted by reference to moral or epistemic, or even aesthetic and prudential, norms. Very often, we are criticised and commended in broad ethical terms, *e.g.* for being simple-minded, parochial, selfish, dishonest, kind, polite, shy, funny. We are criticised in rich ways across several normative dimensions of assessment that cannot be traced back to the moral and epistemic ones only. Unless we allow for this normative plurality, we will be able to capture only part of our practices.

4. The aretaic model: an agenda

The aretaic model attempts to provide a more flexible theoretical framework by shifting the focus from norms to the importance of what beliefs express of one. The challenge is to pin down *what*

precisely they express. For in a sense most of what can be predicated of me can 'express' something of me. Having wide shoulders may show that I used to like swimming as a teenager, but that does not mean that I am responsible for having wide shoulders as Mr. Norpois is responsible for believing that asking questions to the Comte of Paris would be 'unseemly'. Not *everything* that can be said to 'express something of me' is something for which I am responsible. The question is to where to draw the boundary.

The aretaic model's rejection of the control requirement generates scepticism about its very possibility of addressing this challenge. In turn, this has led to the idea that an attributability-centred perspective cannot yield an account of full-blown responsibility. For one thing is to say that my beliefs express something of who I am, but it is quite another to deem me *responsible* for it—especially since it lies outside the boundaries of what I can control—and to evaluate me positively or negatively on that basis. If the aretaic model aims to put forth an account of responsibility, something more has to be said about the normative force of these appraisals.

To see the suspicion manifested towards the notion of 'attributability', consider the following passage from McHugh:

> It seems that we are often responsible for what we believe. To say this is not merely to say that our beliefs are attributable to us—the same is true of our handsomeness or our intelligence—but that we are accountable for them. The buck for our beliefs lies with us. (McHugh 2013, 132)

McHugh explicitly adopts the accountability view, since he treats 'being responsible for what we believe' interchangeably with 'being accountable' for them. So it's not surprising that he explains why we cannot be responsible for our handsomeness by saying that: "while people can be assessed for their handsomeness, the standards of handsomeness are not demands to which people can legitimately be held" (McHugh 2013, 133). The conditions of accountability cannot be met: handsomeness is not under our control, and there are no corresponding norms to which we can hold each other. But what's interesting here is the claim that handsomeness is *attributable* to us: remarks like this contribute to the impoverishment of the term 'attributable', which becomes synonymous with 'that can be predicated of a subject'. So demoted, 'attributability as predicability' is clearly unable to provide an account of 'responsibility'.

On a similar vein, Neil Levy writes:

> But it neither is, nor ought to be, common ground that we are responsible for everything that can be attributed to us. There is, after all, a natural alternative. We can hold that assessment of attitudes is simply that: the attribution of qualities to the agent that are good or bad, admirable or repugnant. After all, *prima facie* there ought to be conceptual

> space for such assessment. We ought to be able to say that something is *bad* without saying that it is *blameworthy*. (Levy 2005, 5)

According to Levy, the problem with an attributability-centred perspective is that it cannot distinguish between something being 'bad' and 'blameworthy', given that it doesn't take on a control requirement.[17] And yet, he insists, not all negative assessment counts as blame. We should be able to draw such a distinction.

In general, both McHugh and Levy insist that an attributability-inspired account is unable to distinguish *responsibility-involving* criticism and commendation from non-responsibility-involving evaluations. In fact, I agree that we want to draw a line between ascriptions of responsibility and 'mere grading evaluations', where the latter are assessments of the quality of something against some standards. That is, we need to be able to separate genuine ascriptions of responsibility from evaluative judgements like 'the weather is bad', 'the chair is uncomfortable', 'the car is slow'. When we say things like these, we are not taking the weather, the chair or the car to be responsible for anything. We are simply registering that they're not optimal as weather, chairs and cars go. However, the aretaic model *does* distinguish the kind of 'badness' involved here from the 'badness' at play in genuine ascriptions of responsibility. The former refers to the fact that something is defective, damaged, broken, substandard, deficient. By contrast, when we deem someone responsible, we don't simply take them to be defective in some respect but we think badly *of them*. Being bad as a car is not comparable to being bad as a person. This is so not because only the latter exercises some sort of control over their actions and beliefs, but rather because only the latter expresses their own values in acting and believing as they do.

I will develop this line in chapter II. Briefly, my suggestion is that, amongst the things that can be taken to express something of us, we can be deemed responsible for the ones that express our *evaluative orientation*. Someone's evaluative orientation is their distinctive sensitivity to practical and theoretical reasons, oriented by their values. When we deem someone responsible for their beliefs, we appraise them aretaically for the kind of person they are; for the values that orient their sensitivity to reasons.[18] This explains why we cannot be deemed responsible for our handsomeness: it does not, in and by itself, express any of the values that orient our sensitivity to reasons. Reference

[17] He quotes for instance Adams: "To me it seems strange to say that I do not blame someone though I think poorly of him, believing that his motives are thoroughly selfish. Intuitively, I should have said that thinking poorly of a person in this way *is* a form of unspoken blame" (Adams 1985, 21); and Watson: "In one way, to blame (morally) is to attribute something to a (moral) fault in the agent; therefore, to call conduct shoddy *is* to blame the agent" (Watson 1996, 266).
[18] On somewhat similar lines, Williams writes "if one acknowledges responsibility for anything, one must acknowledge responsibility for decisions and actions which are expressions of character" and "to be an expression of character is perhaps the most substantial way in which an action can be one's own" (Williams 1981, 130). Also "an individual person has a set of desires, concerns or … projects, which help constitute *character*" (Williams 1976, 5).

to values is what allows us to distinguish between ascriptions of responsibility and mere grading evaluations.

Along these lines, Levy might be right that I don't separate 'bad' from 'blameworthy'. I don't do so because I think blameworthiness is appropriate only within the practices of accountability: it involves reference to someone's owing it to others to behave in certain ways, and distinctive disappointment and reproach when one's expectations and demands are unsatisfied. Thus, I think that 'blameworthiness' is too narrow to capture responsibility-involving criticism in general; it refers only to a part of ascriptions of responsibility, the one concerned with accountability. Put otherwise, Levy's worry seems to stem from an overly weak notion of 'bad' and an overly strong understanding of 'blameworthy', where the former is interchangeable with 'merely defective' and the latter involves reference to sanctions[19] and control. But this begs the question: it assumes that only sanctions and reference to control can identify the proper *locus* of responsibility.

Yet, at this point my opponent is likely to be still unsatisfied; they would contend that I simply pushed the question one step down the line. For why does the fact that my actions express my evaluative orientation make them something for which I am *responsible*? It has to be specified, they would say, why reference to my evaluative orientation can guarantee that we are operating within the realm of responsibility in a full-blown sense. Once again, it seems to me that this question seems urgent only if we accept an accountability-centred perspective. As such, it might not be genuine. In fact, asking the advocate of the aretaic model why I should be responsible for what expresses my evaluative orientation would be like asking the advocate of the accountability view why I should be responsible for what is under my control. In a sense, this is asking *too much*. It is difficult to see how the question *could* be put to rest. They might respond something like 'Well, *that* is precisely why; it's under my control!'. Similarly, I am tempted to say, I can be responsible for my actions *precisely because* they express my evaluative orientation. When we evaluate someone for their beliefs, we assess who they are as a person. We commend or criticise them for being someone who's generous, kind, dishonest, envious, arrogant. When we do so, we are not simply admiring their thick hair or high cheekbones. We are evaluating the quality of their values, how they see the world, what moves them. Reference to values delimits the range of the things that can be predicated of me in the relevant way.

In effect, it might even be possible to turn McHugh's and Levy's worry around on them: by understanding responsibility-involving evaluations as dependent on compliance with norms, it's *their* proposal that risks treating ascriptions of responsibility as akin to grading evaluations.

[19] For criticism against the view that blame is or necessarily involves sanctions see for instance Hieronymi (2004) and Owens (2012, ch 1).

Someone is blameworthy insofar as they rate poorly against some standards, precisely as a slow horse rates poorly as a horse, during a race.

Before proceeding, let me make one final preliminary point. When I say that someone is responsible for who they are as a person, I'm restricting 'who someone is' to their evaluative orientation and to their values. This is a technical use. For of course also someone's appearance, or ethnicity is part of *who they are*. This is particularly clear in extraordinary cases, *e.g.* someone who's particularly short or particularly tall, someone who's blind, or colourblind, someone who's particularly attractive or unattractive. Of course, these features would play an important role in who someone is as a person, and yet they are not values, nor expressive of one's values, in and by themselves. I don't think this poses any concerning worries for the view I'm suggesting. In fact, what my view is committed to say here seems to reflect our intuitions quite accurately: I am not responsible for being a dwarf or a giant, insofar as *that* doesn't display any of my values, though I am responsible for the beliefs and actions that express my evaluative orientation, which may be affected in important respects by my height. To put another way, if I resist the view that we need to control something in order to be responsible for it, and I say that a very attractive person is responsible for being arrogant, I am not saying that they are responsible for being born attractive. They are responsible for being arrogant, which may have been affected by their physical appearance.[20] But it's not that they are not responsible for being born attractive because they couldn't control it; rather it's because being attractive does not express their values, in and by itself.

Once we accept that I am responsible for what expresses my evaluative orientation, the important question arises why I am sometimes criticisable and sometimes commendable for some of the things for which I'm responsible. The advocate of the accountability view addresses this question by reference to norms; their violations and abidance. I discuss my alternative in chapter III. First, I articulate the notion of 'evaluative orientation' in terms of virtues and vices, which I take to be distinctive patterns of being sensitive to reasons. Each person's evaluative orientation incorporates different virtues and vices, though it's not exhausted by them. In fact, there are also values and sensitivities to reasons that are not obviously good or bad to have, *e.g.* valuing etiquette, and that can still be expressed by our beliefs. Thus, having certain values can justify both positive or negative appraisals, *e.g.* if we value etiquette or privacy. By contrast, having some values *always* issue positive appraisals, and those are the virtues. And vices, which can be either *anti*-values or *lack* of values, always call for negative appraisals.

[20] That is, the fact that one's values might be affected by factors that we cannot control does not threaten my view.

In chapter III, I also show how taking virtues and vices seriously can explain why we are not always *epistemically* (or morally) responsible for our beliefs. In fact, a virtue-centred framework introduces a plurality of dimensions of assessments that go well beyond the merely epistemic or moral realms. It is because of this that I can argue that *doxastic* responsibility should not be reduced to *epistemic* responsibility. This is an important point, for even if someone resisted the claim that the kind of aretaic appraisals on which I focus are genuine markers of responsibility, I will still have made a case for the key thesis that beliefs can be normatively assessed in richer and more varied ways than merely in terms of what one ought to believe, or in terms of truth, justification, or knowledge.

In chapter IV, I discuss the control requirement and argue that it originates from the tendency to start a reflection on responsibility by looking at the case of intentional actions. Even though we need to exercise some sort of control over our actions in order to be responsible for them, I think we should be weary of generalising its scope across other domains, and I deny that it applies to the case of beliefs. After having discussed the roots of the control requirement in more detail, I clarify how the aretaic model can explain responsibility for beliefs without resorting to it. In the second part of the chapter, I turn to assessing what I take to be the two most influential views on the kind of control involved in responsibility for beliefs, *i.e.* the indirect voluntary control views and the direct nonvoluntary control views. I raise some worries regarding both and argue that it's unclear whether they can actually meet the control requirement they set out to meet. If I succeed in showing that these proposals fail to address the question how we can be responsible for our beliefs, *given that we cannot exercise control over them*, I will have provided one more reason to think that the question itself is ill-posed.

In the first four chapters I remain fairly neutral on the precise strength of my thesis, *i.e.* on the question whether the aretaic model is meant to *supplant* or *complement* the accountability view. That's because I focus on the practices of attributability for beliefs, which are largely neglected by the advocates of the accountability view. However, nothing of what I claim in the first four chapters entails that the accountability view doesn't offer the best possible explanation of at least *some* cases, *e.g.* those belonging to the practices of accountability for beliefs. I consider this question in chapter V. First, I investigate whether there are any practices of accountability for beliefs, and how they'd look like. I argue that holding others accountable for what they believe is a genuine phenomenon, which involves the relevant sort of demands and expectations. I then offer two different ways of understanding these practices. Amongst these, I lean towards the idea that we wrong someone when our beliefs are *about* them; the extent to which they can expect and demand us to revise our beliefs depends on several factors, amongst which the nature of our relationship. This explains why

in PLANT my housemate is entitled to *expect* it from me that I don't jump to conclusions when I assess whether *p*, and *p* is about them, especially since we are in a certain kind of relationship. My belief can be taken to wrong my housemate, and I *owe* it to them to revise it. Once it's established that we can hold others accountable for their beliefs, I suggest a way of making sense of it from within the aretaic model. In particular, I suggest we do so by distinguishing between *personal* and *interpersonal* virtues and vices. We need not buy into the accountability view to illuminate these cases.

The aim of my thesis is not to argue *against* the accountability view. Rather, it is to draw attention to cases that are central to our lives and yet seldom discussed. My hope is to widen the scope of the discussion on 'responsibility for beliefs' and to bring out the complexity of a rich and important phenomenon.

II. The evaluative orientation of agents and believers

In chapter I, I made a case for the idea that there are two ways in which we can understand responsibility, *i.e.* the accountability view and the aretaic model. The former is committed to a control requirement, focuses on someone's adherence to or infraction of some relevant norms, and typically interprets the relevant criticism and commendation as involving demands that can be imposed on those who are held accountable. By contrast, the aretaic model is built on the idea that someone is responsible for what expresses something of them. The overarching aim of this chapter is to clarify this intuition and lay out the aretaic model in more detail.

One of the most serious attempts to articulate the expression intuition is Gary Watson's *Two faces of responsibility* (Watson 1996). In section 1, I start by looking at Watson's discussion of the notion of 'attributability' and at his claim that, from an attributability point of view, an agent is responsible for an intentional action insofar as it expresses their adopted ends.[1] Call this claim Watson's Thesis (WT). Put otherwise, WT states that there is something about *expressing my adopted ends* that makes my actions something for which I am responsible. This is a technical use of 'expressing'. My glasses could also be said to express that I am short-sighted but I am not responsible for them as I am for some of my actions. I call 'expressing relationship' the responsibility-involving relationship that holds between actions and agents. My task is to investigate its nature and argue that it holds also between *beliefs* and believers.

[1] "But if what I do flows from my values and ends, there is a stronger sense in which my activities are inescapably my own: I am committed to them. As declarations of my adopted ends, they express what I'm about, my identity as an agent. They can be evaluated in distinctive ways (not just as welcome or unwelcome) because they themselves are exercises of my evaluative capacities" (Watson 1996, 233).

Very briefly, I explain WT by developing the notion of '*evaluative orientation*'. My evaluative orientation is my sensitivity[2] to see certain things as reasons[3] for actions and beliefs,[4] oriented by my values. I am responsible for the intentional actions that express my adopted ends *because* my adopted ends are exercises of my evaluative orientation. That is, it is because expressing my adopted ends guarantees that my actions express my values that I can be open to responsibility-involving evaluations for them. In fact, my evaluative orientation, and the values orienting it, is a constitutive part of who I am, which is the primary target of aretaic appraisals.

In light of sections 1 and 2, I discuss the case of beliefs in section 3. My suggestion is that we don't need to look for an *analogue* of WT that is applicable to beliefs. For if someone is responsible for the actions that express their adopted ends because adopted ends count as exercises of their evaluative orientation, then we might simply say that someone is responsible for the beliefs that express their evaluative orientation. In both the action and the belief case, we are in the realm of responsibility as soon as our actions and beliefs express ourselves and our values in the relevant way. Not only practical but also theoretical reasons can display someone's values. I develop this thought by considering Bigotgate, a media scandal that involved Gordon Brown('s beliefs) a decade ago.

To conclude (section 4), I address a possible objection to my view, *i.e.* the *one-instance worry*. In fact, someone might resist the claim—central to my proposal—that we need to shift focus from local evaluations of particular beliefs to global appraisals of persons, on the basis that we are rarely in the position to look at someone's *whole* set of beliefs. If so, how can we know that someone's

---

[2] I talk of 'sensitivity' rather than of 'capacity' because it makes better sense of the cases in which you are *not* sensitive to certain situations. For instance, we can explain the fact that you don't notice my sadness by reference to your distinctive sensitivity, because being insensitive or tactless can be part of it, but we cannot explain it by reference to your *capacity*, because the 'capacity of being insensitive' does not qualify as a *capacity* at all. However, I will still say that one's adopted ends and attitudes are *exercises* of one's sensitivity, in the sense that our sensitivity is operative (rather than temporarily impaired).

[3] When I say that the evaluative orientation that is displayed by your actions is your sensitivity to *reasons*, I refer to *motivating* reasons, *i.e.* to (putative) facts in the light of which you acted. Your evaluative orientation identifies the way in which *you* take certain things to count against or in favour of acting as you do. Sometimes, 'I did A for the reason that P' is true because P is the correct non-normative explanatory reason for the action. This *has* to be so, for I will argue that we can deem you wicked, lazy, dishonest, or unkind by looking at your evaluative orientation as it is displayed by your actions. And what you rely upon in acting in wicked, lazy, dishonest or unkind ways cannot be 'normative' reasons, though you act in their light and they can explain your conduct (for the distinction between motivating and explanatory reasons see for instance Alvarez (2010, ch. 2).
However, in this thesis I remain neutral on the question what kind of thing a 'normative' reason to act is, *i.e.* whether it must be a fact, a mental state or a mental fact, or either a fact or a false belief. For a defence of Factualism see, amongst others, Alvarez (2010; 2018), Dancy (2000), Hyman (2015), Littlejohn (2014), Raz (1999, ch. 2). For a defence of Psychologism see esp. Davidson (1963). For a disjunctivist position see Dancy (2014), Hornsby (2008), Comesaña and McGrath (2014).

[4] I take reasons for beliefs to be standard truth-related ones. Thus, here they cannot be *motivating* reasons. Crudely, the idea is that the distinctive way in which you are sensitive to reasons for beliefs identifies the way in which you take certain things to count in favour of or against the truth the *p*. I discuss this point in section 3 and I say more about the differences between theoretical and practical reasons in chapters III and IV.

particular belief is really representative of their overall evaluative orientation? I think this worry lies on a misunderstanding of my proposal, which I hasten to dispel.

1. Watson's Thesis and attributability

According to WT, an agent is responsible for their intentional actions insofar as they express the agent's adopted ends.[5] Watson introduces WT as an alternative to the accountability view. In fact, the expressing relationship between my actions and my adopted ends fulfils the same responsibility-grounding role that the control requirement plays within the accountability view. They are two different ways of justifying why *I* can be deemed responsible for an action or attitude. Those who accept a control requirement claim that I am responsible for something insofar as I can (or could have) exercise(d) control over it. Those who reject a control requirement have to provide an alternative criterion. The attributability sympathiser resorts to WT: I can be responsible for something insofar as it expresses something of me in the relevant way. In virtue of this very disagreement, Watson claims, accountability and attributability capture different 'parts' of "our idea of responsibility" (Watson 1996, 229). The two kinds of ascriptions of responsibility are made intelligible by different features of the person who is responsible.

To illustrate: suppose that we are very close friends and I accidentally come to know that your partner is less honest with you than you expect. As soon as you entertain some doubts, you come to me and look for reassurances, which I promptly provide. The thought of seeing you hurt once again is unbearable, and I convince myself that maybe things are not as bad as they look. After some time, however, not only do you come to know of your partner's dishonesty but also of my connivance. I knew and let you in the dark. I betrayed you. Nevertheless, your reactions are likely to be ambivalent, and the distinction between accountability and attributability is meant to make sense of these tensions. On the one hand, you are likely to resent my lies. The advocate of the accountability view might say that friendships is regulated by norms that determine what is permissible and prohibited within its boundaries, even though they are not legally binding. You might blame me for violating them. I could have told you, and I should have done so. However, by lying, it's not only that I violated the norms of friendship. My lies matter insofar as they are testament to my values and commitments. And according to *why* I lied,[6] my lies can mean different things to you. For instance, if I lied because I didn't want you to be hurt, they indicate affection toward you. I did something wrong in order to shield you from a bitter disappointment, maybe

[5] That I start off from the case of action might seem at odds with the fact that I insisted, in chapter I, that the accountability view is problematic precisely because tailored on actions. However, what does the explanatory work in WT is not characteristic of agents and actions. Thus, starting from the case of action does not entail that I'm modelling my proposal on it; it simply offers an easier way to see what underlies the case of beliefs as well.

[6] That is, according to what considerations I took to be good reasons for me to answer your questions as I did.

whilst hoping that your partner would have behaved better in future. By contrast, if I lied out of fear of your reactions, then my lies may be evidence of cowardice, or of my fear of confronting strong emotions. They can be evidence that I am not a good friend to you—I am someone who deems it more important to escape a difficult conversation than your long-term well-being. Because of this you'd be disappointed, think I'm uncommitted and detached, lose faith in me.

Accordingly, the kinds of responsibility-involving evaluations issued by the accountability view and by the aretaic model have different features. One is this. The emphasis on control leads the accountability view to be concerned with quite *local* assessments; it is generally to do with occurrent evaluations. Depending on whether you controlled (and could have controlled) φ-ing, I hold you accountable for this specific instance of φ-ing. For instance, you might be a very reliable and loyal friend, but you are to blame this time because you broke your promise. The intensity and durability of my evaluations of your behaviour depend on its gravity (a broken promise or a backstabbing would carry different weight) but the positive or negative evaluations of your conduct do not necessarily transfer to you *as a person*.

By contrast, aretaic appraisals shift the focus to the agent, their ends and values. They are *global* appraisals, of the whole person. In Watson's words,

> …conduct can be attributable or imputable to an individual as its agent and is open to appraisal that is *therefore* appraisal *of the individual* as an adopter of ends. (Watson 1996, 229) [My italics]

To clarify, that my action expresses my adopted end does not mean that I must have performed an intentional action of adopting ends, of which my adopted end is the result. Ultimately, this thought would lead to the collapse of the aretaic model onto the accountability view, since our assessment of someone's ends would risk depending on how they carried out the action of adopting them.[7] Rather, *adopted* ends are those that reflect one's sensitivity to reasons, or that are informed by one's values.[8]

I think that one helpful way of rendering Watson's argument is to make the intermediary role of the notion of 'values' more explicit. Consider the following:

1. My action is a "declaration" of my adopted end.
2. My adopted ends express my values.

[7] Alternatively, we could then ask *why* they carried out the action of adopting them. This would bring the focus back on the values that orient one's actions. However, it would also risk leading to a regress.

[8] This might be one point of disagreement with Watson, who seems to have a stronger reading of what it is to adopt ends. For instance, he writes: "…to adopt some ends among others is to declare what one stands for" (Watson 1996, 233) and "To stand for something is to take a stand, to be ready to stand up for, to defend, to affirm, to answer for" (Watson 1996, 234). I suspect that Watson leans towards these strong formulations because he is talking of *moral* responsibility specifically. Below, I clarify in what sense I take there to be an element of endorsement in having values.

3. My values are inescapably my own.

4. I am responsible for what is my own.

5. I am responsible for my values.

6. I am responsible for my adopted ends.

7. I am responsible for the actions that express my adopted ends.

Watson presents 1-4 as self-evident. To deduce 5 and 6, he employs a sort of responsibility-transitivity derivation rule according to which, if I am responsible for what is my own and my values are my own, then I am responsible for my values, and if I am responsible for my values and my values determine what my adopted ends are, then I am responsible for my adopted ends too. The argument works if we grant a certain link between me, my values, my ends and my actions. I am responsible for the actions that express my ends *in virtue of* the fact that my ends reflect my values, which are part of who I am, which is the ultimate target of aretaic appraisals.

This version of the argument can be extrapolated from several passages of Watson's work. For instance, he commits himself to 1, 3 and 6 in the following passage:

> …if what I do flows from my values and ends,[9] there is a stronger sense in which my activities are inescapably my own: I am committed to them. As declarations of my adopted ends, they express what I'm about, my identity as an agent. (Watson 1996, 233)

He suggests 3-5 in this one:

> …the point of speaking of the 'real self' is not metaphysical, to penetrate to one's ontological center; what is in question is an individual's fundamental evaluative orientation. Because aretaic appraisals implicate one's practical identity, they have ethical depth in an obvious sense. (Watson 1996, 234)

And I think he attempts to elucidate 2 here:

> …ends in the relevant sense are necessarily self-disclosing, since they are direct exercises of moral capacity. (Watson 1996, 234)

However, Watson's remarks remain more underdeveloped than one would hope, especially on the exact relationship between someone's 'values' and their 'ends'. One might also wonder whether Watson's formulation that actions 'flow' from the agent's values means that actions instantiate values, are caused by them or constitute them. Similarly, the nature of someone's 'evaluative orientation' and of their 'moral capacity' is quite indeterminate. And pending a more precise

[9] *Cf.*: "The free agent has the capacity to translate his values into action; his actions flow from his evaluational system" (Watson 1975, 216).

account of the latter, it's difficult to grasp what it means for adopted ends to instantiate it, and on what basis it is a *moral* capacity.[10]

To clarify, these are not primarily exegetical worries; I am not concerned with providing a correct account of Watson's own position so much as with offering an account of responsibility that draws on his remarks. In the next section, I attempt to shed some light on 1-6. In particular, on what it means for my ends to *express* my *values* and for my actions to *flow* from them.

2. *The evaluative orientation of agents*

I suggest that we illuminate the relationship between one's ends and values by developing Watson's notion of the 'individual's fundamental evaluative orientation' in the light of John McDowell's *Virtue and reason* (McDowell 1979). There, McDowell argues against a theory according to which it's possible to codify how one should live. He rejects the idea that there are principles and clearly formulated rules that someone could and should follow in order to behave well. Rather,

> Occasion by occasion, one knows what to do, if one does, not by applying universal principles but by being a certain kind of person: one who sees situations in a certain distinctive way. (McDowell 1979, 347)

To be a certain kind of person is to see situations in a certain distinctive way. And, I want to say, the distinctive way in which I see situations is my evaluative orientation—the distinctive way, oriented by my values, in which I am sensitive to reasons for actions (and beliefs).[11] According to what my values are, I take different things to be good reasons to do (or believe) something, and I am a certain kind of person.[12]

In sections 2.1 and 2.2, I clarify what it is to 'see situations in a certain distinctive way' and what I mean by 'values'.

---

[10] I address questions related to the *significance* of responsibility-involving appraisals in the next chapter, *i.e.* whether it's moral, epistemic or otherwise.

[11] Of course, phrasing things this way should not lead one to think that the distinctiveness of each of us can be reduced to how we *see* situations. First, we are different kinds of persons also insofar as we *interact* with the world in different ways. Yet, this can be explained by reference to the fact that we 'see things' differently. Second, having a certain evaluative orientation is also a matter of responding *in certain manners* to the things that we see in a certain way. For instance, if you value friendship, not only will you give a sincere opinion to a friend who asks for one: you will do so in a certain way, *e.g.* happily, willingly, without hesitation. As Hursthouse puts it, "we think of honest people as people who tend to avoid the dishonest deeds and do the honest ones in a certain manner—readily, eagerly, unhesitatingly, scrupulously, as appropriate" (Hursthouse 1999, 11).

[12] See Owens: "One is held to account for one's merit as a person; one gets blamed for those things (both moral and nonmoral) which are thought to make one a bad person" (Owens 2000).

2.1. 'Seeing situations'

That being a certain kind of person is a matter of 'seeing situations in a certain distinctive way' can mean one of two things. First, that you are sensitive to certain situations as conclusive *reasons* to act. Second, that you see some *features* of situations rather than others.

As for the former, it is common to understand reasons for action as facts that constitute a case for (or against) the performance of an action; that count in favour of performing it or show it worth doing. Conclusive reasons for actions are all-things-considered reasons; not overridden by other practical considerations. Yet, precisely what is meant by 'seeing', 'perceiving' or 'taking' something 'as a conclusive reason' for action is a difficult question. One suggestion is that you take *r* to be a conclusive reason to φ if:

a. you would φ if you thought that *r* were the case, lacking overriding reasons not to φ;[13]
b. you would not φ if you thought that *r* were not the case, and there were no other conclusive reason to φ;
c. you would be able to mention *r* if asked why you φ-ed.

where a-c. do not entail that, when you φ, you need to hold a conscious belief that you take *r* to be a conclusive reason to φ. Along these lines, to 'see situations in a certain distinctive way' is to be a person of such a kind that you would φ because of *r* (and be able to mention *r* if someone asked you why you φ-ed) but you wouldn't φ because of, say, *s* or *p*.[14]

For instance, suppose that you see someone's hunger as a conclusive reason for you to share your lunch with them.[15] And suppose that I see their hunger as well, which I too take to count in favour of sharing my lunch with them. Yet, I don't take it to provide a conclusive reason; for me, it is ultimately overridden by my own hunger, and I eat what I brought by myself. Now, both you and I see someone's hunger as counting in favour of sharing our lunch. But we are sensitive to it in different ways. Only you see it is as a conclusive reason to do so. Only you are a

[13] In fact, practical reasons are *presumptively sufficient*. See Raz (2011, 45): "For a reason to φ at a particular time to be presumptively sufficient is for it to be the case that if there is no other reason either for or against so acting then (a) φ-ing at that time is justified, and (b) if the agent rationally believes that the reason applies, and that there is no other, then his failing to try to φ is akratic". Raz argues that epistemic reasons, by contrast, are not presumptively sufficient. For if you have only one reason to believe that a particular proposition is true, and no reason against that belief, "coming to have that belief on the strength of that reason may be unjustified and irrational". That is, one good and undefeated reason to act is a good reason to perform the action; whilst one good and undefeated reason to believe might be an insufficient reason to warrant a belief.

[14] Here I remain neutral as to whether agents need to see reasons 'under the guise of the good'. For a defence of the claim that reason-explanation is not itself normative, see Setiya (2007), esp. part I. He writes: "One can act for a reason one does not see as good. Taking something as one's reason is taking it to be a reason that explains what one is doing, not a reason that justifies it" (Setiya 2007, 67). Setiya claims that versions of the claim that reasons for acting must be seen under the guise of the good can be found, amongst others, in Anscombe (1963, 21–25), Darwall (1983, 205), Bond (1983, 30–31), Broome (1997), Korsgaard (1997, 221), Wallace (1999), Velleman (1992).

[15] The reader will fill in the example in the appropriate way, *e.g.* you have a regular eating routine whilst they're undernourished.

person who would take someone's hunger to offer a conclusive reason to share their lunch with them.

Second, that to be a certain kind of person involves seeing situations in a certain way also means that different kinds of persons may find different features of situations salient. For instance, suppose that you still see that someone is hungry but that, in this case, I don't even notice that they are. It is not that I do not take someone's hunger to provide a conclusive reason to share my lunch. Rather, I am utterly incapable of reading the signals; it doesn't cross my mind that someone might be; or perhaps I simply never look at people because I find them uninteresting. That is, I am socially inattentive, or an optimist, or cold and careless. If you tried to make sense of why I don't even notice that person's state, you'd explain it by reference to 'what kind of person I am'. I am someone who doesn't notice this or that factor.

2.2. Values

I said that someone's evaluative orientation is their *distinctive* way of seeing situations. This doesn't mean that I am programmed to be moved by certain features of situations in each and every occasion. Rather, an appeal to 'distinctiveness' merely emphasises that there are characteristic and important similarities between the ways in which I see things at different times. And this is what allows us to say that I am *a certain kind of person* to begin with. I can be said to have a certain kind of sensitivity, in the singular, precisely because I see situations in ways that are consistent with each other.

My suggestion is that patterns of responses are consistent within themselves, and we can see them as instantiating a 'distinctive' sensitivity, because of the connection that someone's sensitivity has with their *values*. The thought follows from two main claims. First, persons have fairly *stable* values.[16] Second, they see situations according to what they value; someone's distinctive way of seeing situations is their *evaluative* orientation. Having fairly stable values is what makes us see things and react to them in traceably coherent ways; it is what allows us to be deemed a person of a certain kind.[17] Let me say something more about each of these two claims.

First, to understand the thought that persons have fairly stable values, we need to clarify the meaning of 'value'. If we set aside all the contexts in which it is used to talk about money, *e.g.* 'the

[16] One well-known challenge to virtue theory is *situationism*, the view in psychology according to which a great deal of human perception, deliberation, and behaviour is explained by trivial situational influences more than by personal traits. For an overview of the situationist challenge to virtue ethics see *The Case against Character* in Appiah (2008). For recent developments on *epistemic* situationism, see Fairweather and Alfano (2017).

[17] The particular *nature* of one's evaluative orientation determines *what kind* of person they are, as I argue in the next chapter.

value of a property', or in general to quantify over some measurements, *e.g.* 'the value of a musical note', there remain two main usages, *i.e.* when we say that something is *of value* and that when we say that someone *has certain values*. I am interested in the second sense.

When we say that *something is of value*, we take an object to have a property or a (positive) quality that renders it valuable. Insofar as it has certain characteristics, it has also the property of being valuable. Here, value belongs to the object. Sometimes, something is of value *absolutely*, or at least in a way that is largely accepted by a certain community. For instance, the Divina Commedia or Bach Violin Concertos are of value, full stop. They are of value (at least to all of us) because of what they are, how they have been written and composed, irrespectively of who reads or listens to them. If someone doesn't find them to be of value, that is not proof that they are not; it is proof that that person is mistaken. Other times, something is of value *relatively*, *to someone* in particular. For instance, your dog is of value to you. That is, your dog has the property of being valuable to you, and that property emerges on some of its features, *e.g.* its being affectionate, loyal, or simply yours.[18] Because of how it is, and because of how you are, you take it to be (more or less) valuable.

There is an incredibly large literature that discusses what it is for something to be valuable, and related difficulties. Philosophers work on the differences between intrinsic, extrinsic or instrumental value; the relationships between the property of being valuable and the property of being good; whether there are different kinds of values (*e.g.* moral and aesthetic) and whether they are commensurable. However, these debates are relevant for my purposes only tangentially, as I am primarily concerned with the second way in which we talk of 'value'.

The second sense is at issue when we say that *someone has certain values*.[19] For instance, we say that someone has different values to us; that honesty is amongst your values; that I have old-fashioned values. People possess values. More precisely, the fact that you have a value does not depend on your having certain other characteristics,[20] as being a dog that is valuable (to you) depends on its being affectionate and loyal toward you. The dog is valuable in virtue of having certain features that you find valuable; but you do not in the same way *have* a value in virtue of having certain features, nor in virtue of finding it valuable to have that value.

Precisely what it means for someone to have values remains a difficult question to address. It is easier to say what it does *not* amount to. For one thing, having a value does not mean that you submit yourself to a principle or norm that you need to follow. You might not even be able to

[18] *Cf.*: "It would be a serious mistake to believe that the importance of an object to someone is not fully genuine unless it is independent of his caring about the object. Consider the fact that many of the people we care about most would not affect us in important ways if we did not care about them" (Frankfurt 1988, 93).
[19] In the rest of the thesis, I will use the phrase 'someone has a value' in the semi-technical sense that I delineate here.
[20] To clarify, persons can also be said to be *of value*. For instance, a good football player is of value to their team in virtue of having speed, stamina and lateral sight. In this sense, they themselves are valuable, in virtue of having some other characteristics. But this is not to say that they have a value in the second sense I'm trying to individuate.

formulate your value as a norm. In fact, it seems that people can have values without even knowing that they do, under that description.[21] It is possible that someone is sensitive to certain reasons, in a way that we would recognise as being oriented by a value, without being able to master the corresponding value-concept.[22]

In fact, reference to your value *need not figure* amongst your reasons for your action or belief, in order for your action or belief to count as expressing it. For instance, suppose that, during an exam, I am given the chance of copying from the answer sheet, and I refuse. If asked why I do so, I might say that 'that would be cheating!'. There's no need for me to mention that it would be dishonest, and dishonesty is bad. It suffices that I see the fact that copying would be cheating as a good reason not to copy, for my action to express honesty. The point is that I perceive certain situations as reasons *because* honesty is amongst my values.

If I mastered the concept 'honesty', I could answer the question why I didn't copy by saying *e.g.* 'Because cheating is dishonest!', where this would imply that I consider being dishonest to be bad. However, I could *not* address that question by saying *e.g.* 'because honesty is amongst my values', or 'because I am honest'. The fact that I am honest cannot be the answer to the reasons-seeking question why I didn't cheat. If I did answer in those ways, that would either mean that I misunderstood the question and provided an explanation of my conduct instead; or it would cast some doubts on my being honest, since an honest person would not do something in the light of the fact that they are honest.

At this point, someone might resist the claim that I don't need to master the value-concept in order to have that value on the basis that there seems to be an element of *endorsement* in having a value. And if I must endorse having the value in order to have it, then I must be aware that I have it. Now, I agree that having values involve endorsing something. But I think that, first, this objection is based on a misunderstanding as to *what* we are supposed to endorse and, second, it gets the explanatory order the other way round. First, I suspect that the idea that we endorse having a value comes from the fact that, in several situations, we heatedly say things like 'I stand by this!', 'I'm committed to it!'. However, when we do, we insist on the rightness and force *of our reasons*, not (directly) of our having a certain value. When I say that copying would be cheating, or that copying is dishonest, I insist on it passionately; I stand by it. But here what I endorse is my reason: that the fact that copying is cheating is a good reason not to copy; or that the possibility of cheating is not a good reason to cheat. I don't endorse *being honest*, or *the fact that* honesty is amongst my values.

[21] Notoriously, for instance, valuing humility might even *require* that one does not know that they are humble.
[22] Nor is it a matter of *desiring* to have that value. For sure, desiring to have a certain value might suggest that you already have it to a certain extent, but the two things can come at different stages and the former is not sufficient for the latter.

Rather, by endorsing and standing by my reasons, which I see in a certain way precisely *because* I am honest, I instantiate and express my value. *Given that* I am honest, I endorse certain reasons in a certain way.[23] Being honest makes me see being dishonest as a bad thing. What we endorse when we have a value is not having a value itself but rather the way in which having that value makes us see things.

Rather than saying that we need to endorse having a value in order to have it, I suggest we say that having a certain value requires that you *would* endorse it if asked—on condition that you master the corresponding value-concept—or at least that you don't believe (nor desire) that you should (nor do) *not* have it.[24] But once again, if, when asked whether I endorse honesty, I give a sincere positive answer, that is because honesty is amongst my values to begin with.[25]

This much should shed some light on the sense in which we can be said to have certain values. However, to explain why your way of seeing situations is *distinctive* of who you are, I suggested that your values need to be fairly *stable*. For only if they are fairly stable, we can explain why we respond to diverse situations in ways that we can see as being consistent with each other. And as a matter of fact, it seems that we *need* to have fairly stable values. If there were no coherence between how we perceive situations, not only would we be unable to interact the way we do; also the very idea that we have a personality would seem to require important revisions.

To illustrate: let's say that every morning my housemate offers me some coffee from her Italian coffeemaker. Now imagine that one morning I gratefully accept and rejoice for her kindness; the morning afterwards I suspiciously decline and, as soon as she leaves for work, scrupulously examine the coffeepot, persuaded that it must be poisoned; and today I am convinced that her offer is nothing but a foxy attempt to induce me to offer her something in return. And suppose that I do not have any other interaction with my housemate during the day. Now, first of all, the lack of coherence between the ways in which I read her behaviour would make it impossible for her to say that she knows *me*. It wouldn't even make sense to postulate that "the relationship

[23] This opens interesting questions about education and acquisition of values that I cannot get into here. Briefly, however, I think it possible to show others why certain reasons are good reasons by looking at them and explaining how they work. This amounts to showing others what is the value that orients them.

[24] In *Unprincipled virtue*, Nomy Arpaly explores Huck Finn's struggles with his conscience, as depicted in Mark Twain's *The Adventures of Huckleberry Finn* (Arpaly 2003, 75–79). Huck Finn helps a slave despite believing what he's doing is wrong. Very crudely, Arpaly builds on this case to argue that we do not need explicit endorsement of moral principles in order to behave morally. If so, then my 'minimal' requirements would still be too strong. Someone might count as having a value despite believing that they don't, and even desiring that they don't. However, it's not clear in what sense someone could count as *having a value* whilst believing and desiring that they do not (Frankfurt 1982, 271).

[25] We could also say that someone has a value insofar as they *care* about something, where 'care' is understood along the lines of (Frankfurt 1982): "A person who cares about something is, as it were, invested in it. He identifies himself with what he cares about in the sense that he makes himself vulnerable to losses and susceptible to benefits depending upon whether what he cares about is diminished or enhanced. Thus he concerns himself with what concerns it, giving particular attention to such things and directing his behavior accordingly. Insofar as the person's life is in whole or in part *devoted* to anything, rather than being merely a sequence of events whose themes and structures he makes no effort to fashion, it is devoted to this".

between me and my housemate is the same in the three situations except from my reaction" because, precisely given the unpredictability of my reactions, it would be impossible for her to establish any meaningful relationship with me to begin with. Even more basically, regardless of the possibility of interacting with others, the absence of a fairly stable evaluative orientation would threaten the very idea that I am *a* person. I would not be able to understand my past behaviour, nor to predict my future one. I couldn't make plans, feel remorse or satisfaction, pride or disappointment. I would be alienated from my own memories. Even though not without exceptions, having a value will have to be a fairly stable aspect of who we are.

To conclude, it seems to me that the very fact that we can individuate *tensions* between ways in which I behave proves that we expect my actions to instantiate the same values. We expect them to point towards the same direction, so to speak. Even more, I think that we also expect our actions *and our attitudes* to instantiate the same values, and I take this to support the idea that both our practical and theoretical reasons are oriented by the same set of values. If there were two capacities of perceiving theoretical and practical reasons and they were not only distinct but also independent of each other, my taking a situation to be good enough a reason to believe that *p* and to act as if *non-p* would not raise any concern altogether.

Our values, so understood, can explain why our evaluative orientation is distinctive of who we are only if they do orient the ways in which we see situations. This is the second claim I set out to discuss. We see situations according to our values in this sense: if you *have* a certain value, you find certain things to be *of value* and you consider certain reasons compelling; certain actions worth doing; certain situations worth bringing about. You seek out instances of that value; you are interested in them when confronted with some; you contribute to instantiate it.

For instance, if I value friendship, I pay attention to signals of affection or lack thereof; I am attentive to my friend's well-being; I take their request for a sincere opinion to be a good reason to give them one. In general, I perceive situations in such a way that considerations that bear on friendship have more weight than other ones. And in fact, if someone claimed to value friendship and yet failed to be responsive to other people's well-being, showed disregard for their opinions and needs and were in general a cold-hearted and individualistic person, we would cast some doubts on the legitimacy of their self-ascription. Of course, this does not entail that, if I am prone to noticing violence, that shows that I value *it*. If having a value makes me receptive to what is relevant to it, then not only will I be attentive to what instantiates it, but also to what threatens it. I will notice my neighbours' eight cars not only if I value luxury but also if I value modesty. And if I react

to and notice every small instance of violence, it will not necessarily be because I value violence but also because I despise it.

I think that also the stronger claim is true, that you will not find certain things to be of value and certain reasons compelling if you do not have the corresponding value. An unkind person might carry out the same action as a kind person would, and they might even share some of their beliefs. But insofar as kindness is amongst the values of only one of them, they won't see situations in the very same way. They won't be sensitive to the very same reasons.[26]

To clarify, the claim that, if you have certain values, you will find certain things valuable does not mean that to have certain values *is* to find some things valuable. This would conceal a circularity[27]—if we accept both that someone sees situations according to what they value and that having a value is to see situations in certain ways. Instead, I take having values to be prior in the explanatory order. The fact that I have a certain value can explain why I see the world in a certain why, whereas the fact that I see the world in a certain way does not *explain* why I have a certain value. For instance, the fact that I value solidarity explains why I perceive someone's hunger to provide a good reason for me to share my lunch. But that I see someone's hunger in that way does not explain why I value solidarity. It expresses it, and shows that I do, but it does not explain it.[28]

The fact that values orient one's evaluative orientation also explains the sense in which someone's adopted ends express their values. In particular, I adopt ends either on the basis of reasons that I find compelling, or without considering any, simply in virtue of seeing situations in a certain way. In many cases, I adopt ends based on what I take to be compelling reasons. For instance, I decide to share my lunch with someone on the basis that I see that they are hungry as providing a conclusive reason to do so, and I act accordingly. But in other cases, I adopt an end for which I

[26] As far as kindness is concerned, that is.

[27] Notice that, in any case, I am not sure a circularity would be fatal to my project. It seems to me that it would be a challenge only for someone who's looking for an *analysis* of the notions of 'sensitivity' and 'value'. But my goal is not to provide as analysis as much as to illuminate the complex relationships between some key notions.

[28] Even though I claim that having a value has explanatory priority over seeing things in a certain way, insofar as the former can explain the latter but not viceversa, it is unclear to me whether it has more than *explanatory* priority. For it's unclear whether it's possible to have a value without being sensitive to reasons in the corresponding way. Rather, it seems that having values *needs* to be *instantiated* in someone's seeing things in a certain way. For instance, suppose that I value honesty in a world in which no one is able to lie. Here, I *could not* be responsive to reasons according to my value; there would be nothing that my valuing honesty could make me attentive to. Having or lacking that value would not affect who I am. In such a case, the fact that my value remains unexercised challenges the idea that I have it to begin with. And this makes it unclear whether I could be said to have honesty amongst my values at all. Of course, someone might argue otherwise by emphasising the importance of modal claims. But I am not sure whether being disposed to value honesty in another world would suffice to say that honesty is amongst my values now. Along these lines, though I maintain that having a value has explanatory priority over the way in which having that value makes me see things, I resist the view that one can have a value without seeing things in the corresponding distinctive way. At least in one sense then, I agree with Setiya's claim that "the connection between reason and virtue runs in both directions: it is a matter of reciprocity, not priority" (Setiya 2007, 5).

have no reason other than that I find it worth pursuing. For instance, suppose that I want to learn how to play the violin as an adult, with no expectation and no will to play for or with anyone else. I have no reason to learn other than that I like it and feel inclined to do so, or that I find it worth my time. I adopt that end because I am someone who values music, beauty, challenges, or self-improvement. Having certain values makes me someone who finds learning how to play the violin a good end to adopt. Either way, both if I adopt ends by considering reasons or without considering any, my values orient my adoption of ends. If I were someone else, with different values, I would adopt other ends.

Along these lines, I can also give a better sense of the idea that having an 'adopted end' is not a matter of carrying out an intentional action of adopting ends. In fact, I do not see situations in a certain way intentionally, nor do I intentionally consider certain situations to be reasons. When we talk of *adopted* ends, we do not mean that I have some ends as a result of something I do intentionally. Rather, my ends are adopted by me in the sense that they reflect how I see things and, ultimately, what values I have. I think *this* is what it means for my actions to "flow from my ends" and "express what I'm about, my identity as an agent". And I think Watson might be making this very point here:

> To adopt an end, to commit oneself *to a conception of value* in this way, is a way of taking responsibility. (Watson 1996, 234) [my italics][29]

Adopted ends reflect a certain 'conception of value'. Values are the key middleman between me and my actions. They are integral part of who I am, and they guide me in adopting certain ends, which issue corresponding actions. Because of this, my intentional actions, either as means to achieve an end or as constituting such achievements, make manifest the kind of person I am.

2.3. My evaluative orientation and 'who I am': possible counterexamples

I want to make a final point about evaluative orientations, before turning to the case of beliefs. So far, I have assumed a constitutive link between 'who I am' and my evaluative orientation. If this doesn't hold, then someone could grant that my actions express *my evaluative orientation* and yet resist the claim that *I* am responsible for them. Some familiar situations could look like valid counterexamples to this connection.

First of all, someone might suggest that there are some actions that 'express who I am' despite lacking any connection to my evaluative orientation. For instance, suppose that someone you know, Matt, is easily excitable. Every time something thrilling happens, Matt becomes frenzied.

[29] Even though in this passage Watson suggests a very strong interpretation of what it is to adopt ends, see fn 8 above.

He gets louder, starts to gesticulate and move in a very agitated way. As a result, Matt becomes very clumsy. He is likely to drop, hurt or damage something. His friends know this feature of his. They amiably mock him for it, predict his behaviour and are surprised if he doesn't act as clumsily as they'd expect. They would all agree that being clumsy "is part of who he is". Importantly, they would also sometimes criticise him for his behaviour. For instance, when he scares away a newcomer, or embarrasses them in front of someone important, or damages something. And yet, my opponent would insist, Matt's clumsiness does not express his evaluative orientation in any clear sense. For he has not adopted the end of being clumsy. Being clumsy does not express Matt's evaluative orientation as one of his intentional actions does. Thus, Matt seems to offer a case in which some feature (clumsiness) is part of who he is, he is criticisable for it, and yet it doesn't express his evaluative orientation.

I think that this case seems to provide a counterexample only insofar as it is underspecified. In some circumstances ascriptions of responsibility would simply be out of place, whilst in other ones Matt could be criticised in a responsibility-involving way for his clumsiness. The distinction between the two sets can be drawn precisely by reference to his evaluative orientation; responsibility-involving appraisals will be appropriate only when his clumsiness tells us something of Matt's values.

When Matt is reprimanded for breaking a glass whilst gesticulating, that will not be the kind of responsibility-involving criticism I am trying to capture. Obviously, he caused the damage and is liable to compensate the owner for it, but his action does not reflect badly on him; no one would evaluate him *as a person* for that. If he's held responsible for the damage, he is only in the causal sense in which weak embankments and heavy rain can be responsible for a flood, or a dog for stumbling on a fragile vase.

But in other cases, Matt's clumsiness *can* reflect poorly on him. Suppose that Matt is your partner and you invite him to an important business dinner with the high ranks of your office. He feels the pressure and, in an attempt to look brilliant and elegant, seeks out serious conversations with your boss. He gets heated up, starts to lose his temper and raises his voice. Whilst dining, he spills the wine on the tablecloth several times, stains his shirt with the roast's gravy and—at that point quite agitated—enters the ladies' restroom by mistake. Now, even though his conduct was not intentional, and even less did he exercise *control* over it, you would be embarrassed. Maybe you would avoid the topic with your colleagues the day afterwards, and deliberately avoid inviting him to future events. Also, depending on your character, your embarrassment might grow into resentment: you might indeed criticise him for having been so awkward and clumsy. Or, by contrast, you might take it as a sweet (though ineffective) attempt of being accepted. I think that

the reason is this. Matt's clumsiness does not express his values in the sense that *being clumsy* is amongst Matt's values. Yet, it can express something of his evaluative orientation insofar as it shows something of the way in which he sees situations and reasons. For instance, that he perceives a business dinner as an anxiety-generating situation; that he wants to impress his partner's boss; that he cares about giving a good first impression to her colleagues; that he cares about being held in high esteem. And we can evaluate him in deep ways *for that.*

My opponent could then insist that there are also cases in which one of my actions is informed by my values and yet does not express who I am. One familiar instance of this kind of cases could be a person who's been brought up within a certain society, or family, from which they have absorbed a general view of world. We can think of a criminal who's been raised in an abusive and violent way, or of a housewife who lives in the suburbs of Palermo in the 50s, or of a home-schooled kid raised in a hippie sect in California. Each of these persons will have quite a peculiar sensitivity to reasons. Respectively, they will take your weakness to be a good reason to assault you, their husband's opinion to be a good reason not to take on a job they like, or schools' rigour to be a good reason to protest against their monopoly. Now, my opponent would say, despite expressing their evaluative orientation, their actions would not express who they are. For they didn't *choose* to have that particular orientation, nor were they in the position to control it. They could not help it. At most, the values they express and the ends they adopt would be indicative of the nature of the *society* in which they are immersed.

I will come back to the importance of avoidability when discussing control in more detail (chapter IV). But let me anticipate a main point here. It's not clear to me why a certain sort of upbringing can justify the claim that "their actions would not express who they are". For one thing is to say that they could have not done more to become better persons because of a certain education and environment—or that they could have, despite it. But another matter is to claim that, *because* they could not help it*,* then those actions cannot really express who they are. First, if generalised, the claim risks generating the implausible thesis that no one's actions can express who someone really is, since *everyone* can be understood as irremediably determined by some upbringing or another. To avoid a generalisation, it would be necessary to find a principled way to distinguish between the Palermitan housewife's upbringing and my own *such that* hers makes her actions something for which she cannot be appraised aretaically whilst mine does not affect my responsibility for my actions. If the distinction is supposed to be that she was a victim of a certain evaluative system, it seems to me to sin of historical arrogance on our part. Second, it is not clear what someone who pushes this line might mean with 'expressing who they really are', if someone's

historical and social context and education is taken to *threaten* someone's identity rather than contributing to its constitution. There seems to be an assumption that, if someone only had the chance to grow up within another community, then they would have finally been able to become or express 'who they really are'. But this, at best, would return quite a narrow and metaphysically mysterious understanding of who someone is.

Finally, there are those cases in which the link between someone's actions and their evaluative orientation is temporarily severed. The question is whether, for instance, someone's altered state would rule out the possibility of deeming them responsible for something. For even if those actions expressed *some* values, would they be *the agent's* values? One interesting instance of this kind of cases is drunkenness. Being drunk is typically taken to provide a valid excuse from ascriptions of responsibility: what I do when I am intoxicated cannot be something for which I can be deemed responsible in any strong sense.

The accountability model explains the legitimacy of this excuse by appealing to the fact that alcohol impairs someone's capacity to *control* their actions. That capacity being impaired, the control requirement cannot be imposed and therefore responsibility cannot be ascribed. At most, they'd say, someone can be indirectly responsible for their actions according to whether they could have controlled getting drunk in the first place. For instance, someone would be open to criticism if they drank too much whilst being aware of their poor resistance to alcohol. However, they would *not* be open to criticism if they were intoxicated by someone else, *e.g.* if someone alters their drink and they are neither aware nor in control of it.

This is definitely a distinction we want to be able to explain. And yet, the accountability view's proposal seems to leave another, important part of the story unexplained. Usually, we are not merely concerned with how and why someone ended up being drunk, or with whether they could have avoided it; we often criticise others also for what they do and say whilst drunk. My claim is that the aretaic model is as able as the accountability view to explain the difference between how we evaluate someone who drank too much and someone who was forced into getting drunk and that, furthermore, it can also illuminate why we appraise them for how they behaved whilst being drunk.

To begin with, the aretaic model can explain the difference between someone who drank too much and someone who was forced into getting drunk by drawing on the fact that the former's action, but not the latter's, is connected in the relevant sense to their evaluative orientation. There are innumerable situations in which I can happen to drink too much, and I can do so for as innumerable different reasons. It might be because I am very happy and want to celebrate, or very

sad and willing to drown my sorrows. My drunkenness can express something of the person I am; it can tell others something of what strikes me as unbearable, or of what makes me feel over the moon. By contrast, if someone else gets me drunk, my state does not depend on my evaluative orientation in the slightest sense. I would have never adopted an end that involved, required or allowed being drunk in this situation. I would have not taken any considerations to support the action of getting drunk.

But someone who's drunk can also be open to criticism or commendation for what they do and say whilst drunk, sometimes regardless of how it came to be that they are drunk. Here it seems worth distinguishing between degrees of drunkenness. Let us say that I am tipsy. Alcohol removes certain inhibitions that are usually an important aspect of my personality. Now, if during a business dinner I start talking with your boss and end up underlining what I think are your major defects, it's unlikely that the following day you will criticise me solely *for having drunken* too much. Even if, at that point, I was not in control of what I was saying anymore, there is still a strong sense in which you would resent what I said itself. Conversely, but similarly, if I confessed my love for you after a beer too many, the fact that I am in an altered state would not prevent you from feeling happiness (if you reciprocate) or awkwardness (if you don't). The same goes for any sign of affection that I show towards you. If I hugged you, or gave you my coat whilst waiting in the cold, or spoke highly of you with someone else, you may have the feeling that we are closer, that you know me *more*. The relationship between us would be likely to be tighter and more open, even though I did not control my actions. After all, that *in vino veritas* has been well-known since the ancient Latins. Sometimes lack of control, to a certain extent, does not hinder the connection between my actions and my evaluative orientation but rather makes it more visible. An altered state like tipsiness does affect my judgement as to whether or when to say and do something, and it can make me *over*react or take some situation to be *more* salient than I would normally consider it. However, it might not affect *which ones* I would take as salient, or the nature of what I would do and say.[30]

It is hard to say *when precisely* tipsiness stops being revelatory and starts to undermine ascriptions of responsibility, and probably this is not so interesting a question either. Of course, if someone is barely conscious, taking what they do to be indicative of who they are is misplaced. But the point is to decide *why* that is the case, rather than when. And my claim is that it need not

[30] A difficult case to assess is the one of dementia. People with the diagnosis of fronto-temporal dementia have been known to act in very hurtful ways to their loved ones. The accountability view excuse them from ascriptions of responsibility on the basis that they cannot control their behaviour in the relevant sense. In fact, for many this provides a strong reason in favour of the accountability view. However, it seems to me that, if we say that their attitude *can never* be something for which they are responsible, or that it can never be something that *reflects who they are*, then we risk not treating them as persons at all.

be explained by reference to their lack of control. Rather, it may be because being severely drunk impairs the connection established by WT between their actions and their evaluative orientation.

3. The evaluative orientation of believers

I want to argue that the expressing relationship that figures in WT, as I have delineated it in section 2, can explain responsibility *for beliefs* as well as responsibility for actions. Crudely, as someone can be responsible for an action insofar as it expresses their adopted ends, which are exercises of their evaluative orientation, so someone can be responsible for a belief insofar as it expresses their evaluative orientation. In this section, I clarify what it is for a belief to express someone's evaluative orientation.

3.1. Bigotgate

Consider an example. In 2010, Prime Minister Gordon Brown found himself at the centre of a media scandal, which is now known as 'Bigotgate'. During his electoral campaign, he was interviewed live on TV in Rochdale about Labour's plans to cut the deficit. A woman in the audience, Mrs. Gillian Duffy, repeatedly challenged Mr. Brown on hot topics such as immigration and public debt, but her intervention was ignored. Pressed by his entourage, Mr. Brown then met Mrs. Duffy and engaged in conversation. After what must have been quite a long and tiring discussion, Mr. Brown was inadvertently still wearing a radio microphone as he got into his car and was caught calling the pensioner "a bigoted woman". More precisely, he said: "She was just a sort of bigoted woman who said she used to be Labour, I mean, it's just ridiculous". The story generated widespread media attention and seemed to have had a serious negative impact on the election results. Mr. Brown was heavily criticised for his belief.[31] He was criticised, for instance, for being arrogant, judgemental, or at least unkind. His belief betrays a sense of superiority, perhaps a prejudice towards a certain social class.[32]

Before proceeding, let me clarify what I take this case to be an example *of*. In chapter I, I claimed that we can distinguish between the practices of attributability and the practices of accountability. The marker of the latter is that someone *wrongs* or does right by someone else. In chapter V, I will argue that the most typical case in which someone wrongs someone (else) by believing is when they form or hold an unjustified belief about them—one that is not supported by evidence, or that they formed hastily, without caring about getting things right. Along these

[31] Of course, the fact that people criticised him does not prove that he is *rightly* open to criticism. I will come back to the normativity of aretaic appraisals in chapter III, esp. section 3.1.

[32] To some, the very fact that he used the term 'bigoted' might come across as a sign of arrogance, since it overly denotes a certain kind of unbringing.

lines, it might seem that Bigotgate is an example of the practices of accountability rather than attributability. I want to make two points here. First, even if it were,[33] it wouldn't constitute a problem for my view. In fact, I will argue that the aretaic model can illuminate the practices of accountability as well as the practices of attributability. But more importantly, second, Bigotgate doesn't seem to involve a *wrong* in any clear sense. Gordon Brown's belief is *not unjustified*. According to the Oxford Languages Dictionary, a bigoted person is someone who's "obstinately or unreasonably attached to a belief, opinion, or faction, in particular prejudiced against or antagonistic towards a person or people on the basis of their membership of a particular group". Duffy's remarks, especially the ones on immigration, do support the idea that she is prejudiced against certain groups. And the difficulties encountered in reasoning with her do support the idea that she is perhaps *unreasonably* attached to her beliefs; at least, that it's very difficult to convince her to revise them. In fact, according to many people, Gordon Brown was not believing *falsely* those things about Mrs. Duffy. If so, then the case belongs to the practices of attributability, and the kind of criticism at issue is not an instance of 'holding accountable'.

It seems difficult to make sense of the sort of criticism involved in Bigotgate from within an accountability perspective. In fact, the accountability view could explain why Mr. Brown can be held responsible for *uttering* the recorded sentence. Brown's collaborators, especially those who organised his electoral campaign, might have resented him for ruining months of work out of carelessness. After all, he *could have* double-checked that his microphone was disconnected, or waited to steam-off in the safety of his own home. A similar emphasis on Mr. Brown's *utterance* was placed by some of those who tried to defend him. For instance, Mr. Burnham commented on *BBC 2's Daily Politics* that "we make off-the-cuff comments that we regret afterwards". Sometimes we *say* things that we wouldn't normally endorse. Hastiness, tiredness or exasperation may lead us to make remarks that we regret afterwards. Mr. Brown should have not *said* what he did about a voter, and he regrets voicing it.

However, suppose that Brown attempted to apologise to Duffy by saying that he "didn't meant to say it". Suppose that he stressed that it was a private conversation and he had no idea it was recorded, and that, if he knew, certainly he would have avoided uttering that sentence. Mrs. Duffy would not be satisfied with this. Rather, to clear his name, Brown had to say that he didn't *believe* that she was a bigoted woman. His spokesman announced soon after the accident: "Gordon has apologised to Mrs. Duffy personally by phone. He does not think that she is bigoted". Other members of his entourage stressed that "he was just letting off steam and did not believe Mrs. Duffy was a bigot". And Brown himself claimed that he had mistakenly believed she was bigoted

[33] In chapter V, I allow that Bigotgate may count as a case of accountability *from Mrs. Duffy's perspective*.

because he had failed to understand some of the things she said. Remarks of this sort suggest that what is problematic, *i.e.* what (according to some people) he is criticisable for, is not that he uttered a sentence but that he *believed* it in the first place.

But it's not clear how the advocate of the accountability view could explain why Mr. Brown is open to criticism *for his belief* itself. For it's not obvious that he violated any norm of believing in forming the conviction that Mrs. Duffy is bigoted. Mr. Brown does debate for quite a long time with her, they exchange opinions and argue for their own points of view. If after *that* he sincerely finds her to be narrow-minded, intolerant and opinionated, and forms the belief that she is bigoted *on that basis*, and if he is capable of supporting his belief with reasons that bear on the truth of the content of that belief alone, then it's difficult to say in what sense he would have violated a norm of believing. He might have been a bit quick in his judgement but not irrational to the point of being so heavily criticised.

Someone might deny that he complied with norms of believing on the basis that, as his spokesman says, "he was just letting off steam". What Mr. Brown thought in the car was not a belief to which, upon reflection, he would commit himself but just an expression of temporary exasperation. As such, he did not comply with the norms of believing in forming it. But if it was just an *expression* of exasperation, then it's hard to say in what sense he was *in control* of it.[34] And if the control requirement imposed by the accountability view cannot be met, then it seems that Mr. Brown cannot be held accountable for it. So, either his belief was the result of the correct application of norms of believing or it was not in his control. And either way, the accountability view seems to encounter some difficulty in explaining why he is open to the kind of criticism he is in fact open to.

3.2. Beliefs and values

On the accountability view, we look at what someone does or believes; we evaluate the action or belief itself. If it's not as it should be, we look for someone who's in the position to modify it—where 'being in the position' is intentionally ambiguous between 'being able' and 'being in charge'. As soon as we find them, if they cannot be excused, we hold them responsible and demand that they make up for it. Along these lines, the discussion on responsibility becomes a research for the guidelines that guarantee that we find the right person on which to impose our demands.[35] It works backwards from the object to the author, so to speak.

[34] I will come back to the various possible ways in which someone can be thought to control their beliefs in chapter IV.

[35] *e.g.* the control requirement.

I want to offer an alternative explanation. On the aretaic model, we criticise *someone* for something that they do or believe. There's nothing like a preliminary evaluation of an action or belief that we need to trace back to their authors. From the very beginning, the evaluation is of the person. Their actions and attitudes are expressions of who they are rather than products that they bring about. We do not *infer* something about the agent from their actions and attitudes; rather, we get to know the person by looking at what they do and believe, and we assess *them*. Along these lines, we don't judge the belief that Mrs. Duffy is a bigoted woman as one that is bad having and then verify whether Mr. Brown meet all the relevant conditions; we evaluate what kind of person Mr. Brown is in light of what he believes.

To see the point, consider the comments made by the Shadow Chancellor George Osborne on the BBC: "We have found out the Prime Minister's internal thoughts and I think they speak for themselves … The thing about General Elections is that they reveal the truth about people. … What people will see is the contrast between what he was saying publicly and what he was saying privately". Mr. Osborne's remarks validate the idea that we don't evaluate negatively Mr. Brown's *belief* as much as we criticise *him* for what he believes. His "internal thoughts" are indicative of who he is. They tell us something important about him. Knowing someone's "internal thoughts" even equates to knowing "the truth" about him.[36]

Mr. Brown is criticisable for being a person that believes that Mrs. Duffy is bigoted on the basis of what she says. For instance, Mr. Brown must have taken Mrs. Duffy's question "But all these eastern Europeans that are coming in, where are they flocking from?" to indicate bigotry on her part. For him, the fact that she asks that question is a good reason to believe that she is bigoted. And the fact that he takes that question to support the truth of that belief tells us something important about *his* way of seeing things; about his evaluative orientation.

For some, the fact that he holds that belief shows that he is arrogant and judgemental. Mr. Brown takes himself to be in the position to judge others uncharitably; he thinks too highly of himself. Some might deem him unkind. He does not give the benefit of the doubt to others; he does not focus on the good in people but rather hastens to highlight their defects. He is opinionated, even patronising. He does not have time to spend in such 'ridiculous' matters, as he calls them. Others still may agree with Mr. Brown and think that his belief is not unfair at all. It expresses that he values liberal and progressive ideals, a good education, open-mindedness.

Crucially, the fact that Mr. Brown takes Mrs. Duffy's question to support a charge of bigotry can tell us something about his evaluative orientation precisely because it is not obvious

[36] Moreover, the recorded conversation doesn't mirror Mr. Brown's public statements, and the mismatch between the two proves his insincerity, dishonesty, or untrustworthiness.

that *everybody* would consider it that way. If you yourself think that economic immigration should be more regulated and controlled, for instance, then Duffy's question would support a different sort of assessment. She would not be bigoted to you but rather reasonable, right-thinking, sensible, clear-headed. By contrast, if you value free circulation and consider Europe as a unified nation, you will baulk at the suggestion that the UK shouldn't let immigrants from East Europe enter the country freely. What is for you a good reason in support of the truth of *p* will vary according to what you value.

To put it in more general terms, my claim is that not only practical but also *theoretical* reasons can express someone's evaluative orientation. This might sound puzzling, since practical reasons are typically thought to be connected to value in a way in which theoretical reasons cannot be. The connection between practical reasons and value is generally understood as follows: practical reasons derive from the value of the action for which they are reasons. A reason to φ shows φ-ing worth doing. Or at least, it shows φ-ing worth doing *for those* who find that reason compelling. Thus, we can know what someone finds of value by looking at what they consider to be conclusive practical reasons. For instance, by looking at what I do intentionally or, to be safer, at my practical reasoning, you can get to know that I value friendship, or discipline, or honesty.

By contrast, epistemic reasons cannot derive from the value of having the belief they support.[37] No standard reason to believe that *p* shows *having the belief* that *p worth* believing. I cannot believe that *p* on the basis that it would be valuable for me to be in that state; that would be irrational.[38] Rather, standard reasons to believe that *p* are truth-related; they count in favour of the truth of *p*. Only these can warrant a belief. Thus, theoretical reasons cannot indicate what beliefs someone finds of value in the same sense as practical reasons can indicate what situations someone finds worth bringing about. When I believe that *p*, it's because *p* is the case (at least for me). It's often said that I am *saddled* with the belief: it's not up to me what to believe. I cannot choose to believe a belief over other possible ones. *And therefore*, or so it seems, what I believe cannot display what *I* value. If so, then the claim that Mr. Brown's reasons to believe that *p* can show something of his *evaluative* orientation sounds puzzling.

---

[37] For an initial characterization of the distinction between object-given and state-given (or right-kind and wrong-kind) reasons, see Gertken and Kiesewetter (2017) and Schmidt (2020). For a detailed explanation of the wrongness of those reasons see esp. Hieronymi (2005). For some doubts about whether this distinction is applicable to intentions, see Heuer (2018).

[38] For an overview on evidentialism, *i.e.* the view that nothing besides evidential considerations—facts indicating the truth of a proposition—can be reasons for beliefs, see Dougherty (2011). For arguments in its support, see, amongst others, Adler (2002), Feldman and Conee (1985), Feldman (1988; 2000; 2005), Hieronymi (2008), Kelly (2003), Shah (2006), Velleman (2000), Way (2016), Wedgwood (2002), Williams (1970). By contrast, pragmatists argue for the view that also practical considerations (moral or prudential) can count in favour of holding a belief. They include James (1897), Leary (2017), McCormick (2015), Reisner (2009; 2018), Rinard (2015a; 2015b; 2017; 2019). For a survey of the debate, see Reisner and Steglich-Petersen (2011). For an argument that we have *practical* reasons for thinking that only epistemic reasons can license belief see Whiting (2014).

I think that the puzzlement is due to the ambiguity of 'value' and to an overly narrow focus on certain aspects of believing. In light of the reading of 'values' I suggested in 2.2, I want to say here that what I value is not displayed exclusively by what I find 'worth doing' or 'worth believing'. This is how *practical* reasons are connected to value; they show the thing for which they are reasons to be valuable. But this is not necessarily the only way in which reasons are connected to value. *Truth-related* reasons can be connected to value in other ways.

To introduce the point, consider this passage by Joseph Raz:

> The hint is that the primary significance of reasons, for emotions, beliefs, actions, or whatever else, is to make certain responses eligible, appropriate. Sometimes the presence or absence of additional facts in any particular situation may make the response non-optional, may make one response the required response. That is more likely in the case of reasons for belief and for action than of reasons for the emotions. But on many occasions, I am inclined to say in normal circumstances, there would be more than one response supported by reasons, with none of them supported to a higher degree than any of the others. Needless to say, in all situations our actual response is more definite than that. That is it goes beyond what reasons require of us. And that is the key to the understanding of the limits of reasons, and of the richness of the sources of our responses to the world, which include very much more than our rational capacities. (Raz 2011, 5)

The key thought is that, in normal circumstances, there's more than one (warranted, rational, justified) belief that we can form. There are many ways in which we can avoid being irrational and many possible truth-related reasons that we can consider when we look at the world. And *which belief* we form, and for what reason, can tell a lot about us, and it matters enormously when we evaluate each other.

More precisely, I think that there are three ways in which truth-related reasons can be connected to one's values. First, often we disagree with each other even though "none of [our beliefs is] supported to a higher degree than any of the others". We look at the very same body of evidence and we form different beliefs. And usually, the very way in which disagree says something about our evaluative orientation.[39] For instance, suppose that we sit next to each other on a fairly

[39] The view that, at least sometimes, it's possible for two persons to rationally adopt different doxastic attitudes towards a proposition given the same body of evidence is commonly labelled *epistemic permissivism*. As Rosen puts it, "It should be obvious that reasonable people can disagree, even when confronted with a single body of evidence. When a jury or a court is divided in a difficult case, the mere fact of disagreement does not mean that someone is being unreasonable. Paleontologists disagree about what killed the dinosaurs. And while it is possible that most of the parties to this dispute are irrational, this need not be the case. To the contrary, it would appear to be a fact of epistemic life that a careful review of the evidence does not guarantee consensus, even among thoughtful and otherwise rational investigators" (Rosen 2001, 71–72). For more arguments in support of epistemic permissivism see Schoenfield (2014; 2019), Brueckner and Bundy (2012), Douven (2009). However, not everyone agrees. Feldman argues for the *uniqueness thesis*,

long train journey. The man sitting opposite us wears an expensive suit and is visibly nervous. I form the belief that he is a businessman who's stressed about some money-related affairs. You form the belief that he's on his way for an important meeting. Our beliefs are not guesses; we have our reasons to believe as we do, though they might not be conclusive. In fact, neither belief is supported to a higher degree than the other. We simply read this stranger's behaviour differently. The fact that we do, and the precise way in which we do, says a lot about who we are; about our background beliefs, values and sensitivity to reasons. That I am prone to rationalise his behaviour by reference to money-related affairs indicates that that is the most plausible explanation I can give, or the first plausible one that comes to mind. It might indicate that I myself am anxious about financial volatility, or that I pay more attention to having a job than to interpersonal relationships. By contrast, that you think that the stranger is nervous about meeting someone might indicate that you tend to give importance to interpersonal relationships. You pay less attention to his expensive suit and more to how you think he feels, to his expectations and anxieties.

Sometimes, especially when it comes to *evaluative* beliefs, our disagreement can be highly indicative of the quality of our evaluative orientations. For instance, imagine that Mr. X is the President of our country. On the basis that he stands for certain policies, I might form the belief that he is a great President whilst you might think that he is a terrible one. And depending on the content of our belief, we will display different values. Suppose that I believe that he is a great President because he is against immigration, abortion and divorce, whilst you find all these positions anachronistic and unjust. In this case, you would deem me unjust, privileged, insensitive; someone who values more their own well-being than equal rights for everyone. I am close-minded and value traditions, and I don't trust those who talk of 'progress'. Viceversa, I would deem you radical and a hippie, a libertarian who doesn't respect their own cultural roots. But suppose that I like Mr. X because he supports free health care, subsidies to those who live in a state of poverty and aids to those who lost their jobs, whilst you think that these policies are the wrong ones to adopt. Here, you might deem me preposterous, naive, hypocritical, whilst I might take you to be someone who values money more than justice.[40]

Along these lines, depending on what's the subject matter, disagreements will be more or less indicative of the nature of our evaluative orientation. That I form a certain belief and you form another one, on the basis of the same considerations, shows the distinctiveness of our ways of

according to which a body of evidence cannot support contradictory propositions; reasonable disagreement between two epistemic peers who have shared all their evidence is impossible (Feldman 2006). For arguments in support to the uniqueness thesis see Matheson (2011), White (2005).
Importantly, notice that scepticism about reasonable disagreement seems to be never concerned with *evaluative* beliefs.
[40] Strikingly, *if we agreed*, regardless of what policies Mr. X supports, we would commend each other for our beliefs and think that we both are just, reasonable, uncorrupted, full of integrity.

seeing things. But importantly, the reasons that we take to support our beliefs are all truth-related. Neither of us forms their belief on the basis that it would be a good belief to form. We form them because we take some considerations to count against or in favour of some truth.

Second, we might hold the *very same* belief but for different (truth-related) reasons. Once again, *what* considerations we take to be conclusive is revelatory of who we are and what we value. For instance, we might both believe that the man sitting in front of us is married. However, you think that he's married because he wears a wedding ring, whilst I think that he's married because he carries a packed lunch and his shirt is accurately, but not professionally, ironed. Even though we hold the same belief that *p*, *only I* am open to certain kinds of evaluations for believing that *p*. *My* believing as I do displays—at best—that I am old-fashioned, whilst yours doesn't. And this is because of the reasons we take to support the truth of *p*; only mine displays that I am old-fashioned.[41]

Importantly, what my belief shows does not depend on its truth or falsity. I'd be an old-fashioned person for believing that he's married on the basis that he wears a well-ironed shirt both if he were indeed married and if he were not. Even more interestingly, I'd be an old-fashioned person even if it were true that his shirt had been ironed and his lunch prepared by his wife, and I believed by taking those as reasons. For my belief to display that I am old-fashioned, and for me to be open to positive or negative evaluations for that, it need not be irrational, unsupported, false or unjustified. And I don't need to have violated any norms in believing as I do.

In some cases, not only do we hold the same belief for different reasons but also one of us is *more convinced* of its truth than the other one. For instance, it would be easier for one of us to revise it once faced with some counterevidence. Thus, also *how convinced* we are by our reasons can display the nature of our evaluative orientation; it can display the extent to which we value something. For instance, recall UNWITTING BELIEVER and WITTING BELIEVER from chapter I. Suppose that we both agree that Mr. X is a great President. However, I do so because I've heard many people say that his policies are helping the economy; whilst you do so because you read Mr. X's declarations daily and follow all his political moves. Again, suppose that our belief is true. During his Mr. X's Presidency, and due to the policies he advocates, the economy is booming, and people are generally content. If someone asked us why we believe Mr. X is a great President, I would repeat what I heard people mentioning, whilst you'd give a pondered account of what you've read and reflected upon. My reasons are second-hand, whilst you are profoundly convinced by yours. I think that in cases like this your belief displays your values more than mine does. Not only

[41] In effect, it could also be denied that we hold the *same* belief after all, precisely because we hold it for different reasons.

the fact that you believe that *p* but also *on what* reasons you believe it and *to what extent* you are convinced by their strength show that you are really convinced that those policies are the right ones to stands for. You are someone who values certain policies over other ones, and this can tell us a lot about you. By contrast, my belief cannot display my political views as much as your belief displays yours. Insofar as I believe it only upon hearsay, it doesn't show much about *my* values. If I were presented with further evidence, or if I reflected more on the issue, I might modify my opinion.

It seems to me that the accountability view can capture the difference between UNWITTING and WITTING BELIEVER only by reference to norms of belief-formation. I am criticisable for my belief in ways in which you are not because I violated some norms of belief-formation with which you properly complied. Yet, the fact that I relied on testimony rather than investigating things myself doesn't prove that I have been a 'sloppy or reckless' believer. Obviously, relying on testimony doesn't count as a violation of any epistemic norm in and by itself. Furthermore, even if we agreed that I should not form beliefs *about politics* through testimony, an explanation of the criticism to which I am open that refers exclusively to the fact that I violated some norm of belief formation would deliver quite a shallow explanation of the criticism at stake. The believer of UNWITTING BELIEVER is not criticised simply for forming a belief upon inconclusive evidence. Rather, *if* evaluated negatively, they would be criticised for not caring about politics, or for being lazy, or naïve.

Even more importantly, it's false that UNWITTING BELIEVER is always open to criticism, because of their carelessness, whilst WITTING BELIEVER is always open to praise, given that they put more effort in forming their belief accurately. For instance, suppose that Mr. X were in fact an awful President, who enhances inequality and discrimination. Here, the fact that I am fully convinced by his views would open me to *harsher* criticism, *especially because* I have investigated in depth into his policies. I am not open to criticism because I follow others' opinions but because I am someone who doesn't disvalue inequality and discrimination, and rather take them to be *good* ends to pursue. The fact that I am thoroughly convinced by my position can tell others more about me than your superficial belief can tell about you. But sometimes what my convictions tell us is far from being commendable.

Along these lines, both what beliefs I take to be supported by some (truth-related) reasons and what (truth-related) reasons I take to support conclusively that *p* is the case can display my evaluative orientation. I think that there's also a third way in which truth-related reasons are linked to values, *i.e. for what propositions* we think that there are conclusive reasons. That is, our evaluative orientation is displayed not only by how we settle the question '*Is p true*?' but also by how we answer

the question '*What is true*?'. Our values are displayed by what we think there are good reasons to believe. For instance, imagine once again that we sit next to each other on a train. I believe that the man sitting on the corner is married, that the old woman to the left is a widow and that someone else is a doctor. You believe that trains are not regularly cleaned, that the man sitting on the corner looks suspicious and maybe doesn't have a ticket, and that it might start to rain soon. Suppose that all our beliefs are true. Also, that we hold different beliefs doesn't necessarily mean that we would reject the other's ones, if asked. I might draw your attention to the man's wedding ring, and you might agree with my suppositions.

The point is that the fact that *I* formed those beliefs and *you* didn't—regardless of what beliefs we *could have* formed—is indicative of who we are; it says something about what features of situations each of us finds salient. The fact that I hold beliefs about *people* might indicate that I am interested in others more than you are; because of this, I might be deemed curious, attentive, observant. Yet, the fact that I focus on their marital or social status might show that I am old-fashioned or prude, or at the very least nosy. By contrast, your beliefs all betray that you tend to see the glass half empty. You are a complainer, a pessimist. We can think that I value marriage or a secure job; or maybe that I am sensitive to related factors because I despise both. And we can say that you value cleanness, safety and transparency, and because of this you have a lower threshold of what counts as 'suspicious and fishy'. Because you value safety highly, you are very sensitive to what you think risks endangering it.

In all these cases, we are talking about standard truth-related reasons.[42] And in all these cases, the differences between the reasons to which people are sensitive cannot be organised merely in terms of how strongly they support a certain belief. The *quality* of the reasons to which we are sensitive matters too, besides the *degree* to which they justify beliefs.

4. *The one-instance worry*

In sections 1-3, I have attempted to show that the project of individuating the centre of responsibility in someone's evaluative orientation is worth developing. A key move of my proposal is to shift the focus from local assessments of beliefs to broader assessments of persons. I want to conclude this chapter by clarifying what I mean by 'assessing *persons*'. I do so by rebutting one possible objection to my view.

---

[42] I might have practical reasons *to carry out an inquiry* about whether *p*. And these too can say something of my evaluative orientation, as do my practical reasons to carry out intentional actions, in general. In this case, they show that I find carrying out an inquiry about *p* worth doing. They display my values in this way. But importantly, that I find this inquiry worth carrying out involves the *belief* that it's important to believe something truly about *p*. And *this belief* will be supported by theoretical, truth-related reasons, which in turn show something of my evaluative orientation, in a different way than practical reasons do.

In section 2.2, I argued that, if you have a value, you will see situations in certain ways and find certain reasons compelling. Also, I claimed that you will not find certain things to be of value unless you have the corresponding value. Furthermore, our values are fairly stable: if you have one, you will tend to be sensitive to reasons that express it in different situations. This is part of what guarantees that each of us is *a* person. And yet, none of us is the perfect bearer of a certain value. I can be an honest person who values sincerity and yet find myself lying sometimes. Moreover, in discussing the assessments to which we are open on the basis of our beliefs about the President or about strangers on the train, in section 3.2, I used the conditional: I said that I *might* be deemed nosy; you *might* be deemed a pessimist. The worry then is: how can we know that I *am* in fact *a nosy person*? How can we actually shift the focus from the belief to the person? For we are seldomly in the position to assess someone's *whole* set of beliefs. Typically, we look at just a few of them at a time. And even though a belief of mine usually displays my evaluative orientation, it seems that you can never be sure whether *this* one is a paradigmatic instance of my evaluative orientation or an exception to it. As McCormick and Schleifer put it, "Can we really even assess whether someone possesses a particular virtue based on one instance? It seems not, but we can still blame him in this one instance" (McCormick and Schleifer 2006, 79). Call this the *one-instance worry*.

To anticipate, I think that the one-instance worry lies on a misunderstanding of what it is to *assess persons*. In fact, that we cannot assess someone's evaluative orientation just by looking at few of their beliefs is true if and only if we take someone's evaluative orientation to be constituted by their *full* set of values. Along these lines, the one-instance worry assumes a *monolithic understanding* of 'evaluative orientation', such that, to be assessed fairly as nosy, gentle, kind or polite, one needs to be so in each and every situation. But I am not committed to this assumption. When I stress the importance of shifting from local to global assessments, what I mean is that we need to look at the person rather than at the single attitude. However, this does not mean that we need look at *the whole set of* the person's virtues, vices and values. In effect, I think that I can be a nosy *person* sometimes, without being nosy in all and every situation and without nosiness being 'the most important' feature of who I am. Moreover, I think that it is fair to deem me nosy for my nosy beliefs, if I have some nosy beliefs. I will expand on my response after having presented the worry in more detail.

Before doing so, let me also briefly register that we rarely appraise people in a vacuum, where the *only* thing we know about them is this one belief they hold. More often, we know something about someone; who they are, what they do, what other things they believe. And if we do not, then our ascriptions of responsibility won't have the pretence of certainty. We will be ready to doubt the veracity of our opinions and revise them. And if we do have the pretence of certainty, *e.g.* if I am fully convinced that Angelina Jolie is unkind, then I am being unfair towards her. But

this doesn't seem to be a problem for the view I am proposing as much as a problem for *me* as a member of this society.

4.1. A sceptical worry

There are two ways of understanding the one-instance worry. One is as a sceptical worry. In fact, it shares the same structure of well-known sceptical arguments: we sometimes do and believe things that are not in line with our character. If so, then it will be possible to find consistency of values amongst most of our beliefs, which make up our evaluative orientation, and to find some anomalies in the pattern. And if so, then we cannot safely take any particular belief to display someone's evaluative orientation, for there will always be the risk that we are, in this particular instance, experiencing something exceptional.

Interestingly, the worry does not seem to depend on whether we know the person we are assessing. Suppose that I am considering the belief of someone I don't really know. Suppose that, as a matter of fact, they are a very kind person. They care about others' well-being, do not put themselves first, do not cheat, are gentle. I get to know them during a dinner party and have occasion to notice that, during our exchange, they have a good word for everyone and always think the best of others. However, how can I judge them to be a kind person for what they believe, on the basis of such a limited sample on which to draw? Judging *them* to be kind on the basis of the beliefs they form and hold[43] tonight seems too risky. McCormick and Schleifer consider the converse case. Suppose that I witness someone who's generally very wise and kind whilst, driven by an exceptional outburst of jealousy, they form on insufficient evidence the belief that their wife is unfaithful.[44] McCormick and Schleifer gloss:

> Would it really make sense for us to say "shame on you, you are lacking merit as a person as you clearly [are not wise][45]?" (McCormick and Schleifer 2006, 79)

After all, they *are* in fact a wise and kind person; we are simply seeing them in this particular and infelicitous circumstance.

Things are not smoother when we know the person we assess. Suppose that I know well the wise and kind person who's now driven by jealousy. Situations like this are extremely common.

---

[43] Let's also assume that they are sincere and voice only genuine beliefs.

[44] Brown makes a similar point: "It wouldn't help to appeal to the notion of a bad character trait. For, our judgement that Maud is blameworthy doesn't seem to be a judgement about her character traits. For instance, in judging that Maud is blameworthy, we don't examine whether she is generally disposed to dogmatically ignore the evidence. Indeed, she would be blameworthy even if she is not generally disposed to be dogmatic. So the sense in which she is blameworthy is not that she either has or manifests a bad epistemic character trait" (Brown 2020, 18, fn. 21).

[45] The original quote reads "have not cultivated the virtue of wisdom". However, this puts emphasis on self-cultivation, which is an intentional activity, and therefore raises different issues. The worry I am considering here has to do with the possibility that someone *is not* a wise person if they are *not always* unwise.

We often say and are told things like "I would have never expected this from you!"; "I don't recognise you!"; "Oh, this is so not you!". Thoughts of this kind might even be self-directed: I might find myself having beliefs that are not aligned with who I think I am; that don't display what I take to be my values. These situations also lend themselves to sceptical worries. For even if we (think we) know someone (or ourselves) quite well, we can still doubt whether their (or our) belief displays their (or our) evaluative orientation. That is, we might be tempted to take it as manifesting a long-standing (and so far hidden) tendency. For instance, if I find out that a friend of mine believes that I got a job offer mostly because I am a woman, I might either be surprised because they think something that is alien to who they are or I might start to believe that perhaps I hadn't really understood who they were in the first place. Their belief might be an instance of the bad case or one of the good case: it might be a belief that does not display their evaluative orientation, or one that does display it and helps me rectify my previous understanding of it. But the problem remains, how I can ever know whether *this* particular instance is a good or a bad one.[46]

I suggest a two-step strategy to resist the sceptical worry. First, we should deny the sceptic's understanding of someone's evaluative orientation as a monolithic entity: when we assess someone's belief, we always assess them as a person but this does not mean that we always assess *the totality* of the values they have as a person. That is, even if one single instance may not suffice to give us a *complete* portrait of who someone is and what values they have, it does suffice to tell us *something* about them *as a person*. If you are a kind person and you hold unkind beliefs sometimes, or about certain topics, then I will assess you as being a kind person who's sometimes unkind.

The idea that we can get to know at least *something* of someone by looking at one of their actions or beliefs might sound still too strong. For there might be cases in which an action or a belief would tell us *very little* about them, to the point of seeming irrelevant. In these cases, it seems wrong to overemphasise the importance of what we *should* simply see as an isolated 'bad case'. But the point is that, even if we agree that this one malevolent belief does not display anything of this person after all, still it doesn't justify sceptical worries: sometimes we simply are mistaken in our assessments. This is the second step of the strategy. If we form the wrong impression about someone because we happen to engage with them this *one* time they are unkind, we might need to revise our view of them once we realise that they are not, and be mistaken in the meantime.[47]

---

[46] In fact, the sceptic would not be satisfied with *any number* of beliefs. No set of beliefs or actions, however large, would suffice to give us a sufficiently complete portrait of who someone is and what values they have.

[47] This mistake can take different forms. Two possible ones are these. First, we might fail to understand the role someone plays in society or the relationship we have with them. If I take Angelina Jolie to be unkind because she doesn't believe that my birthday is the 30th of December, I am mistaken in taking her belief to display unkindness. Second, we might fail to understand precisely what value a belief expresses. For instance, we might take someone to be unkind whereas they are simply shy. This is because a belief with a certain content, formed in a certain situation, is not such that it can express uniquely one kind of evaluative orientation.

However, this does not mean that we should *always* refrain from deeming someone unkind because *we might* be wrong. Nor does it mean that we would always be wrong if we assessed them when they are unkind, even if they are typically kind. For they can be kind, and unkind sometimes, or in certain circumstances.

4.2. A worry about fairness

Reference to fairness introduces a second possible way in which my opponent could rephrase their worry. In fact, even if they gave up the sceptical line, they might still insist that shifting the focus from assessments of beliefs to assessments of persons at least generate concerns about fairness. It's *unfair* to assess someone based on one or few of their beliefs.

Again, the crux of the issue seems to be that it would be unfair to appraise me on the basis of a certain belief, if it is a belief that it is surprising for me to hold *given my overall* evaluative orientation. This thought betrays once again a monolithic understanding of 'evaluative orientation'. As soon as we allow for the idea that evaluative orientations are complex and seldomly captured by one-word descriptions, we are not embarrassed by the fact that some beliefs of ours display a virtue, whilst others a lack of it. For instance, I might be very kind in my actions and beliefs and yet become quite aggressive when discussing the behaviour of cyclists on the road. My friends could be surprised at my uncharitable remarks. Now, would this cast doubts on their idea of me as a kind person? Or would it entail that I am not kind after all? It needn't be so. Rather, it would provide a more complex picture of who I am. I might be someone who's a kind person and yet unkind towards cyclists on the road.[48] Also my unkind beliefs display who I am, but who I am is not *uniquely* displayed by them. We just need to allow that I can be a kind and sometimes unkind person; or that I am a kind person but unkind about certain topics.[49]

Thus, it wouldn't necessarily be unfair to deem me unkind for believing unkind things. Rather, it would be unfair to *infer* from my unkind beliefs that I am an unkind person *overall*, and that I must be lying every time I voice one of my kind beliefs. Similarly, it would be unfair to *insist* that I am unkind after having got to know me better, met other people that know me, understood the context. Here, the source of unfairness would be that you don't give my point of view due credit. You would be wronging me in sticking to your judgement. But these cases can be reduced

[48] Obviously, things might well be different if I am kind with everyone except for, say, Asian people. See chapter V for further discussion.

[49] In chapter III, I suggest that not all virtues can be relativised in this way. For instance, if you are dishonest with respect to certain things, you cannot be deemed honest at all.

to the cases of unfairness I discuss in chapter V, *i.e.* unjustified beliefs about you. The source of unfairness would not be the fact that you deem me unkind based only on a few unkind beliefs.[50]

Finally, it would be unfair to deem me unkind for holding a belief if that belief does *not* display my evaluative orientation. For instance, this could be the case if I am numbed by tiredness or stress. Here, it might indeed be unfair to take too seriously something I say to believe. I might be simply parroting someone else. Conditions like tiredness seem to hinder the relevant expressing relationship between my attitudes and my evaluative orientation. They do not 'lower one's filters',[51] so that my beliefs are more sincere manifestations of who I am, but rather *add* disturbance, to the effect that my beliefs are not a manifestation of any of my values at all.

To conclude, the key point is that the one-instance worry does not threaten the project of shifting the focus from assessing someone's beliefs to assessing *them* as a person. Even when we consider only one or few of someone's beliefs, as we often do, we look at the *person* who holds them. To make sense of this, we just need to allow for a more compound understanding of 'persons' and of the nature of our evaluative orientation. This also clarifies that, when I claim that the accountability view risks being overly local, the worry is not that it looks at beliefs one at a time, for I allow that we can do *that*. Rather, the worry is that it looks at them *in isolation*, *i.e.* detached from what they display of their believers. It focuses on the belief without looking at the person. But when we consider one or few of someone's beliefs, we do not simply check whether they are instances of compliances or violations of norms; we look at what those violations *mean*. At what they tell us about the person who forms and holds them.

In this chapter, I have investigated the relationships between who someone is, their values and their actions and beliefs, and I have argued that someone is responsible for their actions and beliefs insofar as they express their evaluative orientation, *i.e.* their distinctive sensitivity to practical and theoretical reasons, oriented by their values.

In chapter III, I address two questions. First, I investigate the *conditions* under which I am open to commendation or criticism for some of the things that display my evaluative orientation (the *conditions question*). Crudely, my suggestion is that we are to commend for the beliefs that express virtues, and to criticise for the beliefs that express vices. Second, I explore the *nature* of those evaluations (the *significance question*). In fact, it is often said that we are *epistemically* responsible for our beliefs, and that our beliefs display our *intellectual* virtues and vices; I am intellectually humble,

[50] Also, ascriptions of responsibility should not be seen as final inescapable verdicts. If someone appraises us wrongly, we are able to resist their judgments, and they are (or should be) ready to modify them. Assessing others' evaluative orientation and having one's own assessed involve a great deal of understanding and negotiation.
[51] Like drunkenness sometimes might, see 2.4 above.

intellectually courageous, intellectually honest, or open-minded. I argue that we have good reasons not to cluster virtues and vices into 'intellectual' and 'moral' ones. Rather, virtues and vices are connected with varied and rich dimensions of assessments, that cannot be reduced to the epistemic and moral realms only.

III. The normativity of aretaic appraisals

In chapter II, I discussed Gary Watson's claim that someone is responsible for an intentional action insofar as it expresses their adopted ends (WT). I articulated WT by developing the notion of 'evaluative orientation' as someone's sensitivity to see certain things as reasons for actions and beliefs, oriented by their values. Along these lines, on the aretaic model, someone is responsible for *both* their attitudes and their actions insofar as they express their evaluative orientation. Responsibility for beliefs ceases to be a secondary phenomenon, derivative on the case of actions.

Once we have determined how to distinguish the things for which we can be deemed responsible from the ones for which we cannot, two important questions arise. First, it has to be explained why we are open to *criticism* for some of them, and to *commendation* for others. That is, we need to understand precisely on what basis certain actions and beliefs open me to positive appraisals, and some to negative ones. Call this the *conditions question*. Second, it has to be clarified *what sort* of criticism and commendation are involved in ascriptions of responsibility; to which dimension of assessment they belong. This question regards the nature of the relevant evaluations. Borrowing the term from Angela Smith, I call this the *significance question*. Typically, the evaluations involved in responsibility for beliefs are thought to be *epistemic* in nature: according to what I believe, I am said to be 'epistemically blameworthy' or 'epistemically praiseworthy'. And yet, as I have shown in chapter I, there are cases in which we seem to deem someone responsible for their belief even though no 'epistemic' breach is detectable (*e.g.* NORPOIS).

I start off by considering what I take to be the standard way of thinking about the conditions and the significance questions, *the simple accountability view*.[1] The simple view is one version of the accountability view insofar as it addresses the conditions question by reference to violation of (and compliance with) norms. Believers are criticisable when they violate norms with which they should and could have complied, and they are commendable when they comply by them even when it is particularly difficult to do so. But the simple view is a *particular* version of the accountability view insofar as it addresses the significance question in one particular way, to which not all supporters of the accountability view are necessarily committed.[2] On the simple view, agents are morally responsible for their actions, whilst believers are epistemically responsible for their beliefs. 'The moral' and 'the epistemic' are treated as separate, largely independent domains.

[1] Hereafter, only 'the simple view'.

[2] That is, I say that the simple view is one version of the accountability view insofar as it offers *a* way—and I believe, the most common one—of specifying *what kind of norms* believers violate. However, there is nothing amongst the distinctive commitments of the accountability view (control requirement, reference to norms, local assessments and demands and expectations) that entails that the norms relevant to the belief case are epistemic norms.

More precisely, the simple view consists of two distinguishable theses. One is that there are two distinct domains that are governed by different sets of norms, *i.e.* the epistemic and the moral; the second is the idea that someone's beliefs belong to the epistemic, and are to be assessed against the corresponding standards, whilst someone's actions belong to the moral, and are to be assessed correspondingly. Call these theses DISTINCTION and SYMMETRY respectively.[3]

I resist the simple view on the basis that it is overly simplistic and compartmentalised, and, as such, unable to accommodate several central instances of responsibility for beliefs. In the second part of the chapter, I develop my proposal. Briefly, the aretaic model addresses the conditions and the significance questions by articulating the notion of 'evaluative orientation' in terms of virtues and vices, which are distinctive patterns of being sensitive to both practical and theoretical reasons. First, someone is open to commendation if their belief expresses a virtuous evaluative orientation and to criticism if it expresses a vicious one. Second, virtues and vices open up varied and diverse dimensions of assessment that cannot be reduced to the moral and the epistemic dimensions only. Along these lines, I reject both the dichotomy between the moral and the epistemic implicit in the simple view and its reduction of *doxastic* responsibility (responsibility for beliefs) onto *epistemic* responsibility (responsibility from an epistemic point of view).

More precisely, in section 1, I investigate what is commonly meant by 'the moral' and 'the epistemic'. I follow T. M. Scanlon and Pamela Hieronymi, amongst others, in claiming that the significance of 'the moral' lies in the importance we place on the relationships of mutual regard. 'Significance' here is used as a technical term, as introduced by Angela Smith. I then consider several possible ways of pinning down the significance of 'the epistemic'. One natural option is that abidance by epistemic norms matters because we value truth or knowledge themselves. In section 2, I raise some questions for the simple view so understood. I try to accommodate them on its behalf, by specifying the significance of the epistemic in ways other than by drawing on the importance of truth or knowledge. However, none seems satisfactory, and the explanatory power of the simple view is shown limited.

In section 3, I introduce a pluralist virtue-centred alternative and elaborate on how it can account for the cases I discuss in section 2. I then address the question how we should reassess

[3] Although I take the simple view to identify a widespread tendency, it's worth acknowledging that many contemporary philosophers would distance themselves from it, especially on the basis that we can be also morally responsible for our beliefs. For a recent overview on pragmatic encroachment, see Kim and McGrath (2018). For references on pragmatism about reasons for beliefs, see fn 38, ch. II.
However, that some argue that we can also be morally responsible for our beliefs doesn't undermine the value of a discussion of the simple view. First, pragmatic encroachment is far from being universally accepted. Second, it widens the jurisdiction of the moral domain, but it doesn't allow, for example, for the possibility that we are epistemically responsible for our actions. Third, importantly, also pragmatists about reasons for beliefs accept the distinction between the moral and the epistemic domains. So even though they challenge SYMMETRY, they grant (and build on) DISTINCTION.

DISTINCTION and SYMMETRY from within an aretaic perspective. I claim that, first, a shift to virtues and vices involves a rejection of SYMMETRY, since each virtue and vice is manifested in someone's sensitivity to reasons for *both* actions and beliefs (section 4). Second, even though DISTINCTION *could* be reformulated within a virtue-centred framework, *e.g.* by drawing a distinction between moral and intellectual virtues, I argue that there seems to be no good reasons to reduce the diverse and varied dimensions of assessment that correspond to virtues and vices to the epistemic and moral domains alone (section 5). I conclude by clarifying my stance on the notions of 'epistemic' and 'moral' (section 6).

1. The simple view

To introduce the simple view, I consider DISTINCTION and expand on the meaning of 'the moral' and 'the epistemic'.

1.1. The 'moral'

We generally share strong intuitions as to what morality is concerned with. For example, we know that it is not concerned with how many walks I take every week, even though I do so intentionally, but rather with whether I greet my neighbours, lie, take care of my elderly parents, recycle. The latter actions, but not the former, have to do with what is *good* to do from a moral point of view.

To set up an inquiry into the meaning of 'moral goodness', Angela Smith suggests distinguishing 'depth' from 'significance' (Smith 2008, 384–87). Crudely, a reference to 'depth' draws the distinction between responsibility-involving evaluations, which are deep insofar as they meet certain conditions,[4] and mere grading evaluations, which simply rate someone's performance against some standards. By contrast, 'significance' is concerned with the *kind of importance* we attach to different sorts of responsibility-involving evaluations. Sometimes deep evaluations have a moral significance; some other times an epistemic significance; and so forth. In these terms, we can say that this chapter focuses on the *significance* of deep—responsibility-involving—evaluations of believers.

Of course, giving a simple and universally accepted account of the significance of the moral lies far beyond the scope of this thesis. Here, I follow a broadly[5] Scanlonian framework according

[4] Notice that Smith, however, expands on both 'depth' and 'significance' by reference to norms. Criticism or commendation depends on compliance with some relevant norms, and significance "depends on the particular standards upon which the demands for reasons are based" and "concerns the kind of importance that attaches to his failing to meet the normative standards that apply in the particular case" (Smith 2008, 385).

[5] 'Broadly' because I take from Scanlon the general idea that morality is essentially interwoven with interpersonality, though this doesn't commit me to the details of his contractualist proposal, *e.g.* the idea that "an act is wrong if and only if any principle that permitted it would be one that could reasonably be rejected by people with the motivation

to which the subject matter of 'morality', in a narrow sense, has to do with what we owe to each other.

> …the relation with others the value and appeal of which underlies our reasons to do what morality requires…might be called a relation of mutual recognition. Standing in this relation to others is appealing in itself—worth seeking for its own sake. A moral person will refrain from lying to others, cheating, harming, or exploiting them, "because these things are wrong". But for such a person these requirements are not just formal imperatives; they are aspects of the positive value of a way of living with others. (Scanlon 1998, 162)

When we act morally, our practical reasons derive their normative force from the value of the relationship of mutual recognition. We do what morality requires insofar as that allows us to instantiate the positive value "of a way of living with others". Along similar lines, Pamela Hieronymi individuates the significance of the moral in 'the relationships of mutual regard':

> to criticize [a thing] is to say it fell short of some standard. But…I don't resent you because you failed to achieve some standard of human excellence or failed to live up to the norms internal to willing. Rather, I resent you because your action or attitude communicated disrespect or disregard for me. Likewise…I am grateful because you chose to benefit me in a way that you needn't have. I am more concerned with myself, or with the relation in which we stand, or with your concern for me, or with how I figure into your world, than I am concerned with your performance, as such—with how you fared against a standard of moral excellence or rectitude or goodness. (Hieronymi 2004, 124)

She continues:

> The force of a judgment of disregard is not adequately captured simply by the thought that some especially important or deep capacity of mine has failed, or even that I have failed in its exercise. Rather, the significance of moral failure, I suggest, is to be found in importance of standing in certain sort of relationships with others: call them relationships of mutual regard. (*ibid.*)

The key point is that we don't criticise someone morally insofar as they fail to approximate to an ideal behaviour but rather because they fail to show regard for *us*, or at least a sufficient degree of respect for our well-being. This line has obvious connections with the Strawsonian tradition. There is an inversely proportional relationship between my mild annoyance at a seriously aching toe, following a clumsy stumbling, and the burning resentment I feel for a mild pain inflicted

---

just described (or, equivalently, if and only if it would be disallowed by any principle that such people could not reasonably reject)" (Scanlon 1998, 4).

intentionally. This is ultimately explained by the fact that stumbling expresses a poor performance but no disregard, whilst an intentional trip shows no failure in performance but serious disrespect of mutual regard.

Along these lines, the relationship between 'depth' and 'significance' is refined: the importance we place on mutual regard is not only what distinguishes the moral from the epistemic domain. Also, it is part of what explains why the corresponding evaluations are *responsibility*-involving, rather than mere grading evaluations. It is because my action can be good or bad in a certain *moral* way, which is not the way in which my hair can be good, that I can be responsible for it. If we did not value mutual regard as we do, not only would it be difficult to isolate something as *morally* relevant but also to understand moral *responsibility* altogether. It would lose its force. If so, then an investigation of the significance of a particular normative domain is not to be carried out after we decide on the conditions of responsibility but rather it must be integral part of an exploration of those conditions themselves.

Finally, notice that Hieronymi is referring to *a certain sort* of relationships with others. In effect, 'relationships of mutual regard' in general underlie more than only moral actions. For instance, passing the ball to members of your team, or giving a client what they ordered, might be seen as actions that are essentially interpersonal and that require a certain mutual regard. What is distinctive of the moral dimension is that it has to do with a *certain* (moral, indeed) kind of interpersonal regard. One way to unpack this thought is to say that, in order to be morally good, an action has to demonstrate that the agent *values* the relationships of mutual regard; others' well-being has to figure somehow amongst the agent's reasons to act as they do. Even though also winning a match or getting the food I ordered would somehow contribute to my well-being, your passing me the ball or giving me my order are not morally good actions because you do not act in those ways *because* I will benefit from them. You might be glad that what you do causes someone's pleasure rather than pain, but that is not your reason to do so.[6]

[6] Two objections might be raised. The first one is that, at least in some cases, we would consider an action to be moral even though the agent did not act out of regard for others, but rather because they ought to do so, where the duty is moral rather than legal. However, the force of the moral ought is itself grounded in the relevant kind of regard: if I ought to tell you the truth because it is morally good, that is precisely because telling the truth shows respect for your right not to be deceived. That sincerity is a value explains why we consider being truthful a moral demand; and sincerity is a value because we stand in certain relationships, *i.e.* because we regard each other's right to know.
Another challenge is this. Someone might say that articulating the relevant kind of regard in terms of reasons for which someone acts works better in the positive than in the negative case: it is not obvious that when I commit an immoral action I take someone's *ill-being* as a reason. I might simply be negligent. Here, I agree that immoral actions need not necessarily take someone else's ill-being as a practical reason. However, my proposal is only committed to the idea that immoral actions do *not* take someone's well-being in consideration. There need not be trace of someone's positive tendency to ill-treat others; it suffices to find tracks of any considerable lack of consideration for their well-being. Of course, what counts as a 'considerable' lack of consideration remains an open question, and a more precise understanding of negligence depends on its answer.

1.2. The 'epistemic'

Philosophers tend to agree that the epistemic domain bears certain structural similarities with the moral one. Typically, both are taken to be governed by normative requirements, though of a different kind. As Smith puts it:

> Failures to respond to the normative requirements in other domains (e.g., aesthetic or epistemic) will have different kinds of significance, depending upon our reasons for caring about the demands in question. (A. M. Smith 2008, 385)

Several others agree. For instance, consider McHugh and Hieronymi:

> Only when we are responsible for something in this way can we legitimately be held to demands or requirements with respect to it. The demands to which we are subject take various forms. For instance, our actions are subject to *moral* demands; we are *morally responsible* for them. The primary sort of demand to which our beliefs are legitimately held are *epistemic* demands. For example, there is a demand on you not to believe something just because your horoscope said it. We are *epistemically responsible* for our beliefs. (McHugh 2013, 133)

> On such an account, to be epistemically responsible for a thing would be to be open to epistemic assessment or judgment on account of that thing, to be epistemically praise- or blameworthy for it, to be open to certain reactions from others on account of it, and perhaps to be open to certain sanctions from one's epistemic community. The difference between "moral" and "epistemic" responsibility, on such an account, would simply be which standards of assessment are applied, and, correspondingly, which reactions are warranted. (Hieronymi 2008, 363)

In both the moral and the epistemic contexts, someone is responsible for something against certain demands and standards of assessment, the nature of which determines the kind of responsibility at issue. The nature of epistemic responsibility depends on what counts as an epistemic norm, which yields an understanding of what counts as a blameworthy violation of it and as praiseworthy adherence to it.

Yet, there is no agreement on the nature of epistemic norms, *i.e.* of the norms that govern beliefs from a distinctively epistemic truth-related perspective, as opposed to a moral or prudential one. The disagreement concerns two related but distinct questions. The first one regards the *content*[7] of epistemic norms, which would identify the boundaries of the epistemic realm. The second one regards the *significance* of the epistemic domain, *i.e.* what are "our reasons for caring about the

[7] And *form*, *e.g.* whether they are categorical or hypothetical imperatives.

demands in question". Or put otherwise, what is the distinctively epistemic value that we instantiate in recognising and following good epistemic reasons.[8]

The two questions are tightly related, since it's quite natural to identify the content of epistemic norms according to what one takes to be the nature of epistemic value. In effect, this is what we seem to be doing in the moral case too: the fact that someone needs help, provided that I can easily assist them, is a good reason for me to do it *from a moral* point of view *because* of the value we place on the relationships of mutual regard. Yet, I'll attempt to maintain the answers to the two questions as separate as possible.

To begin with, we can say that the epistemic realm is concerned with norms that specify the conditions under which we should believe things from an epistemic point of view. Given that beliefs are correct if and only if the proposition believed is true—or, as it's often said, given that 'beliefs aim at truth'[9]—, epistemic norms must identify a certain connection with truth. According to a deep-rooted tradition,[10] the most basic epistemic norm is that we are under the obligation to know the truth and avoid falsehood.[11] For instance, William Alston writes:

> What distinguishes epistemic principles from moral principles? Well, the "epistemic point of view" is characterized by a concern with the twin goals of believing the true and not believing the false. … epistemic principles for the assessment of belief will grade them in the light of these goals. (Alston 1988, 259)

[8] A third related but distinct question regards the *source* of epistemic normativity, *i.e.* on virtue of what reasons for beliefs are normative reasons. One common view is *constitutivism* about epistemic normativity. The idea is that epistemic norms have this normative authority because belief constitutively aims at truth. The nature of belief can explain why we have reasons to believe, or disbelieve, propositions in light of evidence possessed for, or against, them. For discussion, see Côté-Bouchard (2016), Cowie and Greenberg (2018), McHugh and Whiting (2014). For the claim that belief has a constitutive norm, see amongst others Adler (2002), Engel (2005), Gibbard (2005), Millar (2004), Shah (2003), Wedgwood (2002), Whiting (2012).
Côté-Bouchard, contra constitutivism, suggests that "Perhaps…our concept of an epistemic norm is not, after all, that of a norm that necessarily entails good reasons. That is, perhaps epistemic norms have the same kind of normative authority as e.g. norms of etiquette, fashion, games, and the like. If this is right, then there can still be facts about what we epistemically should believe. But just like facts about what we should do according to etiquette, they are not genuinely normative facts" (Côté-Bouchard 2016, 3195–96). On a somehow similar vein, Feldman suggests that epistemic oughts are akin to *role oughts* (Feldman 2000).

[9] This expression was first coined by Bernard Williams (Williams 1970, 137) to identify a set of properties of beliefs: first, that truth and falsehood are dimensions of assessment of beliefs as opposed to other psychological states and dispositions; second, that to believe that *p* is to believe that *p* is true; and third, that to say "I believe that *p*" carries, in general, a claim that *p* is true; that is, it is a qualified way of asserting that *p* is true.

[10] Cf. William James: "There are two ways of looking at our duty in the matter of opinion, — ways entirely different, yet ways about whose difference the theory of knowledge seems hitherto to have shown little concern. We must know the truth; and we must avoid error — these are our first and great commandments as would-be knowers" (James 1897, VII, p. 17). To be more precise, Williams claims that the two laws are separable and that the former is more important than the latter: the possibility of being mistaken should not have us refrain from believing. "In a world where we are so certain to incur them in spite of all our caution, a certain lightness of heart seems healthier than this excessive nervousness on their behalf."

[11] For the view that acquiring *significant truth* while avoiding *significant falsehood* is the only aim of belief, see David (2001).

This view is still widely accepted, though it has been questioned on multiple grounds. Some have argued that epistemic norms don't determine what's *required* of us, but rather what we are *permitted* to believe.[12] Others have insisted that no epistemic norm can impose a duty to believe trivialities, even if they are true.[13] And others have resisted this understanding of epistemic norms as a dyad: sometimes believers should just suspend their judgement.[14]

By complying successfully with epistemic norms, *i.e.* by believing truths, avoiding falsehood, suspending judgement when necessary and so forth, our beliefs are 'epistemically valuable'. Some identify epistemic value with truth: true beliefs are valuable and holding them is valuable.[15] For others, *knowledge*, rather than truth, is the chief epistemic value. We should believe what we know.[16] Yet others argue that we can avoid the problems associated with identifying epistemic value with true belief or knowledge if we maintain that *rational beliefs* are what has epistemic value. "To do well as a believer, to achieve a kind of epistemic excellence, one must form only rational beliefs" (Feldman 2000, 685). Non-factive justification is the condition of epistemic success for belief.[17] Finally, some reject value monism and highlight the importance of other achievements, *e.g.* understanding or wisdom.[18]

This doesn't yet suffice to answer the question about the *significance* of the epistemic domain. In fact, whichever of these views one prefers, *i.e.* whether one thinks that truth, knowledge, reasonableness or understanding identifies the chief epistemic value, it remains unclear what are 'our reasons to care about the demands in question'. That is, it still has to be clarified what can be the analogue of 'the importance of the relationships of mutual regard' for the epistemic domain; what force underpins epistemic evaluations. For flouting epistemic norms, however interpreted, is

---

[12] See especially Schoenfield (2014; 2019), and White (2005) for criticism. *Cf.* Audi (2001, 105): "The results of this paper do not prevent our sustaining … an epistemic deontologism conceived as the view that to be justified in believing that p is for believing p to be (epistemically) *permissible* for one" [my italics].

[13] See Harman's *clutter avoidance principle* (G. Harman 1986). *Cf.* Friedman (2018; 2020), Goldman (1978).

[14] For instance, within an evidentialist framework, Feldman (Feldman 2000) argues that if a person is going to adopt any attitude toward a proposition, then that person ought to believe it if his current evidence supports it, disbelieve it if his current evidence is against it, and suspend judgment about it if his evidence is neutral (or close to neutral).

[15] For the view that truth is the norm of belief (*veritism*) see Ahlstrom-Vij (2013), Gardiner (2012), Gibbard (2005), Pritchard (2011; 2014), Railton (1994), Shah (2003), Velleman (2000), Whiting (2010), Williams (1970).

[16] For the view that knowledge is the norm of belief, see Adler (2002), Bird (2007), Engel (2005), Littlejohn (2013), McHugh (2011a), Sutton (2007), Williamson (2000), and Pritchard (2007) for discussion. Fricker suggests that one fundamental value of knowledge *is* in fact reducible to the value of truth, even though it remains independent from the value of the truth of the constituent true belief. That is, the value of knowledge consists in something about knowledge that helps us retain our true beliefs over time, *i.e.* its superior resilience (Fricker 2009).

[17] See Adler (2002), Feldman and Conee (1985), Feldman (1988; 2000; 2005), Gibbons (2013). For the opposite view that epistemically good beliefs need not be based on sufficient evidence see for instance Brogaard (2014).

[18] For an argument against value monism see Axtell (2008), DePaul (2001) and Kvanvig (2003; 2005), who argues that "epistemic goals include knowledge, understanding, wisdom, rationality, justification, sense-making, and empirically adequate theories ....[T]he class of epistemic goods is manifold, as wide as the class of cognitive successes" (Kvanvig 2005, 287). In effect, according to Kvanvig, understanding has *more* value of both knowledge and truth.

insufficient for a believer to be open to *responsibility-involving* criticism, pending a clarification of why abiding by those norms matters in the first place.

I think that the significance of the epistemic, so understood, will depend on whether we take epistemic value to be final or instrumental. If it's final, then the idea is that we need to abide by epistemic norms because truth or knowledge are valuable in themselves.[19] If it's instrumental, then true beliefs or knowledge are valuable insofar as they are necessary to obtain something else, *e.g.* to succeed in our projects.[20] In section 2, I take into consideration both options and assess the explanatory power of the simple view, understood in the two corresponding ways.

In effect, I elaborate on the first option at greater length (2.1-2.4), since it reflects the most common way of understanding the simple view: SYMMETRY[21] is usually taken to mean that we are epistemically praiseworthy for achieving the epistemic good, and epistemically blameworthy for failing to do so, because knowing the truth is valuable in itself, from an epistemic point of view. We need to abide by epistemic norms because it's good to know the truth, and the kind of goodness involved is epistemic in nature. Call this *the narrow reading* of the significance of the epistemic; someone is epistemically responsible in the *narrow* sense. In 2.5, I explore briefly the second option, *i.e.* whether the simple view can achieve greater explanatory power if we understand epistemic value as instrumental. Here, SYMMETRY states that we are epistemically commendable for achieving the epistemic good insofar as it allows us to achieve some other ends. This option would offer a *wider reading* of the significance of the epistemic. However, it faces some difficulties. First, it struggles with the fact that believing evidentially supported propositions isn't always of practical value. Second, epistemic value risks depending on the moral or prudential value that is attached to the projects we set out to achieve. Finally, even understood along these lines, the simple view is equally unable to explain all the cases that we'd recognise as genuine instances of responsibility for beliefs.

2. The explanatory power of the simple view

In this section, I assess the explanatory merits of the simple view by probing how it can account for some key cases beliefs for which we can be responsible. In 2.1-2.4, I focus on the narrow reading; and I consider the alternative wider reading in 2.5. In both versions, it is unclear whether the simple view can offer a satisfactorily comprehensive account.

---

[19] For criticism see for instance Raz: "It is not the case that there is always (even a pro tanto) value in having a true belief, whatever it is. Nor is it the case that it is always a disvalue to have a false belief. If that is so, and since there can always be reasons for believing at least any true proposition, it follows that reasons for belief are not provided by values in the way that reasons for action are" (Raz 2011, 43). See also Côté-Bouchard (2017) and White (2007).

[20] Foley (1992, ch. 1). For discussion and criticism see Lockard (2013).

[21] Henceforth, I discuss only the part of SYMMETRY that is concerned with epistemic responsibility for beliefs rather than with moral responsibility for actions.

2.1. SHERLOCK, GOOFY

The simple view accommodates some cases quite nicely. For instance, it seems able to explain why we praise characters like Sherlock Holmes (SHERLOCK). Sherlock is good at getting at the truth, and we welcome his achievements with commendation, awe, reverence. He individuates clues, weighs reasons, and gains knowledge even when it is particularly difficult to do so. He can discern between considerations that count in favour of something and those that merely look like they do. Furthermore, he does so consciously, confidently, reliably, quickly, accurately. In *this* sense, Sherlock is good epistemically.

The claim that there is something like 'the epistemic good', and that it should be understood in the narrow way, seems to be supported by the fact that we can praise someone for achieving it independently of their capacities to achieve *e.g.* the *moral* good. In fact, we can praise Sherlock for being extraordinarily smart, rational, reasonable or knowledgeable and yet consistently believe that he is morally deficient in important respects. The compatibility of epistemic praise and moral condemnation—and viceversa—is seen as a proof of the distinction between the two domains and of the legitimacy of each.

Similarly, someone can be criticised for failing to abide by epistemic norms independently of their moral good- or ill-will. For instance, someone might do what they can to abide by moral norms, and yet fail to do so properly because of cognitive limitations. Maybe they're gullible or sloppy; they fail to weigh reasons correctly, give too much importance to some aspects, jump to conclusions. Here, we'd be able to criticise them for failing to achieve the epistemic good, without appraising them morally. And we would criticise them because, in failing to abide by epistemic norms, they fail to achieve something that is valuable in and by itself—something that it would have been important for them to achieve.

For instance, consider evaluations of characters like Goofy (GOOFY). In a memo written in 1934, Disney animator Art Babbitt writes:

> Think of the Goof as a composite of an everlasting optimist, a gullible Good Samaritan, a half-wit, a shiftless, good-natured colored boy and a hick […]. His brain is rather vapoury. He laughs at his own jokes, because he can't understand any others. He is very courteous and apologetic and his *faux pas* embarrass him, but he tries to laugh off his errors. He talks to himself because it is easier for him to know what he is thinking if he hears it first.[22]

[22] From an article on The Independent written by Nick Hasted on Sunday 23 October 2011, https://www.independent.co.uk/arts-entertainment/so-what-goofy-1358726.html.

We can consistently consider Goofy 'half-wit' and certainly not bright; and yet deem him good-natured, and courteous in his actions. And these binary evaluations are possible precisely because, the simple view states, there are two distinct evaluative domains that have different targets. He is not good epistemically, *i.e.* he is not good at achieving the epistemic good.

However, not every time we criticise (or commend) someone for their beliefs, we do so *because* they failed to abide by (or extraordinarily complied with) epistemic norms. Sometimes, no violation of norms can be detected at all (2.2). Some other times, the fact that the belief is false or unjustified is not *what matters* (2.3, 2.4).

2.2. NORPOIS, ABSENCE

We can deem someone responsible for a belief even though no violation of epistemic norms (nor extraordinarily compliance with them) can be detected.

Recall NORPOIS from chapter I, who believes that it would be 'unseemly' to bother the Comte de Paris with certain questions. Mr. Norpois does not violate any epistemic norm in believing as he does. He believes for reasons, and he'd be able to mention considerations that he thinks bear on the truth of his belief. He is not irrational. And yet, someone might criticise him on the basis of what he believes for being so old-fashioned and attached to appearances. They could find him snobbish, even ridiculous. Others might think that he is someone who knows how to play the game, maybe foxy. Or perhaps just simple-minded and polite. But none of these evaluations depends on his belief being true or justified: Mr. Norpois can be old-fashioned or polite irrespectively of the Comte's actual reactions to those 'unseemly' questions. In fact, it might even be *true* that the Comte would be bothered. And lacking a clear violation of a norm, it is difficult to make sense of our evaluations of Mr. Norpois within the accountability view.

Consider another case, ABSENCE. We are colleagues. You haven't come to work today, and you are missing an important meeting, which you told me you were eager to attend. I fail to form the belief that you didn't come to work because I don't notice your absence, though your chair is empty. In fact, your alarm didn't go off, and a simple call from me would have spared you many problems. However, we are not friends; you are always precise and organised; and you've never needed my help before.

It is not obvious that I violate an epistemic norm in this case.[23] I would, if we accepted a very strong evidentialist norm according to which we should form all the beliefs for which we have sufficient evidence. If this were a genuine norm, then I would violate it in ABSENCE. However, it

[23] For one thing, I do not violate James' dictum; it is not that I believe falsehoods, I simply *fail* to notice your absence.

has been highlighted that norms of this kind would be overly demanding. They would impose an incredible cognitive burden or issue an extremely liberal understanding of 'legitimate criticism'. We would be criticisable most of the time for failing to form true and justified beliefs that we could have formed. One suggestion is that I fail to comply with a norm that states that I owe you attention, care and consideration. Norms of this kind might indeed be constitutive of certain kinds of interpersonal relationships,[24] but it seems implausible that they are operative here, since we are not close. And more importantly, a norm of this kind wouldn't be *epistemic* in the narrow sense. So it's hard to see how violating *it* would make me *epistemically* criticisable, within the framework at issue.

And yet, the next day my colleague would still evaluate me negatively for having failed to believe that they were absent. Even if I didn't violate any epistemic norm (and even if I didn't *owe* it to them to think about them),[25] they would think I am self-absorbed and individualistic. I am criticisable for being a kind of person that does not think about others and is not eager to help. I am not particularly generous or kind. Even though my reason-weighing process was not obviously epistemically incorrect, in the narrow sense, I am still open to these kinds of criticism.

2.3. PLANT, MATHS

In some cases, even though we *could* identify an epistemic breach, we would not criticise the believer for *that* as much as for what their epistemic breach tells us about them. It seems to me that there are two main clusters of cases. If the believer holds a belief *about* someone but doesn't care sufficiently about getting things right, then they *wrong* them. These cases belong to the practices of accountability.[26] In all the other cases, we would criticise a believer for a false or unjustified belief but neither because it is false or unjustified, nor because it wrongs someone.

Let me start with the former cluster of cases, where I wrong others by believing. I will be brief, since I discuss the practices of accountability at length in chapter V. We wrong others by believing things about them without caring about getting things right. In doing so, we treat them unfairly. The crucial point for present purposes is that I am open to criticism for *what is displayed* by

[24] I will discuss the case of friendship in chapter V.

[25] This is why ABSENCE is not a case that belongs to the practices of accountability, as I clarify in chapter V.

[26] To anticipate, in chapter V I argue that I can be held accountable and criticised if I hold a belief about you without caring sufficiently about believing truly. Thus, these beliefs are often false or unjustified. And this is why cases like PLANT and MATHS are both instances of accountability and pertinent to present purposes, since I'm looking at cases where an epistemic breach can be detected and constitute a wrongdoing. However, notice that not all the beliefs that belong to the practices of accountability need to be false or unjustified—I can show that I don't care about getting things right about you even if I end up forming a true, and even justified, belief. So, the range of cases that belong to the practices of accountability is broader than the kinds of cases exemplified by PLANT and MATHS.

the fact that I flout epistemic norms, more than for the fact that I fail to comply with epistemic norms in and by itself.

For instance, recall PLANT. I come back home, notice that one of my plants has been damaged and form the false belief that you, my housemate, caused the damage. To make the example more vivid, suppose that we are close friends,[27] that I usually give the benefit of the doubt to my friends and that you are generally quite a careful person. Here, I do form a belief upon inconclusive evidence. I take a broken leaf to be conclusive evidence of your guilt, and in doing so I dismiss other pieces of evidence and alternative plausible explanations. I give too much weight to certain considerations and fail to see the normative force of others, and I settle for the first explanation that comes to mind.

However, I don't seem to be criticisable *for that*. My failure to abide by epistemic norms is not treated and assessed *as a failure to know the truth*, or to respect the intrinsic value of complying with epistemic norms. As I weigh my evidence incorrectly and fail to gain knowledge, I am bad for something more than that incorrect weighing. In fact, you deem me unfair and unjust. You think that I am uninterested and uncaring, or maybe uncharitable and unsympathizing. I am someone who thought the worse of you; perhaps who's lazy. My belief indicates that I haven't paid sufficient attention to how much care you use to my plants; I have an inaccurate opinion of your character. I am open to criticism on the basis of the fact that I violated certain epistemic norms but not because I violated them, in and by itself.

Consider a second case, MATHS. Suppose that we are colleagues in an accountability office. Exasperated after fighting with some calculations, I wonder out loud how to carry out an operation, and you promptly provide an answer. You are a woman, and I find myself suspicious of its veracity. The fact that you are a woman might or might not be consciously recognised by me as a reason, and yet it does play a role when I am settling the question whether your suggestion is the answer to my query.

Here, I do flout epistemic norms in believing that your answer might not be correct, given that my belief is supported by reasons of the wrong kind. My belief formation is hampered by considerations that do not bear on the truth of the content of the belief. However, again, you wouldn't criticise me *for that*. I am not open to criticism because I draw on the wrong kind of consideration to support my belief[28] as much as because I let *those* specific considerations support it. *What* considerations I take to support the belief is what informs the *kind* of criticism to which I

[27] However, I wouldn't be open to criticism only if we were friends. For I owe it to *anyone* about which I have a belief not to get things wrong about them.

[28] In section 6 below I clarify that, of course, I could also be criticised simply for drawing on the wrong kinds of reasons. However, that issues quite a shallow kind of criticism.

am open. In this case, on the basis of the very considerations I take to be good reasons, I can be deemed misogynist, unjust, patronising; someone who fosters inequality. But if I doubted your answer on the basis that you wear a purple T-shirt, which brings bad luck, the very same epistemic breach would have had another meaning altogether. I would be superficial, superstitious, silly.

2.4. ENVY, MASK

Finally, sometimes I violate an epistemic norm and you criticise me for my belief. However, neither do you criticise me for my breach in and by itself, nor because I've wronged someone in believing as I do.

For instance, say that I believe that my sister does not deserve to be more popular than I am (ENVY). We are both teenagers. As a matter of fact, she is very kind with everyone, volunteers for all sorts of initiatives and succeeds in all the activities she takes on. It is difficult to pin down the truth conditions of statements about desert, for different people have different standards as to what is sufficient to deserve something. However, I have set the bar unreasonably high, and I would not think the same of other people if they were popular by doing exactly what my sister does. I am clearly failing to see accurately the strength of my reasons because I do not want to see it. I am deceiving myself; I am not lucid. I am not what a believer should be to count as epistemically good in the narrow sense. And yet, I am not criticised because I form an unjustified belief, or because I am not good epistemically. Furthermore, in this case my belief does not obviously count as an instance of wronging either. As a matter of fact, I do care about getting things right about my sister. If anything, I would want *everyone else* to get things right about her. I follow her movements, listen to what she says, ponder over the matter repeatedly. My belief is not the result of carelessness in any clear sense.

Yet, I am open to a wide range of evaluations for believing as I do. First and foremost, I am envious. I'm bitter, negative, spiteful. You could also think that I am a self-pitying person, who prefers to shift the burden of their own unhappiness onto other people. My belief can express that I have some paranoid tendencies. I am in denial, likely quite fragile and insecure. And all this emerges from the very way in which I see reasons more than from the fact that *I am wrong* in doing so. It emerges from my distinctive sensitivity to reasons; from the fact that I take *certain particular* considerations to be conclusive.

Let me mention one last case. Some time ago in London, I found myself passing by a protest against lockdown. One of the protesters shouted at me that I should remove my mask because 'blue masks contain asbestos'. Call this MASK. Now, this man held this belief; he was not wearing any mask, repeated his statement more than once, was protesting against anti-Covid

measures. The belief that masks contain asbestos is false, unjustified, and likely to be utterly unsupported. And I doubt he would have been able to draw my attention to any reason why he holds it, or why I should. In believing it, he surely violates all epistemic norms, however we want to understand their content. However, we don't criticise him *for this* as much as for being someone who's paranoid, aggressive, uncritical.

Also, it's not obvious that we criticise him because he wrongs someone. Of course, the actions he might carry out in light of this belief are likely to be harmful to others, *e.g.* not wearing a mask indoors. But I will argue in chapter V that we should not derive the wrongness of a belief from the wrongness of the actions to which it can lead. Moreover, the belief itself is not about anyone. And even if we agreed that he does wrong others by believing as he does, the evaluations to which he's open would not (all) be concerned with the fact that his belief is likely to lead to harmful actions. It is on the basis that he *holds* that *belief* that we can deem him gullible, paranoid, impulsive, over-confident. These are not assessments that depend on the fact that he would act in certain ways. And although they are *supported* by the acknowledgement that he holds a false and unjustified belief, they are not reducible to it.

2.5. The simple view again

At this point, a sympathiser of the simple view might stress that its apparent inability to account for the cases discussed in 2.1-2.4 is due to the narrow reading of it, and that a wider understanding of the significance of the epistemic domain might enable it to strengthen its explanatory power. This cluster of views interpret epistemic value as instrumental to achieve some other good.[29]

One possibility is that abidance by epistemic norms guarantees that we are reliable sources of information for others. Ultimately, we value being knowledgeable for others.[30] It is important to believe only truths, or only upon conclusive evidence, because we are members of a community that shares knowledge. By the same token, we also value being *deemed reliable* by others; being recognised as knowledgeable. We abide by epistemic norms because that increases our likelihood of being believed: if I support my beliefs appropriately, I will be in the position to give you my reasons for them when asked and convince you of their truth.

[29] For instance see Fricker: "For it is all too easy to specify the value of truth, and thereby the knowledge that captures it for us, in purely practical terms without reference to our epistemic motives: we need plenty of true beliefs in order to successfully pursue our practical and other purposes in life" (Fricker 2009, 5).

[30] *Cf.* "We don't strictly speaking blame poor epistemic performers, just more or less automatically deduct credibility points from them in our internal scorekeeping, and that shows in how we treat their testimony, inquire together with them, or rely on their reasoning" (Kauppinen 2018, 2). Others argues that one's beliefs should be such that they are properly assertable should the need for such information arise, *e.g.* (Goldberg 2014).

This reading offers an easy explanation for cases like MATHS, where I fail to treat you as a credible and reliable fellow member of this community,[31] and it might illuminate MASK, on the basis that we may criticise the protester for spreading false information. However, so understood, the simple view becomes unable to account for all the cases in which beliefs don't show regard or disregard for others and their testimony. For example, Sherlock is quite disengaged and often exercises his capacities out of curiosity, vanity or boredom, which are self- rather than others-centred. And Goofy is usually the only victim of his own silliness. Mr. Norpois doesn't show regard or disregard for others either: in believing as he does, the questions whether he is knowledgeable for others or trusts others simply do not arise.

In general, the simple view would be unable to explain why we are responsible for the beliefs we *simply hold*, without sharing or acting upon them. And yet, it seems to me that any account of responsibility for beliefs should be able, at least potentially, to explain why I am criticisable or commendable for the beliefs that I simply hold. So that for example in MATHS I am not criticisable only if I fail to trust your answer or testimony but also if I merely believe that, if I had a problem, I should not come to you for help. Even more vividly, even if I never mentioned or acted upon the belief that my sister does not deserve to be popular, I would *still* be envious and bitter, and legitimately open to criticism for that.

There are other alternatives. A second option is that we value abidance by epistemic norms insofar as it guarantees that we *act* better, in the sense that acting knowledgeably lowers the probability of error in action. It's important to pay attention to evidence and reasons because a failure to do so might lead us to commit practical errors. So, the belief that women are not good at maths does not open me to criticism insofar as it's poorly supported but rather because it is likely to make me act in a wrong way, *e.g.* not to hire women for certain important roles thereby fostering inequality. A third possibility is that we care about abidance by epistemic norms because we want to know how we should feel at a given time.[32] The idea is that our emotional life depends on what we believe is the case, for we will be angry, frightened, or joyful according to what we take to be true. The more we know what's true, the more we are secure in our emotional reactions. So,

[31] Something along these lines has been extensively explored by Miranda Fricker (Fricker 2007), who recognises cases like MATHS as instances of what she calls 'testimonial injustice'. Very briefly, the idea is that sometimes my failure to weigh your testimony correctly has a certain significance because the reasons why I do not trust you have nothing to do with your capacities as a witness but are rather concerned with other aspects of your persona that should not matter when I assess your capacities as a witness. Sometimes, that is, I am criticisable for not trusting you because my distrust is due to biases and prejudices against you. In the light of my arguments in sections 3, 4 and 5, I resist naming these kind of cases instances of *epistemic* injustice: the fact that belief formation is involved does not mean that the relevant responsibility is epistemic. Rather, I think it would be more illuminating to simply say that someone is mean, aggressive, or close-minded, and that this is expressed both in how they act and form and revise beliefs.

[32] Owens (2013). He refines the view in Owens (2017).

in PLANT, I do badly insofar as I put myself in the position of feeling resentment and annoyance mistakenly. I could have believed better and avoided unsupported bad feelings.

Along these lines, the simple view would state that I am epistemically responsible for abiding by epistemic norms *insofar as* that affects my conduct or my emotions. In particular, I'm epistemically blameworthy for my failures to abide by epistemic norms insofar as those lead me to act badly or form the wrong emotions; I am epistemically praiseworthy for complying with epistemic norms because that ensures that my actions and feelings are the right ones to carry out and have. However, it's not hard to see that, if the significance of the epistemic responsibility involved in the simple view is to be understood along these lines, then the simple view cannot accommodate all the cases under consideration. For not all those believers are evaluated for their beliefs on the basis that they will issue in certain actions and emotions.

In general, the chief problem seems to be that, however we decide to specify the significance of the epistemic domain and explain the importance we place onto abidance by epistemic norms, our understanding of 'epistemic responsibility' ends up being overly specific, and the simple view essentially unable to account for all the cases that we recognise as instances of responsibility for beliefs. The problem is that the believers in SHERLOCK, GOOFY, NORPOIS, ABSENCE, PLANT, MATHS, ENVY and MASK cannot all be taken to commit the very same kind of epistemic failure, or an epistemic failure that is important for the very same reasons. To put it in more general terms, the beliefs for which we are criticised or commended are not all criticisable or commendable on the same grounds.

At this point, those who sympathise with the simple view might argue that the mistake is mine in assuming that the simple view has to specify the significance of the epistemic in such a univocal way. Perhaps the reason why none of the readings suggested can account for all the cases considered is that it's impossible to give a simple answer to the question why it matters whether we comply with epistemic norms. We are open to criticism in a variety of ways and for a variety of reasons. The simple view could simply accept *all* the readings proposed.

My reply is this. I agree that it's impossible to give such a simple answer. And I agree that it would be attractive for the simple view to accept all the possible readings of the significance of the epistemic. However, insofar as its advocates insist on talking of 'epistemic blameworthiness' and 'epistemically criticisable', it seems to me that the burden is on them to explain in what sense the different readings would *all* be ways of understanding *epistemic* significance. That is, they would have to clarify why, if believers are criticised for violating epistemic norms *insofar as*, in violating them, they fail to either gain knowledge, or act correctly, or feel the right emotions, they would all

be criticised *from an epistemic point of view*. Talking merely of an 'epistemic' dimension of assessment does not seem suitable to foster a pluralist account. Perhaps, it *could* become one, if we allowed for a long disjunction that specifies the significance of the epistemic. But I don't see why we shouldn't be pluralist to begin with.

3. A pluralist virtue-centred framework

In the rest of the chapter, I argue that my aretaic model can offer a pluralist alternative to the simple view—one that rejects the dichotomy that underlies it. More precisely, the aretaic model offers alternative answers to both the conditions and the significance questions, *i.e.* to the question why I am sometimes open to *criticism* and sometimes to *commendation* for the things for which I am responsible, and the question what is the nature of the relevant responsibility-involving evaluations.

First, where the accountability view, and a fortiori the simple view, addresses the conditions question by reference to norm-compliance and norm-violation, the aretaic model explains the fact that we are sometimes commendable and sometimes criticisable for our beliefs by articulating the notion of 'evaluative orientation' in aretaic terms. Very crudely, I am always commendable for the beliefs that express my virtues and always criticisable for the beliefs that express my vices.[33] Second, where, on the simple view, we are evaluated for our beliefs epistemically,[34] on the aretaic model we are assessable for our beliefs in many different ways—as many as there are virtues and vices. In fact, virtues and vices are associated with varied and diverse dimensions of assessment, which cannot be reduced to the epistemic and moral ones only.

To substantiate my proposal, I need to address three questions. First, how to understand virtues and vices; second, what it means to articulate the notion of 'evaluative orientation' in aretaic terms; third, in what sense virtues and vices are associated with as many distinctive dimensions of assessment. I do so in 3.1, before showing how my pluralist framework can illuminate a variety of cases in 3.2. Sections 4 and 5 reassess SYMMETRY and DISTINCTION in the light of my proposal.

3.1. Virtues and vices

According to John McDowell, virtues are those "states of character whose possessor arrives at right answers to a certain range of questions about how to behave" (McDowell 1979, p. 331).[35] In

[33] And when my beliefs express other values, that are not obviously good or bad values to have, I can be open to different sorts of evaluations.

[34] Or, for some (*e.g.* the advocates of pragmatic encroachment), *either* morally *or* epistemically.

[35] McDowell is influenced by Iris Murdoch, who talks of *moral vision*, understood as vision of moral things. The virtuous person is sensitive to the moral features of the world, and the deliverances of this sensitivity are a kind of knowledge (Murdoch 1971). See Clarke (2017) for a recent development. According to Clarke, the "sensitivity conception of virtue rejects the modern aspiration to codify moral requirements but defends the objectivity of those requirements. The virtuous person and the virtuoso embody standards that cannot be definitively embodied any more than they can be

general, I suggest we think about virtues and vices as *distinctive patterns* of being sensitive to reasons. They are 'distinctive patterns' in the sense that in most (if not all) situations there is one way in which the paradigmatically kind or dishonest person would see situations; something they would consider to be salient, some reason they would find compelling. The paradigmatic kind person will be attentive to *certain* aspects of each situation and react to them in ways that are essentially different from how the paradigmatic honest or callous person would. Hursthouse makes this very point in describing an honest person:

> They hasten to correct a false impression their words have led you into which would be to their advantage; they own up immediately without waiting to see if they are going to be found out; they give voice to the truth everyone else fears to utter; they are concerned to make sure you understand what you are signing or agreeing to do for them. We expect a reliability in the actions that reflect their attitude to honesty, too. We expect them to disapprove of, to dislike, and to deplore dishonesty, to approve of, like, and admire honesty, and so we expect them in conversation to praise or defend people, real or fictitious, for their honesty, to avoid consorting with the dishonest, to choose, where possible, to work with honest people and have honest friends, to be bringing up their children to be honest. (Hursthouse 1999, 11)

Being honest is not just a matter of doing or believing *something*. It is a matter of seeing and being moved by reasons in consistent ways, in a variety of situations. The honest person will behave, entertain certain thoughts, hold or reject some opinions in ways that we would recognise as being instances of one and the same sensitivity—one that we can call, indeed, 'honesty'.

Along these lines, I am committed to the claim that virtues and vices are sensitivities to *both* reasons for acting and reasons for believing.[36] For instance, think about generosity. Suppose that you are happy to invite people over for dinner and serve them delicious delicacies. And suppose that you are also happy to share your insights and help others to see different aspects of key issues. Also, you always think the best of your friends and are quite charitable in your judgements. Here, we need not say that you are *both* practically *and* intellectually generous, or that you are generous in your actions and generous in your beliefs. Rather, you are a generous person, who sees the possibility of sharing as a good reason to do so. The fact that different situations require that your generosity gets manifested in different ways (in the practical and theoretical realms) does not entail

codified. The standards have got to be individually appropriated (and they can never be appropriated once and for all)" (Clarke 2017, 1:47).

[36] This will be crucial below, both in reassessing SYMMETRY (Section 4), since it implies that the same dimension of evaluation is open to both actions and beliefs, and in reassessing DISTINCTION (Section 5), for it suggests that most virtues and vices involve both moral and epistemic elements, traditionally understood.

a split between kinds of generosities. Rather, the fact that those different ways can be traced back to a single sensitivity suggests that there is one and only corresponding virtue, namely generosity.[37]

In chapter II, I argued for the idea that each of us has a *distinctive* sensitivity to reasons, their evaluative orientation. And I just claimed that virtues and vices are *distinctive* patterns of sensitivities to reasons. Let me clarify the interactions between these two claims. The key point is that, although I can be kind or callous, none of us is the ideally kind or callous person. My distinctive evaluative orientation involves different virtues and vices in different degrees. More precisely, our distinctive evaluative orientations are *less pure* and *more complex* than any single vice or virtue.

First, it is unlikely that any of us will be responsive to each and every fact that the paradigmatically kind person would be responsive to. I might take a certain fact to provide a good reason to help someone, despite failing to consider certain others to be sufficiently weighty, even though an ideally kind person would. And this does not invalidate my status as a kind person in and by itself: I do not need to always respond as the paradigmatically kind person would, in order to count as a kind person.[38] Rather, the fact that sometimes I don't react as the paradigmatically kind person only suggests that I am not simply or not only a kind person. When we assess why I did not take those other facts to be sufficiently weighty, we might discover that I am also stingy. And there need be no contradiction. I might be a person who's honest, generous and open-minded, amongst other things, whilst you can be honest, stingy and grumpy, and someone else can be corrupt, generous and open-minded. This explains in what sense no single virtue or vice can provide an exhaustive account of the sensibility *I* display in responding to reasons.

Also, it explains why my evaluative orientation is not reducible to *a sum* of virtues of vices: they affect each other in such a way that my kindness is likely to be different from your kindness, according to what sorts of person you and I are. For instance, if I am kind and gullible whilst you are kind and cunning, both you and I share the same virtue, even though we might take different situations as good reasons to act or believe. So, saying that virtues are linked to *a distinctive way* of seeing things doesn't mean that every person who has that trait will see things in exactly the same way. Nor does it mean that, if we see things in the same way, we will react in the same way. Sometimes both you and I recognise the same situation as a reason to do or believe something and

[37] In fact, suppose that I were only generous or kind with my beliefs and not with my actions, or viceversa only in acting and not in thought. It's even difficult to outline a plausible example; it's hard to think of someone who acts kindly towards you and yet believes unkind things about you. But even when it's possible, *e.g.* if we think about someone who's generous with their beliefs (however we understand that) but ungenerous with their actions, we would be hesitant to call them 'generous' at all. If someone is utterly insensitive to the *practical* reasons to which a generous person would be sensitive, then we have reasons to doubt that their sensitivity to some theoretical reasons, which seems to express generosity, really is an expression of generosity.

[38] I argue for this point in chapter II, section 4, where I take issue with the one-instance worry.

yet carry out different actions or form different beliefs. For instance, suppose that both you and I see that Sara is blue and in need of support. However, given other traits of our personae, I might play the clown and overburden her with attentions, whereas you might remain silent and respectfully distant. *Both* our reactions would express kindness, and yet they would do so in different ways, in accordance with the overall nature of our evaluative orientation.

In fact, that one of our virtues or vices cannot be assessed independently of the overall configuration of our evaluative orientation is further proven by the fact that we often evaluate differently the same action or belief according to who's the agent or believer—according to what kind of person they are. Suppose that we both end up visiting Sara. And suppose that we are both kind and caring but I am also generous, cheerful and optimistic, whilst you are also lazy, reserved and moody. Even though we *both* visit Sara, she will react differently to our gestures. She might be surprised at yours but not at mine, or she might be more touched by your visit insofar as you took her moodiness to be a conclusive reason to go to her house even though, given other features of your sensitivity to reasons, it would have been very easy for you to consider respecting social boundaries or a lazy afternoon at home as conclusive reasons *not* to go to visit her. Given *my* evaluative orientation instead, the fact that I perceived Sara's mood as a good reason to visit her is less remarkable because, if not out of kindness, I would have been sensitive to that same fact on the basis that I am, for instance, generous. Other traits of my evaluative orientation would have made me consider Sara's status as a conclusive reason to do something for her.

In chapter II, I also argued that someone's evaluative orientations are oriented by their values. To have a value involves having the corresponding sensitivity to reasons; to recognise and react to certain aspects of situations.[39] I argued that someone can have a certain value even without being able to describe themselves using the corresponding value-concept. To count as having a value, they need to endorse the reasons that having that value makes them responsive to, rather than *the fact* that they have the value. For instance, if I am honest, I need not endorse that I am honest as much as that copying during an exam would be cheating (and thereby bad). If I refrained from copying for some other reasons, *e.g.* that today is Thursday and copying on Thursday brings bad luck, I wouldn't be someone who values honesty because I wouldn't act on the right reasons; on the reasons on which an honest person would act.

Now, virtues are patterns of being sensitive to reasons that are not themselves *oriented* by values, in the way in which my evaluative orientation is oriented by my values. Rather, they *are* values. This is so precisely because they are *patterns*; the purest form of a certain way of being

[39] I defined the meaning of the phrase 'someone has a value' in chapter II, section 2.2.

sensitive to reasons. They capture one consistent way in which things are relevant, in each and every situation. Thus, sometimes I have a value—one that orients my sensitivity to reasons—in that I have a certain virtue. A virtue is one of the values one might have. I can value honesty, generosity and kindness, amongst other things.

By contrast, a vice can either be a *lack* of value or an *anti*-value. With some vices, that you have a vice simply means that you fail to be sensitive to reasons to which someone who has a value would be sensitive. For instance, if you fail to see things as a patient person would, you are *ipso facto* impatient. To count as being impatient, it's sufficient to be insensitive to the reasons that someone who's patient would be sensitive to. When your actions and beliefs display vices of this kind, you are criticisable for not having a value; for being a person who cannot see certain things as valuable.

With other vices, the fact that you have a vice means that you take certain things to be reasons for doing or believing something even if they are not, and that you do so following a pattern that instantiate a *non*-value; an *anti*-value. You are positively moved by bad reasons; you take the wrong things to be valuable. For instance, in order to be malevolent, it is not sufficient that you *fail* to see reasons that a benevolent person would see. Rather, it requires that you see certain things as reasons when they are not. Being malevolent is to have a positive orientation to see things the wrong way, to give weight to the wrong factors. Similarly, the paradigmatically dishonest person is the one who sees someone's accidentally dropping a purse as a good reason to keep it, rather than the one who fails to see someone's dropping a purse as a good reason to pick it up and give it back. When you have a vice of this second kind, you are not, or not only, criticisable for failing to have a value. You are criticisable for having something that is structurally similar to a value, and yet it is *bad* to have.

Importantly, virtues and vices do not exhaust the set of things over which 'my values' can range. As I claimed in chapter II, I can have values that orient my way of seeing things consistently but without issuing obviously wrong or right answers to the questions what to do and what to believe. They cannot be easily positioned on the Right-Wrong dimension. For example, valuing etiquette can orient my beliefs and actions and give stability to my sensitivity to reasons; it can figure amongst my values. Yet, it doesn't obviously bring me closer to the Good, nor does it obviously push me away from it. Similarly, optimism can orient your evaluative orientation and be expressed by your beliefs and actions. And you can be open to commendation for being an optimist in some cases, *e.g.* when it instantiates endurance or good-heartedness, and to criticism in some other cases, *e.g.* if it leads you to self-deception or wishful thinking. The fact that my valuing etiquette or my being an optimist cannot be easily identified as virtues or vices doesn't entail that they cannot count as proper values, or that I cannot be assessed for the actions and beliefs that

express them.[40] *All* values open up possible perspectives from which to assess others. Only, being deemed an old-fashioned or optimist person can figure in both commendations and criticisms, whilst being deemed honest will always[41] be a form of commendation and being deemed dishonest will always be a form of criticism.

For the sake of clarity, when I say that those who are virtuous are sensitive to the *right* reasons, I don't mean to suggest that certain reasons are the right ones, and the virtuous person happens to pick up on them reliably. Those reasons are the right ones precisely *because* they are the reasons to which a virtuous person is sensitive.[42] Conversely, bad reasons are the reasons to which the vicious person is sensitive. In *Reasons without rationalism*, I believe Setiya draws on this very point to argue that we do not need a reason to be virtuous. That is, we should not assume that there are some standards of practical[43] reason that we can understand apart from ethical virtue and then verify whether the "life of ethical virtue…is to be justified by these standards". Rather, "if a virtuous person would be moved by certain considerations, it *follows* that they count as reasons to act" (Setiya 2007, 3). The reasons that correspond to virtues are *good* reasons on which to act and believe.

> So if justice and benevolence are really virtues, they correspond to reasons in their own right: it belongs to good practical thought to give weight to the kinds of considerations to which the just and benevolent person is sensitive. The answer to the question "Why should I be moral?" is not, on this account, supplied by further *reasons* to be moral, which are certified as reasons by a standard other than ethical virtue. It is supplied by the fact that having the moral virtues is a matter of being responsive to considerations that *therefore* count as reasons to act. (Setiya 2007, 3)

Along these lines, the fact that we are open to commendation for the actions and beliefs that display a virtue is not empirical. It's normative. The idea is not that people treat me in a certain way if I do and believe certain things. Rather, I *am to be* treated in a certain way depending on whether I am right in taking certain facts as reasons. This is also proven by the fact that commendation or criticism can be misplaced. You could be wrong in criticising or praising me.

---

[40] Also, it matters *what role* certain values play in my life; even if it is not bad in and by itself to value etiquette, it might *become* bad if I value it *too much*, or if I take considerations about etiquette to override other more important ones. In these cases, I could be indeed open to criticism for expressing *e.g.* superficiality and shallowness.

[41] Of course, if those who judge have a twisted evaluative orientation, then they might praise the vicious and blame the virtuous. But the point here is normative, not empirical: if someone's belief (or action) expresses a virtue, it *should* be commended, and if it expresses a vice, it *should* be criticised.

[42] I take this to be compatible with Murdoch's thought that "virtue is the attempt to pierce the veil of selfish consciousness and join the world as it really is" (Murdoch 1971, 91).

[43] He focuses on the case of actions.

Two clarifications are due. First, above I argued that having virtues and vices is manifested differently by different persons; it cannot be captured by a *list* of actions and beliefs that works for everyone. Your expressions of kindness might be very different from my expressions of kindness. But this is a point about the nature of virtues and how they are manifested. It is to highlight the impossibility of reducing them to simple formulas that are generally applicable. It does not by any means entail that whether you or I are kind is an *arbitrary* matter. It doesn't cast doubts on the fact that having virtues and vices is a normative issue that yields normative evaluations.

Second, above I also claimed that my actions and beliefs can display values that are not obviously virtues, for which I can be either commended or criticised. But again, the fact that different people see optimism or attachment to etiquette in different ways does not mean that it is arbitrary whether I am to commend or criticise. The fact that there's not only *one* way in which I can be normatively assessed for my optimistic belief doesn't mean that the assessments to which I am open are not genuinely normative. Instead, it means that being optimistic in one's beliefs will be a good or a bad thing according to the times, the circumstances and the people involved.

To conclude, let me expand on the thought that virtues and vices are associated with varied and diverse dimensions of assessment. The key point is that, although we call virtues the sensitivities that lead their bearers to answer *well* the questions what to do and what is true, and vices those that lead their possessors to answer them *badly*, there are *different ways* in which they are good or bad answers. The way in which it's good to be generous is different from the way in which it's good to be kind. Virtues correspond to different ways of being good.

In fact, I can assess whether you are good *in one* particular way or another by looking at whether you are sensitive to certain factors or others. These are different if I set out to assess how honest, kind or decorous you are. Accordingly, my commendation would be essentially different in each case; connected with different judgments, feelings and reactions. I might admire your honesty, be touched by your generosity, respect your decorum. Similarly, the reasons to which you are sensitive can be wrong in a variety of ways. And the way in which I criticise you for being unkind is different from the way I criticise you for being obtuse, which, in turn, differs from the way I criticise you for being irrational, hasty, or careless. I might want to prove you wrong if you're arrogant, warn you if you're gullible, avoid you if you're hopelessly close-minded.

To say it with Anscombe in *Modern Moral Philosophy*,[44] an action is typically much more than merely 'right' or 'wrong'; it is kind, unjust, fair, funny. Similarly, I want to say, beliefs are not simply

[44] (Anscombe 1958). *Cf.* Williams' idea that there are 'thick ethical notions', "which seem to express a union of fact and value. The way these notions are applied is determined by what the world is like (for instance, by how someone

true or false, or justified, correct, plausible, rational, supported. They also show that the believer is arrogant, charitable, mischievous, ironic, and so forth.[45] And there doesn't seem to be any good reason why we should reduce the variety and diversity of all these evaluative dimensions to merely two, *i.e.* the epistemic and the moral. I will expand on the latter point in section 5.

3.2. The explanatory power of the pluralist framework

The aretaic model can account for a variety of cases because it doesn't aim to individuate a common denominator that underlies all the instances in which believers are criticisable in the relevant way. There's no single kind of failure that can be ascribed to all believers, nor a failure that has the same significance every time. Rather, believers are criticisable in each case either because their belief expresses a vice, or because it expresses a value that is criticisable under certain respects.

For instance, my failure to believe well has a different flavour in PLANT, MATHS, ABSENCE or ENVY, depending on the broader context, *i.e.* on the content of my belief, on the very reasons I take to be conclusive, and so forth. For instance, you criticise me for believing that you damaged my plant, especially if you are not particularly clumsy, insofar as it shows that I am suspicious and distrustful, someone who does not grant the benefit of the doubt to others, someone who thinks the worst of people. In MATHS, you criticise me for thinking that women are not good at maths because that manifests not only that I am closed-mindedness and superficial but also that I am unjust, prejudicial and patronising; someone who fosters discrimination and inequality. In ABSENCE, the fact that I fail to think about you despite being in the position to help you shows that I am not particularly generous or kind, and probably quite self-centred; I am criticisable for being a person that does not think about others and is not eager to help. Finally, in thinking that my sister doesn't deserve to be popular, I manifest myself as someone who's envious, bitter, uncharitable.

Each of these failures is unique and triggers different reactions in you. You might be hurt, disappointed and offended in PLANT, angry and indignant in MATHS, upset and saddened in ABSENCE. You might try to convince me to revise my belief in MATHS, and either succeed in doing so or simply distance yourself from me, since my beliefs express an evaluative orientation that is incompatible with your values. In PLANT instead, you might try to show me that you are

has behaved), and yet, at the same time, their application usually involves a certain valuation of the situation, of persons or actions" (Williams 2008, 129).

[45] Importantly, even if someone resisted the idea that the kind of aretaic appraisals on which I focus are genuine markers of *responsibility*, I will still have made a case for the claim that beliefs can be normatively assessed in richer and more varied ways than merely in terms of what one ought to believe, or in terms of truth, justification, or knowledge.

trustworthy, if you care about my judgement, or show me that you are very careful, if you want to prove me wrong.

Importantly, your appraisals of my beliefs will vary according to other traits of my evaluative orientation. For instance, if in PLANT, MATHS, and ABSENCE I am also gentle and awkward, you might still criticise me in all of them, and yet in ways that are different from the ones sketched above. You might pity me for being paranoid, narrow-minded, or lonesome respectively. Or again, if I am also gullible and naive, you might be annoyed at my being sloppy, easily deceived, or in the clouds, respectively. In any case, I'm not responsible and criticisable for my belief insofar as it is false, or the outcome of a fallacious reasoning; rather, I am responsible insofar as it indicates that I am a kind of person that sees the world in a certain way and perceives certain situations as conclusive reasons to believe certain things.

The aretaic model is also able to illuminate cases like NORPOIS, which no version of the simple view seemed able to accommodate. In fact, accepting an accountability framework makes it hard to talk of responsibility independently of compliance with or violations of norms. Thus, where no violation of epistemic norms can be detected, there seems to be no room for responsibility-involving criticism. And yet, Mr. Norpois is indeed appraised as someone who's simple-minded and old-fashioned. And this is not because he violates some norms, *e.g.* that we should avoid being *démodé*. Rather, it is because the evaluative orientation expressed by his beliefs displays that he perceives situations as an old-fashioned person would. Importantly, we might disagree as to whether he deserves criticism or commendation because, as for optimism, 'old-fashionedness' is not an obviously positive or negative trait. But these possible disagreements do not entail that we are not in the realm of responsibility altogether; they simply mean that the values that his beliefs express can be read as orienting his sensitivity either well or badly.

Finally, there are examples like SHERLOCK and GOOFY. My opponent might insist that the aretaic model is ill-suited to explain these cases because, even though we commend Sherlock and criticise Goofy, we don't look at the *values* (or lack of values or anti-values) that their beliefs express. Sherlock's beliefs manifest his outstanding intelligence. We admire him for his cleverness and welcome his statements with awe. We deem him especially witty, smart, quick, accurate and reliable on the basis of what he believes; for his ability to discern truths from falsehoods. But it's not that Sherlock's beliefs express that he *values* cleverness, or at least they don't *necessarily* express that. Similarly, we criticise Goofy for being gullible and obtuse. We deem him half-wit because of what (and how) he believes. Often, we pity him, and sometimes even berate him for being so dumb. However, we cannot see dumbness as one of Goofy's *anti*-values.

Here, I am happy to bite the bullet (partially). I would agree that being more or less smart does not express one's values, in standard circumstances—though maybe it does in some cases. When it doesn't, I am happy to concede that our appraisals are not deep responsibility-involving evaluations. We look at Sherlock's rapidity and reliability with the same awe and admiration with which we look at beautiful blue eyes, or at the speed and accuracy with which a wild goat climbs a vertical wall. Sherlock is not responsible for *that*. However, as soon as Sherlock's and Goofy's beliefs do express something of their evaluative orientations, then we are in the realm of responsibility once again. So, for instance, Sherlock's beliefs can sometimes show lack of empathy and arrogance, which we receive with irritation and contempt. Or they can express interest in the resolution of some cases, sensitivity to someone's right to know the truth about the loss of their beloved ones. By contrast, Goofy's sloppy beliefs can express that he is good-hearted, gentle and caring, funny and cheerful, and we might greet his naivety with tenderness. Or they might display that he is incapable of understanding the risks of believing poorly, apathy, laziness. Here we do deem them responsible, *i.e.* as soon as their beliefs express something of their evaluative orientation. I will come back to these points in section 6.

4. SYMMETRY reassessed

It remains to be explained precisely how the aretaic model offers an alternative to the simple view, *i.e.* how it affects SYMMETRY and DISTINCTION. In fact, someone might insist that we simply reformulate the two theses in aretaic terms. I want to resist this suggestion. If we did so, reference to virtues would be a terminological shift rather than a substantial reformation. I focus on SYMMETRY in this section and on DISTINCTION in the following one.

Recall that SYMMETRY is the thesis according to which we are morally responsible for our actions and epistemically responsible for our beliefs. The way in which we are responsible for our actions is essentially different from the way in which we are responsible for our beliefs insofar as actions and beliefs are (at least typically) governed by different norms and evaluated against different standards. And so rarely, if at all, we are morally responsible for beliefs or epistemically responsible for actions.

In principle, it is possible to translate SYMMETRY into aretaic terms. For instance, by saying that some virtues and vices primarily pertain to actions and some other ones to beliefs. However, not only do I doubt that there is a neat way of drawing a distinction of this sort;[46] also, as I said in 3.1, I think that the vast majority of virtues and vices are manifested in someone's sensitivity to *both* practical and theoretical reasons. There is no generosity for actions and generosity for beliefs,

[46] As I argue in section 5.

but the very same virtue, *i.e.* generosity, is manifested in both someone's actions and beliefs. Someone is responsible for their actions and beliefs insofar as they express the very same evaluative orientation; the very same virtues and vices. If this is true, then it'd suffice to reject SYMMETRY, for actions and beliefs would be evaluated in the same light.[47]

Put otherwise, the thought is that even if we had a capacity to identify theoretical reasons as distinguished from a capacity to identify practical reasons, the two capacities would be oriented by one and the same set of values. What I take to be sufficient to support the formation of a belief will be in general[48] coherent with what I take to be sufficient to support the execution of an action. In fact, you would be at least surprised and at most bewildered if I were typically[49] generous, honest, and kind in my actions and typically selfish, duplicitous and uncharitable in my beliefs; or if I were cruel, venal and depraved in my actions and compassionate, considerate and fair in my beliefs. And the very fact that we would identify *tensions* between these beliefs and these actions proves that we *expect* them to instantiate the same values. If there were two capacities of perceiving theoretical and practical reasons and they were not only distinct but also oriented by independent sets of values, my taking a situation to be a sufficient reason to believe that *p* and to act as if *non-p* would raise no concern.

Of course, none of this excludes that some traits might be mostly concerned with reasons to believe than with reasons to act, or viceversa. For instance, that someone is open-minded mostly means that they will take into consideration more possibilities before settling their mind. And, someone might insist, this shows that open-mindedness cannot be displayed by someone's *actions*. Even setting aside possible objections to this line—*e.g.* that someone's open-mindedness affects their practical reasoning and therefore their conduct as well—, the point is that it wouldn't prove that they constitute a distinguishable and independent domain, governed by a distinctive and

---

[47] Perhaps it's possible to read Montmarquet as suggesting a similar move here: "So far as issues of responsibility are concerned, then, the fundamental category is that of the virtues - rather than action or belief. … Instead of attempting to characterize - what has been arguably the most important goal of this branch of philosophy - in purely general terms, such notions as that of an 'intentional act' or a 'voluntary act', the alternative suggested by our discussion would be a continuation of something more like Foot's inquiry into the notion of a 'virtue' and its connection to the will" (Montmarquet 2008, 389).

[48] Of course, our capacity to perceive reasons can be affected by some factors: I can be more or less receptive if I am tired, stressed, anxious, myopic, confused and so forth. Also, revisions of one's values can occur. But both tiredness or value-revision would affect someone's sensitivity to practical *and* theoretical reasons equally.

[49] I add this qualification because having a virtue seems to be compatible, at least sometimes, with not having it in some contexts. This thought draws on my treatment of the one-instance worry. At least certain virtues allow to be relativized to certain topics. Kindness seems to be one of them. Of course, being someone who's kind all the time and unkind with cyclists does create some tensions in my character. But they will be such that I can live by them without being particularly irrational, since the scope of my unkindness is quite narrow. By contrast, it seems that other virtues don't allow to be relativized. For instance, consider honesty. There would be something problematic in admitting that you are an honest person even though this one time you believe you could steal that purse, or avoid paying taxes, or cheat for an exam. If you were convinced of one of these things, it would be difficult for me to grant that you are generally honest but dishonest in believing that it's okay to take money from others. One or few instances of dishonesty are sufficient to cast doubts on your honesty.

separated set of norms. Rather, it would simply mean that certain sensitivities to reasons happen to concern primarily the question what is true rather than also the question what to do. We may say that they are *vacuous* on practical matters, but this doesn't suffice to think that they belong to a separate normative domain altogether.

5. DISTINCTION reassessed

More delicate is the question whether we should retain DISTINCTION, *i.e.* the thesis that the moral and the epistemic are the two primary, and separate, dimensions of assessment. In fact, the advocate of the simple view might grant the argument against SYMMETRY and accept the idea that most virtues and vices are manifested in both someone's actions and beliefs, and yet stress that some virtues and vices are morally relevant whilst others are epistemically or intellectually important.

Quite surprisingly, most of the authors working in virtue and vice epistemology do accept a distinction between *moral* and *intellectual* virtues and vices, which typically mirrors the distinction between the moral and the epistemic domains.[50] Along these lines, intellectual virtues and vices are traits relevant for the attainment of truth and knowledge, *e.g.* intellectual curiosity, intellectual honesty, intellectual humility, gullibility, intellectual courage, intellectual arrogance and so forth. Strikingly, these are typically understood within a broadly accountability framework: it is good for someone to have an intellectual virtue insofar as it helps them achieve the epistemic good or abide by epistemic norms; it is bad for them to have an intellectual vice insofar as it hinders their way to the truth.[51]

To be more precise, this seems to be true for both the so-called virtue 'reliabilists' and virtue 'responsibilists'. According to the former, intellectual virtues are understood precisely in

---

[50] For an example of how DISTINCTION gets typically translated in aretaic terms see Audi (2008, 404): "The point has two sides: I am in a state that reflects a kind of wrong-doing, the kind of conduct that led to holding the belief; and I have a belief that is epistemically ill-grounded. Very similar points may be put in virtue-theoretic terms: first, I have acted contrary to intellectual virtue; second, I have a belief that was neither virtuously formed nor would be held (other things being equal) by a person of intellectual virtue having the same relevant evidence I do."

[51] Cf.: "'Good epistemic character' must involve an orientation towards the corresponding notion of believing what is true and not believing what is not true" (J. Montmarquet 1992, 336).

Notice that understanding epistemic virtues in terms of traits that favour the achievement of certain goals may risk favouring a quasi-consequentialist construal of them. Quassim Cassam for instance conceives of vices and virtues as ways of inquiring, amongst someone's character traits, that are to criticise or commend according to how likely they are to bring about a certain outcome, *i.e.* knowledge. But then the goodness or badness of a character trait depends on what they contribute to bring about. They are not good or bad *per se*. An aretaic model that draws upon *this* understanding of virtues and vices would not be essentially different from the accountability view: on the latter approach, I'd be responsible and criticisable for a belief insofar as I violated the relevant norms; on the former, insofar as I failed to "seek out and evaluate evidence". In both cases, I'd be responsible for my belief according to how well I formed it. This is not necessarily a problem in and by itself. But it may be a problem if we were resorting to virtues to provide a radical alternative to the accountability view. Understanding character traits as being to criticise and commend according to whether they trigger proper or correct procedures of inquiry doesn't seem to help with *that* radical project.

terms of reliable, truth-directed cognitive dispositions.[52] Someone is virtuous insofar as they are competent in gaining knowledge. Virtues, in this context, include cognitive faculties or powers like vision, memory, and introspection. By contrast, virtue responsibilists deny that mere reliability is sufficient to qualify as an intellectual virtue, and some of them even claim that reliability is not necessary at all.[53] And yet, also for the responsibilists virtues are such insofar as they have a certain connection with *truth*. For instance, Montmarquet argues that virtues are the traits of character that someone who *cares about truth* would want to develop (even though they are not in fact reliable). According to Zagzebski, intellectual virtues, unlike the moral ones, have a single underlying motivation that can be roughly understood as the desire[54] to obtain truth and avoid falsehood. Or rather, as the desire to be in "cognitive contact with reality", which includes states like understanding and wisdom, besides propositional knowledge. So also responsibilists think that intellectual virtues put us in touch with some 'epistemic' favourable status.[55]

I say that such a reformulation of DISTINCTION in aretaic terms is *surprising* for two main sets of reasons. First, because a commitment to DISTINCTION seems to me to lose the innovative potential of an aretaic perspective; second, because distinguishing moral virtues and vices from intellectual ones is not as straightforward as it may seem.

To begin with, one of the main advantages of reflecting about responsibility in terms of virtues and vices is precisely that they multiply the dimensions of evaluation that are relevant in assessing someone's beliefs. That is, they allow us to enlarge the pool of possible assessments from 'true', 'false' or 'justified', 'unjustified' to 'kind', 'unjust', 'fair', 'funny', 'arrogant', 'charitable', 'old-fashioned', 'optimistic', and so forth. An aretaic perspective provides us with a richer and more varied palette of explanatory tools. Furthermore, if we accepted to reformulate DISTINCTION in aretaic terms, we would be failing to provide a radical alternative to the accountability view. In fact, DISTINCTION seems to develop naturally from *accountability* premises: it is the idea that different domains are governed by different norms that fosters the compartmentalisation between the epistemic and the moral, since it facilitates an independent investigation of each. In fact, we can explore the nature, scope, source and aim of norms like "Do not kill!" and "Only believe what you

[52] Goldman (2012), Greco (1999), Greco and Reibsamen (2018), Sosa (2000).
[53] Axtell (2008), Battaly (2008), Code (1984), Hookway (1994), Zagzebski (1996; 2004).
[54] *Cf.* (J. Montmarquet 1992, 336): "Let us refer to this underlying desire to believe what is true and to avoid belief in what is false as (epistemic) conscientiousness". For criticism about the idea that characterization of virtues need make reference to someone's *motivation* see (Driver 2003, 380), who states that "what does mark moral virtue is that its traits systematically produce good for others whatever the agent's psychological states".
[55] For instance, (Axtell 2008, 66): "The reduction of epistemology to the theory of knowledge…also functions to inhibit our conception of the role of the reflective virtues in the intellectual life, since firstly we seek interesting truths, and secondly our virtue-relevant goals include not just knowing and believing truly, but also higher or more distinctively human epistemic achievements such as theoretical understanding". Also Baehr (2008) argues that questions pertaining to intellectual virtue that can be addressed in relative isolation from traditional epistemology. And yet, they seem to me to remain in the very same neighbourhood.

know!" one at a time, without any need to investigate their interconnections.[56] And if we start our investigation from the normative standards that govern different domains, then being able to delineate where precisely their *boundaries* lie will be an urgent issue. By contrast, if we start from accepting a plurality of dimensions of assessments, then the project of dividing them into *two* sets is less natural; it's unclear what would be the rationale behind it. Furthermore, if we ground responsibility in someone's evaluative orientation, we'll necessarily be looking at the *interactions* between different virtues and vices rather than at each in isolation. On the aretaic model, in order to understand kindness, we need to verify how it is expressed by different believers, how it interacts with other virtues, how it is affected by their vices.

In effect, there seem to be certain relations of necessity and entailment between traits. In chapter 3 of *Virtues of the mind*, Zagzebski explores some of them. For instance, certain virtues require certain others. Consider the connection between honesty and understanding:

> For example, honesty is on all accounts a moral virtue. It is a virtue that requires that one tell the truth. But it is not sufficient for honesty that a person tell whatever she happens to believe is the truth. An honest person is careful with the truth. She respects it and does her best to find it out, to preserve it, and to communicate it in a way that permits the hearer to believe the truth justifiably and with understanding. But this in turn requires that she have intellectual virtues that give her as high a degree of justification and understanding as possible. She must be attentive, take the trouble to be thorough and careful in weighing evidence, be intellectually and perceptually acute, especially in important matters, and so on, for all the intellectual virtues. The moral virtue of honesty, then, logically entails having intellectual virtues. (Zagzebski 1996, 158)

Even though I suggest we don't categorise virtues as moral or intellectual, this passage is important in that it highlights that sometimes having a virtue is a condition on possessing another one. Someone cannot be both careless and courageous because, without sufficient care and attention, they would only be reckless. The two sensitivities that we call carelessness and courage cannot coexist. Similarly, some vices are entailed by other vices.[57] For instance, "Envy, pride, and the urge to reinforce prejudices can easily inhibit the acquisition of intellectual virtues. A person without sufficient self-respect and an inordinate need to be liked by others may tend to intellectual

[56] In fact, accepting that there are epistemic oughts and moral oughts also generates the problem whether they are commensurable. Mills (1998) discusses whether there can be conflicts between these different kinds of duties. Feldman (2000, 694) and Brown (2020) argue for *incommensurability*, *i.e.* the thesis that if these oughts conflict there is no way to adjudicate between them, no meaningful question about what I ought to believe all things considered. By contrast, some grant that conflicts are possible, *e.g.* Code (1987) and Howard (2020). Of these, some maintain that moral duties always override epistemic duties, *e.g.* Chisholm (1966, 60), whilst others argue that we should all-things-considered believe what is epistemically rational to believe, for instance Wedgwood (2017).

[57] According to Haack (2001), there are certain ways of being unjustified that entail being morally unjustified.

conformity. An egoistic person will want to get his way, and this includes wanting to be right" (Zagzebski 1996, 159).

Along these lines, it emerges that not only the project of reformulating DISTINCTION within an aretaic framework is unattractive; it might not even be *feasible*. First, virtues and vices are interconnected in deep and important ways. Second, as Zagzebski points out, they are so diverse that the difference between an intellectual and a moral virtue might be the same as the difference between any two moral virtues or any two intellectual virtues.[58] Finally, there is little to no agreement amongst scholars as to how to distinguish between moral and intellectual virtues.[59] This doesn't prove that it cannot be done. But it does suggest that the task is less trivial than it might seem at first. Julia Driver, for instance, suggests that moral, but not intellectual, virtues benefit others.[60] She writes:

> The intuition I would like to explore is that intellectual virtues have—as their source of primary value—truth or, more weakly, justified belief for the person possessing the quality in question, and this is what 'getting it right' means for the intellectual virtues, whereas for the moral virtues the source of value is the benefit to others, the well-being of others, and for the moral virtues this is what 'getting it right' means. (Driver 2003, 374)

But for one thing this suggestion seems unable to accommodate several counterexamples. Truthfulness and open-mindedness are usually taken to be intellectual traits, and yet they involve benefitting others in a strong sense. And this way of understanding moral virtues seem to capture just *one* kind of virtue. There are many broadly *practical* virtues that don't fit that definition, *e.g.* prudence, temperance, courage.

6. The 'epistemic' and the 'moral' reassessed

The overarching aim of this chapter has been to argue against the idea that there are two *distinct* domains to which all (or most) dimensions of assessment involved in responsibility for beliefs can be *reduced*. I clarified that virtues and vices are distinctive patterns of sensitivity to both theoretical and practical reasons, and that many of them are interconnected in important ways. At this point,

---

[58] Similarly, on Hume's view, moral and intellectual virtues are the same insofar as they are all virtues, since all virtue traits derive their value from their pleasing qualities. For an analysis of its reception see Fieser (1998).

[59] See Driver (2003) for an overview. She considers and rejects the following criteria: that moral virtues are voluntary whilst intellectual virtues are natural; that moral virtues are formed by habit whilst intellectual virtues by teaching; that the former speak to the emotional aspects of our nature whilst the latter speak to our rational capacities; that moral virtues are 'other-regarding' as opposed to 'self-regarding'.

[60] More precisely, Driver puts forth a consequentialist account; virtues traits should not be characterized in terms of 'characteristic motivation', which may serve to offer some way of marking a distinction but it cannot distinguish them at the level of value-conferring quality. That is, 'characteristic motivation' does not pick out what is distinctively valuable about the traits in question.

my discussion might be understood as an attempt to brush away the very notion of 'epistemic' and 'moral', or to empty them of all substance. But this needn't be so. Insofar as we use them, 'epistemic' and 'moral' are useful notions that we should safeguard, provided that we are clear on what we mean by them.

In effect, and again contra DISTINCTION, if we understand 'epistemic' and 'moral' along the lines I sketched in section 1, we could say that most virtues and vices involve both epistemic and moral aspects. In fact, someone who has a certain virtue or vice will be concerned both with getting to the truth (or knowledge) and with the relationships of mutual regards. However, they will be so *in certain distinctive ways*. Let me start from discussing the 'epistemic' horn.

As I suggested at the end of section 3.2, I lean toward the idea that there might be a dimension of assessment that we can call 'epistemic', and that it's possible to commend or criticise someone in that respect. For instance, Sherlock Holmes could be someone who's epistemically commendable in the sense that (and insofar as) he is extraordinarily attentive, reliable, fast, smart. But this sense of 'epistemic' is substantially narrow; it yields very shallow evaluations. For it is a dimension of assessment concerned merely with someone's capacity to gain true beliefs, irrespectively of all the other aspects whose importance I have been trying to emphasise. Irrespectively, that is, of *what* someone believes, or *why* they care about believing attentively, reliably, truly on such-and-such topics. As such, it would not be able to illuminate the significance of the appraisals involved in responsibility for beliefs.

To clarify, this line of thought would in no way suggest that certain virtues and vices *are epistemic*. Rather, if we accept this shallow 'epistemic' dimension of assessment, it would be *orthogonal* to the dimensions of assessment individuated by virtues and vices, and at least partly independent of those. That is, all sort of people could be epistemically commendable or criticisable in this narrow sense. They could all be more or less reliable, fast, and smart. But as soon as we start thinking about how *each virtue* and vice is linked to truth and knowledge, the ways in which someone 'gets to the truth' are not to be understood in this shallow, orthogonal sense anymore. For the ways in which a kind person 'gets to the truth', compared to how an honest or wicked or arrogant person 'gets to the truth' will be informed by different values, and be very different from each other.

For instance, if I am kind, I will get to know truths that someone who's honest would not necessarily be interested in knowing, or I will form reasonable and justified beliefs that someone who's honest would have not formed. For instance, your daughter's birthday, or whether you recovered from your injury. And if you are honest, you will be particularly careful in figuring out how much you ought to pay in taxes. But this does not prove that you are 'epistemically commendable'. It proves that you are honest. The fact that you are attentive and careful in *this*

situation, *i.e.* filing a tax form, lends itself to more than a shallow evaluation of your 'rational capacities as a believer'.

I think that the same can be said about generosity, courage, humility, honesty, arrogance, avarice, or malice, in that all require that the subject has 'epistemic' concerns *in some distinctive way*. Most (if not all) virtues and vices bring in material traditionally located in the epistemic domain narrowly understood but they do so in very different ways, to the point that they are not 'an epistemic matter' or 'epistemic concerns' but rather 'a matter of kindness', or 'concerns of honesty'.

When it comes to 'the moral', my suggestion is far more tentative. I am tempted to say that there is no orthogonal and shallow sense of the 'moral'—something like a dimension of assessment that targets merely the extent to which people respect mutual regard. And this is so because the way in which people respect others is *immediately* informed by their values and sensitivity. The way in which someone understands what it is to respect and have regard for others will vary according to the quality of their evaluative orientation. It's impossible to separate it from their values.

For instance, what counts as 'mutual regard' for the kind person will be different from what counts as 'mutual regard' for the honest but bashful person. Only for the former, it will involve things like buying a present for a colleague's daughter or throwing a welcome-back party in someone's honour. For a generous person, being extremely hospitable will be integral part of what it is to honour the relationships of mutual regard, whilst that might not be so for the prudent person.

Along these lines, the 'moral' dimension of assessment would isolate the different ways in which we treat and respect others, as oriented by different virtues. As such, it wouldn't be a *shallow* dimension of assessment. Yet, it would be a *narrow* one. First, because mutual regard is not the *whole* point of many virtues; rather, most virtues are concerned with mutual regard amongst other things. Second, not every virtue is concerned with mutual regard at all, *e.g.* temperance.

In the first three chapters, I have presented my aretaic model, a pluralist virtue-centred alternative to the norm-based accountability view. In the next chapter, I discuss the control requirement and explain in what sense the aretaic model is not committed to it. In doing so, I provide an account of fairness that does not depend on control.

IV. The control requirement

In the first three chapters, I developed an aretaic account of responsibility for beliefs as an alternative to the accountability view. In particular, I articulated the idea that ascriptions of responsibility don't necessarily make reference to violations of (or abidance by) norms. This is so for a variety of reasons. First, often we criticise others for their beliefs even if they haven't violated any norm—in any clear sense[1]—in believing as they do. Second, even if it can be said that they violated a norm, we might not criticise them *for that*[2] as much as for what the fact that they violated it displays of them. Third, it's unclear *what kind* of norms would be relevant in the case of responsibility for beliefs: however we understand the significance of 'epistemic norms', it seems that believers are only rarely appraised 'epistemically'.

But in framing the accountability view in chapter I, I claimed that there's more to it than a firm insistence on the importance of norms and, typically, the assumption that doxastic responsibility is reducible to epistemic responsibility. Two further tenets are central to it. One is that, when we hold someone accountable, we can place certain relevant expectations and demands on them, *e.g.* that they do or refrain from doing something, or that they make up for some bad consequences they brought about. I will discuss this in chapter V. The other tenet is the idea that, if someone is responsible for something, then they must be in control of it. Call this general thesis *the control requirement*. Corollaries are that, if someone is not in control of something, they cannot be responsible for it; and that deeming someone responsible for something, despite their lack control over it, would be unfair.

The control requirement is tailored on the case of action, and yet it is widely accepted as holding for responsibility in general.[3] In fact, the fundamental question in the debate on responsibility for beliefs seems not to be *whether* we need to control our beliefs in order to be responsible for them, but rather *what kind* of control is involved in believing, given that we seem to be responsible for them and yet unable to believe voluntarily. Most proposals in the literature are at pains to address the latter question (call it *the control challenge*),[4] whilst assuming a positive

[1] The qualification is needed because, in principle, we could formulate as many norms as we need to make sense of our practices: if we criticise Mr. Norpois for believing that bothering the Comte of Paris would be unseemly, then he must have violated a norm not to believe that bothering the Comte of Paris would be unseemly. But clearly it would be hard to say why dictates like this would be normative—setting aside the fact that this strategy would generate a problematic explosion of 'norms'.

[2] Similarly, even if someone can be said to have abided by a norm exemplarily, we might not commend them *for that*.

[3] In effect, that we cannot ultimately control 'who we are' or 'what we value' is often taken to prove that *attributability* cannot yield an account of full-blown responsibility.

[4] For some examples of its formulation, see Alston (1988), Audi (2008), Booth (2014), Feldman (2019), McHugh (2012).

answer to former. It is this assumption that I oppose. My aim in this chapter is to show that the aretaic model can vindicate the practices of responsibility for beliefs without committing itself to a control requirement. Put otherwise, that we should question the legitimacy of the control challenge rather than trying to address it.

Traditionally, the sort of control involved in the control requirement is a kind of *voluntary control*. So it's worth stepping back and asking why responsibility is supposed to require this kind of voluntary control, how it's typically understood, and what it would mean to exercise it. In section 1, I draw on the work of Michael Frede to address these questions. Briefly, I suggest that the control requirement depends on one way of interpreting Aristotle's claim that we are responsible only for what is 'voluntary',[5] *i.e.* as stating that we are responsible only for what we can *control* voluntarily. This is the *control reading* of Aristotle's thesis, which can be traced back to the Stoic tradition. Yet, Aristotle's claim need not be interpreted this way. In fact, Aristotle writes that we are responsible for the actions we do *hekontes*, *i.e.* neither by force nor out of ignorance. Positively, Frede suggests, this means that we are responsible for what we do when *we* are motivated so to act. Call this the *hekon reading* of Aristotle's thesis.

Crucially, we don't inherit from the Stoics only the control reading of Aristotle's thesis but also *a certain way* of understanding 'voluntary control', according to which we control an action voluntarily if and only if (and insofar as) we *give our rational assent* to the impulses that motivate it. I suggest that this framework underlies also the contemporary understanding of voluntary control. This *must* be so, for otherwise the case of belief would not come across as problematic, and the control challenge would not arise. It is the Stoic interpretation of Aristotle's thesis, paired with the Stoic understanding of voluntary control, that fuels the control challenge. That is, responsibility for beliefs seems puzzling once we accept that we are responsible only for what we voluntarily control and we understand voluntary control in such a way that it cannot apply to beliefs.

I then turn to show the merits of the alternative, *hekon* reading of Aristotle's thesis. This is well suited to my aretaic model: insofar as my actions reflect my own evaluative orientation, rather than someone else's, I am responsible for them. In the rest of section 2, I develop this line to explain responsibility for beliefs and suggest that, rather than being the harder case, responsibility for beliefs is the *easier* one. For our beliefs are *directly* sensitive to the way in which we see reasons; we believe *hekontes*. Concerns about control emerge specifically when it comes to responsibility for actions because of the possibility of accidents and errors in performance. I conclude by explaining how admitting that we might need to (somehow) control our actions in order to be responsible for them does not commit us to the control reading of Aristotle's thesis.

[5] Aristotle, *EN* 3, 1109b30-34.

At this point, someone might object that resisting the control reading of Aristotle's thesis, thereby renouncing the control requirement, makes it impossible for me to account for the difference between fair and unfair ascriptions of responsibility. In section 3, I resist the worry by offering an account of fairness that fits well with the *hekon* reading. In doing this, I hope to show not only that the aretaic model can avoid the difficulties generated by the control requirement but also that we have good reasons for finding that very requirement unnecessary. Responsibility for beliefs can be made intelligible before any appeal to control is made indispensable.

To strengthen my case, in the second part of the chapter I turn to the contemporary debate on responsibility for beliefs and consider the two most influential ways of *addressing* (rather than rejecting) the control challenge, *i.e.* the indirect voluntary control views (4.1) and the direct nonvoluntary control views (4.2). Both proposals accept the control reading of Aristotle's thesis—and thereby the control requirement—and attempt to accommodate the possibility of responsibility for beliefs in different ways. By raising some worries regarding both strategies, I hope to provide further support for the idea that we'd better resist the control reading to begin with. Finally, I consider Pamela Hieronymi's proposal (5), which I believe offers a way of understanding our relationship with our beliefs in a way that can illuminate responsibility, without addressing the control challenge as it's typically understood.

1. Voluntariness, control and the will

The *locus classicus* for the connection between voluntariness and responsibility is the beginning of book 3 of the *Nicomachean Ethics*:

> Since virtue is concerned with passions and actions, and on voluntary ones praise and blame are bestowed, on those that are involuntary pardon, and sometimes also pity, to distinguish the voluntary and the involuntary is presumably necessary for those who are studying the nature of virtue, and useful also for legislators with a view to the assigning both of honours and of punishments. (Aristotle, *EN* 3, 1109b30-34, transl. Ross)

Here, Aristotle draws a distinction between voluntary and involuntary actions and claims that responsibility-involving evaluations are appropriate only with respect to the former. I am responsible for investing in cryptocurrencies but not for being taller than my sister, and this is so because only the former is something I do voluntarily. Relatedly, it would be *unfair* to hold me responsible for being taller than my sister because that's not something I do voluntarily.[6]

[6] For some formulations of the thought see: "It is often said to be unjust to blame someone for what he could not help doing. We think it unfair to adopt an attitude of disapproval toward someone on account of an act or omission, where this was something outside his control" (Glover 1970, 70–73); "It is an illicit generalization to infer that having alternative possibilities is never required for moral responsibility or free will — throughout an entire lifetime. One could claim this only if one could claim that it is fair to hold persons responsible for being what they are even though

Crucially, the claim that

[1] We are responsible for what we do voluntarily.

is typically taken to be equivalent to the claim that

[2] We are responsible for that over which we can exercise voluntary control.

Reference to control seems helpful to qualify the meaning of 'voluntary' in the relevant way and single out the things for which I am responsible from other things, which can be considered somewhat *voluntary* and yet not something for which I am responsible. In fact, often I welcome what happens to me, or what someone does to me, or for me,[7] as something good; perhaps, it's precisely what I wanted to happen. It is something I *will*. And yet, I am not responsible for things of this kind. *Thus*, we need to appeal to control; for I am not in control of what happens to me, or for what is done to or for me.[8]

Precisely what it means to (be able to) exercise voluntary control over something is a difficult question. For now, in very general terms, we can say that someone exercises voluntary control over an action if they chose to carry it out, started to do so and have the power and capacity to affect its course, where 'affecting its course' means that they either could modify its course, if they so wished, or would try to preserve it in case some external circumstances risked halting it. For instance, I exercise voluntary control when I bake this carrot cake insofar as I decided to do so (I chose to spent time baking instead of doing something else, I know what a carrot cake is, no one forced me to bake a cake, I perused several recipes and picked this one); I put myself in the position of doing so (I am in the kitchen, bought the ingredients, etc.); and I monitor every phase of the baking process, so that if the oven is too hot I lower the temperature, if my partner wants to eat all the nuts beforehand I stop him, and so forth. Finally, baking a carrot cake is under my voluntary control because I could stop doing it half-way through, if I so decided. *Therefore*, or so it seems, I am responsible for baking this cake.

Now, I think that the move from [1] to [2] is what makes the case of beliefs look puzzling. In the rest of this section, I argue that the tension that is typically seen between responsibility for

---

there is nothing they could ever have done to make themselves different than they are" (Kane 2002, 697); "If morality depends on luck, then at least sometimes people are judged morally for things that are beyond their control. This seems to be unfair; one does not deserve to be held responsible for what is beyond one's control" (Statman 1993, 2–3). For the idea that legal penalties should not be applied to someone unless they had the capacity and *a fair opportunity* to adjust their behaviour to the law, see the work of H. L. A. Hart, especially (Hart 1968), and the rich debate that followed. See also Brink and Nelkin (2013) and Nelkin (2020).

[7] Of the latter cases, sometimes I will also be responsible. For instance, if I hire someone to do something on my behalf. The point is that that's not *always* the case.

[8] In effect, this suggests that the proper conceptual connection holds between responsibility and *intentionality*, more than between responsibility and voluntariness. Rather than of 'voluntary control', it would probably be more accurate to talk of 'intentional control', or something along those lines. For insightful reflections on voluntariness, intentionality and choice, and their converse, see Hyman (2015).

beliefs and our impossibility of believing at will can be avoided once we accept a different reading of [1].[9] My argument here is indebted to Michael Frede's brilliant book *A free will: origins of the notion in ancient thought*.[10] Frede makes a case for the view that, though Aristotle is committed to [1], he would reject [2]. Of course, this is partly due to the fact that we have different theories of the mind, and I don't aim to suggest we accept Aristotle's view *tout court*, tripartition of the soul included. Yet, I believe that a short digression in the reception and development of the understanding of 'voluntariness' in antiquity will both shed new light on [2] and suggest how we can do without it.

In effect, I propose we tell the story backwards from [2] to [1], *i.e.* starting from the interpretation that is more familiar to us back to Aristotle's view. Frede argues that the notion of the 'will' as we understand it comes from the Stoics. For the Stoics, we have impressions (*phantasia*) of the things we perceive.[11] Each impression has a propositional content and a "coloring", depending on whether we perceive it as pleasant or unpleasant. According to their colouring, impressions can constitute impulses to go after the thing perceived or to avoid it; they are 'impulsive' (*hormētikai*) insofar as they constitute the *impulse* to act.[12] Adult human beings, in contrast with animals and children, are not moved by impulsive impressions *directly* because they also have reason, which scrutinizes impressions critically before they accept them as true and reliable. We are moved by an impression only once reason has *given its assent* to it. In order to give its assent, not only has reason to take the impression to be *true*; also, it has to accept its colouring; find its impulsive character appropriate.[13] Thus, for the Stoics motivation is always rational because, alongside with a passive element, *i.e.* a certain kind of impulsive impression, it always involves an active one, an *assent of reason* to that impulse. An action is "up to us" (*eph' hēmin*) if "its getting done is a matter of our giving assent to the corresponding impulsive impression" (Frede 2012, 45).

---

[9] There are other options. Smith and Hieronymi for instance maintain that [1] can be understood as [2] but deny that voluntariness is required for responsibility. That is, if we understand voluntariness in a narrow sense, then it cannot be a requirement on responsibility because not all the things for which we are responsible are voluntary in this sense. Hieronymi phrases this narrow sense like this: "an activity is *voluntary* just in case you decide to do it for reasons you take to settle the question of whether to do it, therein intend to do it, and, providing all goes well, do it by executing that intention" (Hieronymi 2008, 367). By contrast, I explore here the alternative option: we are responsible for the things we do voluntarily insofar as those are the things we do *hekontes*.

[10] For further discussion see Broadie (1991, ch. 3), Kahn (1988), Kenny (1979).

[11] Interestingly, for Chrysippus, even the impression itself reflects the fact that it is formed by a certain mind. Thus, humans and animals will have different the impressions, and so different people.

[12] Importantly, for the Stoics not every act of the will is a willing or a volition. That is, only if the will assent to 'an impulsive impression', the results will be a willing that leads (or aims to lead) to an action. When we choose to give assent to an ordinary nonimpulsive impression instead, like the impression that it will rain tomorrow, we believe that it will rain tomorrow. This belief too is an act of the will, for it results from assenting to an impression, but it is not a volition.

[13] This is why appetites and emotions are nothing but pathological affections produced by the mind. They are misguided beliefs. For instance, "fear…is nothing but the false belief that an evil is coming, or might come, one's way—a belief generated by assent to an impression which is deeply disturbing because one wrongly takes the situation to be an evil" (Frede 2012, 40).

In the *Discourses*, Epictetus specifies the idea further. He insists on the fact that no external action is entirely under our control. *Choosing* whether to φ is up to us but φ-ing itself is not. And so only giving assent to φ-ing is 'within our reach'. For instance, we can choose to cross the road but we might never make it to the other side. Choosing to cross the road is up to us; making it to the other side is not.[14] It is our disposition to assent (*prohairesis*) that defines us as the kind of person we are. That is our will. So, to reiterate, our will is not a disposition 'to choose to act' in a certain way but rather a disposition to choose to deal with our impressions in a certain way, *i.e.* assenting to them, rejecting them and so forth. As Frede glosses, "This indeed is the first time that we have any notion of a will" (Frede 2012, 46).

Along *these* lines, in order for me to exercise voluntary control over acting in a certain way, I need to have given my rational assent to it; I need to have deliberated that I want so to act.[15] More precisely, that I want to act in this way *rather than* in other ways. *Given* all the reasons that are available to me, I determine *which one* to follow. Through reason, I can decide what are for me the strongest reasons. I choose whether to φ *or not* to φ. This is the Stoic understanding of 'choice' and 'voluntariness' ('Stoic-choice' and 'Stoic-voluntariness' for short). I believe that Stoic-choice and Stoic-voluntariness inform the way in which 'voluntary control' is typically understood. I exercise voluntary control if I have given my assent to a certain motivation over another one; if I considered different options and chose which one to follow.

Now, if we accept that I need to exercise voluntary control in order to be responsible for something and that exercising voluntary control requires this kind of Stoic-choice—one between φ-ing *and not* φ-ing—then of course responsibility for beliefs is puzzling. For that kind of choice is unavailable to me, in believing. I cannot choose whether to believe or not to believe; nor can I choose which reasons for belief I want to follow.[16] In believing that *p*, I cannot peruse my reasons for *forming* the belief that *p* and *decide* which one is the strongest, or which one to follow.[17] Thus, the control challenge arises.

---

[14] For an authoritative analysis of tense and aspect of verbs see Vendler (1957).

[15] For an example of a strongly volitionist account, according to which an agent is responsible for something (an act, omission, attitude, and so on) just in case that agent has – directly or indirectly – *chosen* that thing, see Levy (2005). For an argument against the tendency of overemphasising the importance of 'deliberation', see Arpaly (2003). Arpaly's key thesis is that people act rationally without deliberation, and act irrationally with deliberation. Philosophers tend to over-emphasize deliberate agential control and autonomy at the expense of the emotional, dispositional, and affective aspects of their personae. By contrast, Arpaly argues that, to assess someone's moral worth, we need to verify whether their reasons for causing some action were indeed the right moral reasons, even if they happened to have been non deliberated upon, and how strongly motivating they were.

[16] On the basis that we lack the requisite sort of control required for our beliefs to be formed deliberately, David Owens argues against the possibility of epistemic *akrasia*: it's impossible for us to believe "either in accordance with our judgement about what we should believe or against those judgements" (Owens 2002, 395).

[17] I expand on this point (and the difference between theoretical and practical reasons) in chapter II, section 3.2.

But in Aristotle there's no such notion of 'the will', as a disposition to assent to impressions by reason, nor a similar understanding of what it is 'to choose'. This is largely because Aristotle allows for radically different forms of motivation, or desire. We have both nonrational and rational motivations, which originate in different capacities, abilities, or 'parts of the soul'. One kind of motivation is nonrational; these are the appetites (*epithymia*). For example, we are hungry irrespectively of what we think, even if we believe that it wouldn't be good to eat.[18] The other kind of motivation is rational, *boulēsis*: "a form of desire which is specific to reason. It is the form in which reason desires something" (Frede 2012, 20). Someone's desire of reason is "a direct function of their cognitive state, of what reason takes to be a good thing to do". Importantly, when Aristotle talks of choice (*prohairesis*), he is concerned exclusively with *rational* motivation. For instance, the virtuous person is the one who does the right thing *from choice* (*ek pro haireseōs*). We choose rational desires, we choose to follow them. But this doesn't mean that we have a choice to do *or not to* do so. Rather, we either choose the rational motivation *or we fail to do so*. This is the core of the Aristotelian understanding of 'choice' (henceforth, Aristotelian-choice). As Frede puts it, "the choice one makes in Aristotle is not, at least necessarily, a choice between doing X and not doing X, let alone a choice between doing X and doing Y. It is a matter of choosing to do X or failing to choose to do X, such that X does not get done" (Frede 2012, 29). If someone fails to follow reason and acts on a nonrational desire, it is not because one chooses *not* to follow the rational desire, or because one chooses to follow the nonrational desire. Rather, they fail to choose to follow a rational desire.[19] We act either on a desire of reason (which we chose), or on a nonrational desire.

> Reason is not made to appear in two roles, first as presenting its own case and then as adjudicating the conflict by making a decision or choice. How the conflict gets resolved is a matter of what happened in the past, perhaps the distant past. (Frede 2012, 24)[20]

On this view then, the virtuous person is the one who doesn't fail to act on rational desires, for which rational desires are always stronger reasons than appetites. So, I am not virtuous if I contemplate whether to punch a stranger who insulted me or sit down and drink a beer, *and I choose* the latter (by exercising control). Rather, I am virtuous if I choose (in the Aristotelian sense) to act on the rational desire *over failing* to do so. If I am such that the rational desire always motivates me.

---

[18] Notice that both rational and nonrational motivations can be reasonable or unreasonable. A nonrational desire is unreasonable if, for instance, I am hungry and yet right in believing that it wouldn't be good for me to eat. But if it would be good for me to eat, it is my rational motivation that is unreasonable.

[19] Put otherwise, in cases of *akrasia*, it's not that "we sit there anguished, tormented, torn apart by two conflicting desires which pull us in opposite directions, while we try to make up our mind which direction to take" (Frede 2012, 22). The contemporary debate is divided amongst those who think that weakness of the will isn't possible and those who think it is. The two positions can be traced back to Hare (1963) and Davidson (1969) respectively. For discussion see Walker (1989).

[20] For Aristotle, someone acts wrongly because they failed to train themselves to act on reasons rather than impulses.

Along these lines, when Aristotle says that praise and blame are appropriate only when it comes to *voluntary* actions, he is not saying that we are responsible for the actions that we carry out once we have given our rational assent to an impulse or desire, amongst other ones we have considered and weighed. Voluntariness is not characterised by a form of Stoic-choice. Rather, 'voluntary' and 'involuntary' translate *hekontes* and *akontes*. Things we do *akontes*, and for which we cannot be deemed responsible, are things we do because we are forced to, or because we act out of ignorance. For instance, Aristotle says, if we "were to be carried somewhere by a wind, or by men who had [us] in their power" (Aristotle, *EN* 3, 1110a3-4, transl. Ross). The meaning of *hekontes* is to be understood primarily against *this*: the things we do *hekontes* are those we do neither by force nor out of ignorance. These are the things we do voluntarily, willingly; for these, we are responsible. "Put positively", Frede glosses,

> for us to be responsible for what we do, our action has to somehow reflect our motivation.[21] We must have acted in this way, because in one way or another we were motivated to act in this way, that is, either by a rational desire or a nonrational desire or both. (Frede 2012, 25)

We are responsible for the things that we did because we were motivated so to act, because the source of our action was not outside us.[22] Any form of motivation to act is sufficient for the agent to be responsible for so acting.

Along these lines, we can provide an alternative reading of [1], the *hekon* reading of Aristotle's thesis:

> [2*] We are responsible for what we do according to the reasons we take ourselves to have; for what reflects our reasons.

In order for something to reflect our reasons, we don't need to choose whether we want to do *or not* to do it, or whether we want to follow this or that reason. There's no need for reason to browse through possible reasons, weigh them and decide *which one* is the strongest. An action is not "up to us" *(eph' hēmin)* if "its getting done is a matter of our giving assent to the corresponding impulsive impression". In other words, [2*] does not involve Stoic-choice.[23] Rather, something is up to us if whether it does or does not get done depends on us, rather than on someone or something else. This captures what we can call Aristotelian-voluntariness. Doing something *voluntarily* is doing it

[21] Frede talks of motivation because, like Aristotle, he's concerned with responsibility for actions. In formulating the *hekon* reading of Aristotle, I will talk instead more generally of one's *reasons*.
[22] *Cf.*: "What sort of acts, then, should be called forced? We answer that without qualification actions are so when the cause is in the external circumstances and the agent contributes nothing" (Aristotle, *EN* 3, 1110b1-2, transl. Ross).
[23] Though, to clarify, it does involve Aristotelian-choice in that, in order for me to count as doing the right thing of my own accord (*hekōn*), I need to *choose* it (given that it will be a matter of following a *rational* desire).

for one's own reasons and knowingly, and for this we can be responsible. Responsibility doesn't require that we voluntarily *control* something to be responsible for it, if 'control' involves anything *more* than this.[24]

2. *Actions and beliefs*

[2*] squares nicely with the aretaic model I've articulated in previous chapters. To begin with, to say that I am responsible for what I do voluntarily, and that I act voluntarily insofar as *I act in the light of my reasons*, means that I am responsible for what I do when I act intentionally.[25] As long as I act for (what I take to be sufficient) reasons, I am responsible for so acting. I am responsible because *I* am the agent, rather than someone else, or mere chance. In my terms, I am responsible for the actions that display my evaluative orientation. The reasons to which I am sensitive, and by which I am motivated, indicate what kind of person I am. In effect, this also explains why Aristotle claims that "choice is thought to be most closely bound up with virtue [*oikeiotaton*] and to discriminate characters better than actions do" (Aristotle, *EN* 3, 1111b6-7, transl. Ross). Someone's rational desires tell us more about who they are than their actions.

If this is on the right track, then it is not as surprising that we are responsible for our beliefs as well as for our actions. If I am responsible for what I do insofar as I act for the reasons I take myself to have, then we can say that I am responsible for what I believe insofar as I believe for what I take to be good reasons, *even though* the way in which I act for reasons differs from the way in which I believe for reasons. For of course practical and theoretical reasons do not function in the same way, as discussed in chapter II. But the marker of responsibility in the case of action is not that the reasons for which I act are *practical* in nature, and therefore *motivations*. This is what characterises the case of actions, it's distinctive to it. Rather, what makes voluntariness the marker of responsibility is that we act for reasons that are *ours* rather than someone else's. And if *this* is the criterion that identifies what we can be responsible for, then it can also apply to the case of beliefs. I am responsible for my beliefs insofar as it is *I* who believes them, for reasons that *I* take to be sufficient to support them. My beliefs reflect my reasons, although they don't reflect my

[24] Aquinas had no doubt that faith could be a virtue even though we are not able to choose what to believe (Aquinas 2010, part II–II, questions 1–5). For a discussion on religious belief and believing at will see Pojman (1986).

[25] Even though hereafter in the text I talk of acting and believing 'for reasons', notice that I can act intentionally also if I act for no reason in particular. See Anscombe (1963, §17), Hieronymi (2014, 13), and the debate on arational actions that followed Hursthouse (1991). In previous chapters, I make sense of this by emphasising that my evaluative orientation is displayed both by the reasons I find compelling and, more generally, by the way *in which I see things*. For instance, I said, I can learn how to play the violin for no particular reason and yet do it intentionally. Even though I don't have particular reason to set out to do so, it can still make manifest the way I see things, my evaluative orientation; and I can be responsible for it.

motivations. And I am responsible for what displays my evaluative orientation—the very way in which I see things.

In effect, if we accept [2*], then we can say that I believe *hekon* as long as I am not forced to believe something and I am not ignorant of what I believe.[26] However, as a matter of fact, it is unclear what situations *could* satisfy either of these conditions. First, can anyone *force me* to believe that *p* in the same way in which they can force me to do something? In general, we can say that I act under coercion when the conclusive reason for my so acting is that, if I do not do so, something bad will happen to me, and, if I so act, something good will happen to those who are coercing me. I am forced to do something when there's a threat of repercussion if I don't act in that way, and when I wouldn't act in that way if I weren't subjected to such a threat. I might act intentionally, and yet not for reasons that show the action worth doing for me.[27] If so, then it's unclear whether I can be forced into believing that *p* in a similar way. For I *cannot* believe that *p* unless I take *p* to be true. As it has often been highlighted, I cannot believe that *p* even if, weren't I to believe it, someone would punish (or reward) me. In this sense, I cannot believe that *p* for someone else's reasons. My reasons to believe that *p* cannot but be *mine*, and they cannot but bear on the truth of *p* (from *my* perspective). In effect, someone might say that I am forced to believe that *p* if I am presented with biased evidence. But even here, as long as I believe that *p* for reasons that I take to be sufficient, it is *my* belief. At most, I was forced *into forming* a certain belief[28]—but I was not coerced into believing itself.

Second, it is unclear whether I can *ignore* what I'm believing as I can ignore what I'm doing. The kind of ignorance that excuse me from responsibility for an action is the one at issue when I (non-culpably) lack certain information that would specify the description under which my action is intentional in such a way that, if I took that to be the description of my action, I would stop acting. For instance, I might intentionally set out to drink the content of the glass I see on my table, not knowing that it is transparent petrol; but if I knew that there's petrol in the glass, I would refrain from drinking it. In a sense, my action is the same, but it is intentional under one description, *i.e.* drinking water, and not intentional under another one, *i.e.* drinking petrol. By contrast, my belief does not admit of different descriptions in the same way. I cannot realise that when I believe that something is the case, *really* I am believing that something else is the case. That would equate to

[26] *Cf.*: "…there is a sense in which any expression of the will - in the absence of interfering factors - is as voluntary as any other. But in this sense, notice, an expression of open-mindedness (in belief) will be as voluntary as an expression or courage or kindness (in action)" (Montmarquet 2008, 387).
[27] Besides sparing me repercussions.
[28] I say more on this whilst considering the indirect control views below.

dropping the former belief and gain *another* one, or to revise the former substantially. I cannot believe that *p* and, when asked why I do so, respond 'Oh, I didn't know I was believing that *p*!'.[29]

If this is right, then *most of the time* we believe *hekontes*, and satisfy the conditions set in [2*]. Besides the cases in which I would mistakenly deny that I hold certain beliefs, in standard circumstances I cannot but be responsible for what I believe insofar as, when I believe, I believe for reasons that *I*–nothing nor anyone else—take to be conclusive.[30] Far from being the *hard* case, responsibility for beliefs seems to be the simplest and most basic one. Most of the times, beliefs do reflect our reasons, or, more generally, the way we see things.

By contrast, often actions fail to reflect our motivations. For one thing, we can act under coercion or ignorance. But also, as Epictetus was so careful to point out, things happen in the world that might prevent us from completing what we had set out to do: accidents, glitches, mistakes. It is because of this possibility of performance errors that we think of control as necessary, when we reflect on responsibility for actions. So it is no coincidence that the notion of the will as we typically think of it, which informs the control requirement as we know it, comes precisely from Epictetus. We *need* something like a control requirement when it comes to responsibility for actions because what we do can so easily fail to reflect our reasons. Put otherwise, we don't need the control requirement to make sense of the fact that I am responsible for crossing the road but rather to explain that I am responsible for crossing the road in a world that is populated by forces that might prevent me from doing so. Believing doesn't run the same risks. Beliefs don't risk failing to reflect our reasons.

In effect, I am happy to agree that some sort of control plays a role in responsibility *for actions*, namely a sort of 'monitoring control'.[31] For if I am responsible for what expresses my

---

[29] As I say below in the text, I resist the idea that the beliefs I acquire when I am under hypnosis count as beliefs I am forced to acquire. For it's unclear to me how they can count as 'beliefs' at all, given that I don't hold them for *my* reasons, and sometimes even *against* my reasons. However, some cases might offer a counterexample, *e.g.* implicit biases, if we accept that they are genuine beliefs. Suppose that you ask me why I believe that women aren't good at maths. I might respond 'I don't believe that is true'. If you give me evidence that I do, *e.g.* draw my attention to some of my actions, I might indeed say, 'Oh, I wasn't aware of holding that belief!'.
Drawing on the impossibility of remaining in the grip of a belief if one views the evidence for an opposing belief as decisive, Adler argues against the possibility of epistemic *akrasia* (Adler 2002, 8).

[30] Sometimes things are quite complicated to assess. For instance, suppose that I am aware of having been conditioned, growing up, in such a way that I fear being alone with someone of a different ethnicity. It might happen that some racist thoughts pop up in my head. First, it's not clear whether these have the status of *beliefs* at all. Yet, someone might think that I am responsible for them too. I lean toward saying that it depends on the circumstances (their intensity, frequency, my behaviour, my other background beliefs). According to how I behave and what I believe and fear more generally, I might or may not be a racist person. Put otherwise, that will depend on whether those thoughts reflect any of my values and of my take on things.

[31] To clarify, I think that the kind of *voluntary* control informed by Stoic-choice doesn't apply to intentions, for we cannot choose what to intend any more than we can choose what to believe. Also, perhaps it does not even apply to actions. So, whether Stoic-choice is a real phenomenon is a difficult question. Yet, answering it negatively would require more than what I offer here. My point is that, if it *is* a real phenomenon, it's not what discriminates between what we

reasons and there's the risk that my actions fail to do so, then I will have to *monitor*, in acting, that my ends are indeed being achieved. To exercise this monitoring control over an action is to ensure that it is aligned with one's intention. In order to saw Smith's plank, I will have to make sure that I saw just the plank and not also the stand on which it lies; that it's Smith's plank and not someone else's; that I cut it all the way through, and so forth. Thus, it involves that the agent is aware of the status and development of their action, and that they modify their behaviour if unexpected circumstances risk affecting its achievement. If I am baking a carrot cake, I need to control that it doesn't get burnt, so that I lower the temperature if my oven is too hot. Exercising this form of control does not require explicit awareness of control; it's not that I carry out an intentional action of monitoring my cake over and above the intentional action of baking it. Monitoring that my action gets done as I intend it is embedded in what it is to act intentionally.

The kind of monitoring control at issue is not akin to the voluntary control that I discussed above, the one informed by Stoic-choice and Stoic-voluntariness. For to make sure that things turn out in the world as I want them to, I need not exercise a rational capacity to decide which one of my reasons I want to follow. I need not *give my assent* to a reason over another one. Exercising control over my action, so that it reflects my motivation, does not require me to deliberate *over my reasons*. Thus, allowing for a form of monitoring control does not commit me to the Stoic-inspired understanding of control that, paired with the control reading of Aristotle's thesis, generates the control challenge.

Nor does it commit me to the control reading of Aristotle's thesis, more generally. For it doesn't entail that I cannot be responsible for something *unless* I exercise voluntary control over it. Some form of control might be necessary in the case of action to fulfil the condition set by the *hekon* reading of Aristotle's thesis but it isn't necessary to be responsible for something, in general. To put it another way: although it's true both that we cannot exercise voluntary control over our beliefs and that we cannot exercise monitoring control over them, only the first claim makes responsibility for beliefs seem problematic. In fact, the first claim means that it's *impossible* for us to exercise the relevant kind of control due to the very nature of beliefs and theoretical reasons. By contrast, the second claim is meant to capture the fact that exercising monitoring control over beliefs is *unnecessary*. We cannot monitor that our beliefs reflect our evaluative orientation because they immediately do so; they cannot but being held for the believer's own reasons. Thus, accepting some sort of monitoring control over actions does not generate problems for responsibility for beliefs.

---

are and are not responsible for. It cannot be a requirement on responsibility. Thus, the fact that we cannot exercise Stoic-choice over beliefs should not pose a problem for responsibility for beliefs.

3. Fairness and 'I cannot help it'

When I am forced to act, or act whilst ignoring important facts, it would be *unfair* to hold me accountable for what I do. In effect, it's hard to pin down precisely what is added when we say that an ascription of responsibility is unfair, over and above being inappropriate, or mistaken. I will address this question in further detail in chapter V. Very briefly, I would say here that unfairness involves an element of injustice. As such, it applies to a narrower set of cases. It would always be mistaken for you to deem me responsible for something I do whilst being forced or ignorant, but it would be unfair only if it returned an unjust portrait of who I am.[32] For instance, suppose that I really like reading, and that my friends and family knows it very well. A friend of mine insists that I read few pages of this horrifically cheesy contemporary romance. My sister, who knows nothing about literature, enters the room and registers that I am reading. She would take me to be doing it intentionally, and judge me on that basis, whilst I'd rather be doing something else instead. In cases like this, her assessment of me doesn't seem unfair, though it's mistaken. And I want to say, it is not unfair precisely because it does not ascribe to me values and a sensitivity to reasons that I don't have (nor does my sister ascribe them to me randomly, or by luck—she *knows* that I love reading!). By way of contrast, consider a case, suggested by Basu and Schroder (2019), in which you've struggled with alcohol problems for many years but have been sober for some months. When you return home, your spouse smells the wine that someone spilled on your shirt, and you can see from her eyes that that she thinks "you have fallen off of the wagon". In this case, your spouse's ascription is not only mistaken; more deeply, it mistakes your values. It portraits your evaluative orientation in the wrong way; it ascribes you vices that you do not in fact have. Your spouse's ascription is indeed unfair.

Now, the fact that it would be unfair to deem me responsible when I act under coercion or ignorance is often illuminated by the thought that *I couldn't help it*. If the fact that I couldn't help it is understood in a certain way, *i.e.* in accord with the *hekon* reading of Aristotle's thesis, then I agree that it wouldn't be fair to deem me responsible for it. Here, what I 'can't help doing' is something I do *akontos*; I'm forced to act *against my reasons*. That 'I cannot help it' stands for a proper

[32] Notice that sometimes it's unfair to think (falsely) that I did something because I was forced. It's easy to see this in the positive case (that is, when I am doing something positive): you unfairly think that someone must have forced me into cleaning the dishes, whilst I am doing it on my own initiative instead. But it can also happen in other cases. For instance, I am sixteen and my mum is convinced that someone must have forced me to start smoking. Although smoking is not a thing to be proud of, I might insist that *I* make my choices and consider it unfair that she casts doubts over that. On the aretaic model, the mother's beliefs are unfair because they don't return a faithful picture of the daughter's evaluative orientation. The view I'm opposing seems to have a harder time making sense of it: once you tie unfairness to the idea that you lacked control over something, it's hard to see why it's unfair to think that someone was forced into doing something whilst they are not.

excuse if it means that someone or something makes me act in a way that does not accord with my reasons, motivations, desires. Similarly, that 'I can't do otherwise' makes it unfair to deem me responsible only if it means that someone or something successfully prevents me from doing something else, perhaps what *I* would intend to do.[33] And if we want to say that 'I lack control' in these cases, that must refer to the sort of monitoring control that guarantees that my actions reflect my motivations. If I lack the capacity of ensuring that what gets done in the world is what I was doing intentionally, then what gets done in the world is likely not to reflect my intentions, and it would be unfair to deem me responsible for it.

But that 'I cannot help it' and that 'I cannot do otherwise' are ambiguous phrases, and often they are taken to mean that I cannot be moved by different reasons. In this sense, it's said that I couldn't help reading your diary, eating the last piece of cake, eavesdropping. Understood in this way, I want to insist, they do *not* excuse me. For here I'm not forced to act 'against my reasons'; I don't act *akontos*. I act on weak reasons, wrong reasons, or reasons that I wouldn't endorse upon reflection, but they are still mine insofar as they moved me, at least at that point and time.

The fact that I cannot help doing something in this second, problematic, sense is often taken to mean that 'I cannot control' it. The control I lack when I cannot be moved by different reasons is the Stoic-inspired voluntary control I delineated in sections 1 and 2. But crucially, I want to insist, lack of *this* kind of control does *not* excuse me from responsibility. That I cannot choose my reasons, or by which reasons to be moved, doesn't entail that I am not responsible for acting according to them. Thus, voluntary control cannot be the kind of control required for responsibility, for I can lack the possibility of exercising *it*, and still count as acting on my reasons. I can lack that kind of control and still be responsible for what I do, and fairly deemed so.

Rather, I think that it's unfair to criticise someone for something that doesn't express their distinctive sensitivity to reasons. Ascriptions of responsibility are fair as long as they target what displays someone's evaluative orientation. Along these lines, it might indeed be unfair to deem me responsible for beliefs that are just the result of hypnosis,[34] which I wouldn't hold otherwise. For they do not reflect *my* reasons; they do not display my evaluative orientation. I am forced to believe them in that I might hold them even *against* my reasons.

---

[33] I think that this is essentially in line with Frankfurt's well-known argument that we may well be morally responsible for what we've done even though we could not have done otherwise (H. G. Frankfurt 1969). In fact, I take the core of Frankfurt's insight to be that what matters is someone's *choosing* their reasons, in the Aristotelian sense; what someone sets out to do, and the reasons on which they act, matters more than their possibility of doing otherwise. Even if they could have not done otherwise, they are open to assessments for what they *in fact* choose to do.
For some disagreement on this point see for instance Peels (2017), who takes the fact that the agent could not have not believed that *p* as excusing them from responsibility for that belief. Peels concludes that epistemically justified belief and responsible belief are distinct, precisely because only the latter requires the capacity of believing otherwise.
[34] If beliefs that I acquire after being hypnotized can be treated as my *beliefs* at all.

By contrast, I am fully responsible, and it's not unfair to think that I am, for most of the beliefs that I cannot help having. For instance, suppose that, after careful reflection, I conclude that you betrayed me. Since you ask me to reassess my opinion, I might consider the issue again and again. Yet, I always reach the same conclusion. As much as I'd like to believe otherwise, I cannot help but believing that you did betray me; I am fully convinced that it happened. Clearly, this kind of incapacity to believe otherwise doesn't make ascriptions of responsibility illegitimate.[35] Quite the opposite, it is for these unshakeable convictions that result from careful consideration that I would be most obviously and legitimately open to assessment.[36] For they are not 'against my reasons' in the slightest sense.

This way of understanding fairness also makes sense of another aspect of our practices. Consider an example from Sher in *Blame for traits*. Suppose that you are "so selfish that the suffering of others barely registers on [your] consciousness" (Sher 2001, 154), or that you are "so corrupt that [you see] no point in being honest", whilst I see "the point of being honest perfectly well, but…[this time] give in to a dishonest impulse".[37] If we understand 'control' and 'couldn't help it' as I think we should *not*, then it might result that only I demonstrate some control over my action insofar as, at the very least, I could have chosen to follow another reason. Another reason was available for me to choose. By contrast, it's utterly impossible for you to see things otherwise; you cannot help but being dishonest and acting in dishonest ways. Thus, on this view, it would be *unfair* to hold you responsible and criticise you for something that you could have not avoided, and it would be fair to criticise me—at least more than it is to criticise you.

---

[35] Whether the will itself is free is an extremely broad and difficult question, which I cannot investigate here. However, it's important to mention that it has been often stressed that also the incapacity to *will* otherwise doesn't make ascriptions of responsibility illegitimate. A classic example is Luther, who has famously declared "Here I stand;/can do no other". Frankfurt comments: "It is clear, of course, that the impossibility to which Luther referred was a matter neither of logical nor of causal necessity. …What he was unable to muster was not the *power* to forbear, but the *will* . …the person in a situation of this kind generally does not construe the fact that he is subject to volitional necessity as entailing that he is passive at all" (H. Frankfurt 1982, 264), and "When we accede to being moved by logic or by love, the feeling with which we do so is not ordinarily one of dispirited impotence. On the contrary, we characteristically experience in both cases- whether we are following reason or following our hearts- a sense of liberation and of enhancement" (H. Frankfurt 1982, 267). See also Williams (1993; 1990); according to Williams, one can determine, also through deliberation, that one can or cannot do certain things and these "incapacities can not only set limits to character and provide conditions of it, but can also partly constitute its substance" (Williams 1981, 130). For an insightful discussion on the differences between Frankfurt and Williams on volitional necessities, see Watson (2004).

[36] Hieronymi makes a similar point. She argues that, if I reconsider a question and yet find myself unable to come to a different answer, this is not a manifestation of a limit — unlike my inability to modify by bone structure, for example, which instead manifests a limit of the kind of being I am. Rather, on Hieronymi's view, my inability to change my answer is a *manifestation* of my ability to answer questions. It simply proves that I am *deeply* convinced of my answer.

[37] In the original, Sher says "but [decide]—perhaps after a struggle—to give in". I expunged this for it assumes the Stoic understanding of 'choice'. In effect, this also entails that Sher and I give different diagnoses of the case: he insists that I am not to criticise because I failed to deliberate correctly; I lost in the game of weighing reasons correctly.

But this is clearly not how we would react.[38] It seems to me that we would criticise *more* the person who's utterly dishonest. At the very least, it would be unfair to be *harsher* with me than with you. And I think that this is so because the fact that you cannot see the right reasons does not excuse you. For it is not that, because of *that* incapacity, you end up doing or believing something *against your reasons*. It is not that 'you cannot help it' in the relevant sense. In effect, I am not excused either, since I do act and believe on my own reasons. It's not unfair to criticise me for having been dishonest in this occasion. But it would be unfair to treat me as someone who's *always* dishonest, like you. As Sher puts it,

> Indeed, if anything, many would actually consider it *fairer* to blame the thoroughly corrupt person, whose bad decision requires no thought, than the merely imperfect agent whose transgression represents a surrender or a falling away.[39] (Sher 2001, 154)

4. *Controlling beliefs*

In sections 1-3, I have argued that, by rejecting the control reading of Aristotle's thesis, we can dispel the puzzlement that arises around responsibility for beliefs. We are responsible for both our actions and our beliefs insofar as we act and believe *hekontes*. If my arguments are convincing, then the control challenge would cease to be a pressing issue.

Yet, the control challenge is often *addressed*, in fact. Many authors accept the control reading of Aristotle's thesis and attempt to accommodate the possibility of responsibility for beliefs by qualifying the control requirement one way or the other. There are good reasons to look at some of these proposals. First, someone might simply find my arguments against the control reading unconvincing. Second, if the most popular ways of addressing the control challenge are found wanting, then I will have provided one more reason to doubt that its core question is well-posed. In this section, I consider two clusters of views: the indirect voluntary control views and the direct nonvoluntary control views. Although not all the proposals available in the literature would fit one or the other set,[40] I take them to be the two most plausible and influential approaches.

---

[38] Some disagree. For instance, McCormick and Schleifer claim that it would be unfair to blame someone less for the occasional lapse than someone who often fails to respond to reason (McCormick and Schleifer 2006). First, I think this depends on a monolithic understanding of evaluative orientation according to which, if I am kind, I will have to have always and only kind beliefs. McCormick and Schleifer do not conceive of the possibility that I might be someone who's kind and unkind, in different contexts. I argued against this understanding of evaluative orientation in chapter II. Second, they seem to think that, on views like mine, someone's 'lapse' won't be criticised, and this is unfair. But as I have argued in the text, my unkind belief will indeed be taken to manifest unkindness. It won't be underestimated as an anomaly of little importance. But the point is that, if I am criticised, it is not because I *could have controlled or avoided* something. I will be criticised insofar as I display a lack of kindness.

[39] He continues "Yet once we have acknowledged that it is fair to blame people for bad acts that flow automatically from bad traits over which they lack control, we can hardly deny that it is fair to blame them for those bad traits themselves".

[40] Another cluster of views is referred to as *doxastic voluntarism*. Its proponents accept the control requirement and argue that we are responsible for our beliefs insofar as what we believe can be "up to us", at least sometimes. See especially

4.1. Indirect voluntary control views

In 1985, Robert Adams writes:

> Amongst philosophers interested in the topic of this paper, probably the commonest opinion has been that we are accountable for wrong states of mind only insofar as they are consequences (that should have been anticipated) of our own voluntary acts and omissions in the past—and thus only insofar as they were indirectly within our voluntary control. (Adams 1985, 12)

Even though it is not obvious that the indirect control view is the 'commonest' today, many do deem it the best alternative. Consider for instance:

> The purpose of this paper is to explore what underlies the legitimacy of this practice—the praise or blame of people for what they know or fail to know as a result of investigation or otherwise—namely, the ability to exercise control over one's doxastic states, and, in particular, as I will argue, one's ability to exercise indirect generic control over one's doxastic states. (Meylan 2015, abstract)

> …we can talk of praise and blame for particular doxastic tokens without needing to "outrun" voluntary control, since we can base such judgements on our indirect doxastic control or indirect doxastic influence. (Booth and Peels 2010, 259)

In general, the core idea is that we control our beliefs indirectly by exercising voluntary control over the activities that constitute their formation and revision. We are responsible for our beliefs insofar as we exercise indirect control over them. This view seems to accommodate nicely the cases in which we criticise someone because they 'could have put more effort' in reflecting upon something, or because they 'overlooked pieces of evidence that they could have easily noticed'.

Depending on what one takes to be the relevant activities over which we exercise direct voluntary control, there can be different versions of the indirect control views. According to some, the believer directly controls the activities that constitute an investigation on a certain topic, on which they set out to form an opinion.[41] We want to know whether *p* is true, false, or it is impossible

---

Ginet (2001), Steup (2000; 2008; 2011; 2012), Weatherson (2008). For similar views, see Shah (2002); Russell (2001); Ryan (2003). See also Montmarquet (1986), though he retracts his position in Montmarquet (1987). For a defence of *negative* doxastic voluntarism, see Rott (2017). For criticism see Booth (2014), Buckareff (2014), Feldman (2019), Nottelmann (2006), Pojman (1986), Scott-Kakures (1994), Williams (1970).

Others deny that we are responsible for our beliefs on the basis that we cannot control them. For instance, Levy (2007a) argues that the existing arguments to accommodate the control challenge are all wanting and that our lack of control over our beliefs typically excuses us of responsibility for them.

Richard Feldman proposes yet another way (Feldman 2000). He rebuts the control challenge by denying that oughts imply can; epistemic oughts are to be understood as *role oughts* which hold despite one's incapacity to fulfil them.

[41] See Meylan (2015; 2017), Nottelmann (2007), Peels (2017), Price (1954). See also Audi (2001, 2008) for the suggestion that we should accept an epistemic obligation to be attentive to one's evidence, from where we will be led to an ethics of inquiry, and Chrisman (2018), though the latter distances himself from accounts that view doxastic agency as being

for us to say, and we put ourselves in the best possible position from where to decide whether *p*. We intentionally gather evidence, ask questions, research, and so forth. We can call this the *investigation account*. On the other side of the spectrum, some claim that the believer can control directly the process of forming one particular belief.[42] We want to believe that *p*, perhaps that God exists, and we make sure that we will form it by exposing ourselves to just and only the relevant pieces of evidence. Call this version the *exposure account*. Other proposals fit somewhere in the middle. For instance, some have argued that we can control what beliefs we gain by selecting the sources of information to which we expose ourselves, *e.g.* by getting information from one particular newspaper or TV channel.[43] Here is it not a particular belief we're going after. Yet, we do want to make sure that our opinion is aligned to a set of somehow related propositions, *e.g.* a certain ideology. We can refer to this version as the *inclination account*.

In the light of the arguments I developed in chapters I-III, in general I think that these proposals miss the core of the phenomenon of responsibility for beliefs, since they are not concerned with the particular considerations someone takes to be conclusive to believe something. Rather, the advocates of this line typically end up shifting the focus from the believer's responsibility for *holding* beliefs to their responsibility for the activities of belief formation and revision. In effect, the former is reduced onto the latter. For instance, Meylan writes:

> It seems, then, that we sometimes praise or blame people for what they believe or fail to believe, for what they know or fail to know, etc. as a result of their having investigated the evidence, their having failed to investigate the evidence, or their having neglected some evidence, etc. Cast differently, we praise or blame people for—what I shall

too *external* to belief. Chrisman's central thesis is that epistemic agency is involved in the activity of maintaining a belief-system (as well as in deliberation). Alston (1988) argues that we can, at most, direct our attention to certain evidence but that, once we do so, the evidence will determine what we believe. So he agrees with the advocates of the investigation account but denies that indirect control is sufficient to responsibility. Kornblith (1983) argues that someone's belief is justified if they've done all they should have to bring it about that they have true beliefs.

[42] *Cf.* Pascal (1905, 188): "Labour then to convince yourself ... Learn of those who have been bound as you are, but who now stake all that they possess; these are they who know the way you would follow, who are cured of a disease of which you would be cured. Follow the way by which they began, by making believe that they believed, taking the holy water, having masses said, etc. Thus you will naturally be brought to believe". More recently, Naylor (1985) argues for a similar view: "*S* believes voluntarily that *p* if and only if (i) S chose to believe that *p*, and (ii) Because he chose to believe that *p S* brought himself about, using means he himself chose to use, that he regards *p* as being conclusively evident or as being self-evident" (Naylor 1985, 432).

[43] Levy and Mandelbaum (2014). See also Cohen (2013) for an account of Kantian inspiration. According to Cohen, we have two forms of indirect doxastic control that are sufficient to ground epistemic responsibility: first, the capacity to judge and doubt; second, the ability to choose our epistemic maxims. I take the latter idea to indicate that Cohen's proposal can be considered a version of the inclination accounts, in my terms. On similar lines, Heller says we are responsible for beliefs if we form and assess our beliefs in accordance with our self-endorsed belief-forming dispositions (Heller 2000, 130–41). Even though he presents his proposal as a development of Hobart's idea that what we are really judging when we blame S for a bad act is S's *character*, Heller ends up building an indirect control view, according to which S's epistemic *nature* is constituted by her *desires* to form beliefs in accordance with certain dispositions rather than others.

> globally name for sake of brevity—*the doxastic states, in which they are as a result of their investigative behavior*. (Meylan 2015, 1–2) [my italics]

On this view, someone is responsible for their beliefs *insofar as they are the result* of their 'investigative behaviour'. Thus, there's no room to discuss the very quality of someone's reasons and evaluative orientation. Nor to explain why we can be criticised or commended also for holding a certain belief, irrespectively of how we've formed it.[44]

The point I want to make here, however, is another. Even if we grant that the control requirement imposes a valid condition and accept to take on the control challenge, it's unclear to me how *any* form of *indirect* control can satisfy it. Let me elaborate. In introducing the control challenge, I said that the control requirement is usually taken to be at odds with the fact that 'we cannot believe at will'. Thus, I suggested, the kind of control involved in the control requirement must be a form of voluntary control understood along the lines of Stoic-voluntariness and Stoic-choice, for *this* is what is unavailable to us when we believe. The thought is that, in order to be able to exercise this kind of control over an action, I need to have deliberated on my reasons, determined which one to follow, and acted upon it. I control my action voluntarily in this sense insofar as, by deliberating, I choose to do it over doing something else, and I act accordingly. However, I cannot do *this* when it comes to beliefs. I cannot deliberate on what reasons to believe; I don't peruse reasons and give my assent to what I take to be the best one. Thus, I cannot exercise the same kind of voluntary control over my beliefs.[45] The question is, if so, how can I be responsible for them? The indirect control views purport to offer a solution to *this* question. Yet, it's not obvious that *indirect* control can in fact offer an answer to it.

Consider first the investigation account. Intuitively, this is the most plausible proposal, at least insofar as it aims to cover the largest pool of cases. In fact, I cannot self-manipulate into believing specific propositions every time I form a belief but I can exercise control on a regular basis over the ways in which I form my beliefs. Yet, things get murkier at a closer look. On this view, I decide which topic is worth having beliefs about; or more specifically, the truth of what proposition to examine. So, I exercise the relevant kind of voluntary control in deliberating whether to form an opinion on a certain subject matter. But it's hard to say in what sense I exercise 'indirect' control over the resulting belief. Beliefs seem to be treated like the haircut I would get if I entered

[44] As Owens puts it, "the possession of the virtue (and the absence of the vice) has an ethical significance which is quite independent of the (ethical) feasibility of the actions required to cultivate it. So even if the virtue can be the product of voluntary action, it doesn't follow that our responsibility for our moral character is a mere by-product of our responsibility for our actions" (Owens 2000, 119).

[45] See McHugh (2011b) for the view that judging is not voluntary because it has truth as a constitutive goal, and Hieronymi (2008) for the view that the feature which renders an attitude a fundamental object of responsibility (that the attitude embodies one's take on the world) is also what guarantees that it could not be voluntary.

a salon and asked the hairdresser to follow their inspiration. Of course, I myself decided to have my hair cut, entered the salon and made my requests controllingly. However, given that I just asked for *a* haircut and not for *any* in particular, it's unclear in what sense I exercise indirect *control* over my getting a blond, curly or shaved head. At most, I exercise indirect control over getting *a haircut*. But I wouldn't control becoming *blond*, however indirectly.[46] Similarly, I might be said to exercise direct control over forming a belief on a certain topic and indirect control over *a belief that I have on a certain topic*. But I wouldn't exercise control over *this* particular belief. So, if control is a requirement on responsibility, I am responsible for my belief in the sense that I am responsible for having a belief concerning a certain proposition, but I am not responsible for having this particular belief. I am not responsible for taking this particular proposition to be true or false (or difficult to assess).

The exposure and the inclination accounts are in a better position to explain in what sense we are in control of the beliefs we hold *themselves*, though indirectly. Beliefs here are not treated like a haircut I get blindly as much as like a cake I bake following a recipe. I decide to follow this recipe rather than doing other things, I know that following this recipe will issue this particular result, and I make sure that nothing goes wrong in the meantime. The cake is the result of what I directly control. But given that my activities guarantee it as a result, and I know it, I can be said to exercise indirect control over my cake itself. Similarly, I can select what reasons to expose myself to. Although I cannot choose which reasons to follow *directly*, given the nature of beliefs, I can decide what reasons to have myself follow. In a sense then, I do decide what reasons to follow. I just cannot do so directly.

An initial problem here is that it's not obvious that exposing ourselves to certain evidence will issue the targeted belief as following a recipe religiously guarantees the expected outcome. The success rate of believing what we manipulate ourselves into believing remains at the mercy of the strength of the evidence considered. Thus, we would count as actually exercising indirect control over a belief only if the evidence to which we expose ourselves are sufficiently strong. But we have no control over *that*. Deciding to follow certain theoretical reasons, *e.g.* by manipulating oneself into considering only some, might not be sufficient to forming the corresponding belief.

Furthermore, in order to expose myself to certain evidence or sources that would lead me to believe that *p*, it seems that I would have to *already* believe that *p* is true (on the exposure account) or at least that something in its neighbourhood is (on the inclination account). Conversely, I would avoid the evidence and sources that would likely contradict what I take to be true. But then I would not exercise indirect control 'over a belief I form', because I would hold that belief *already*. I

[46] The proponent of the investigation account could then say that I can be responsible for getting a silly haircut, in virtue of issuing silly instructions. In effect, this seems to be the very intuition that fuels their proposal. Yet, it'd remain unclear in what sense I can be said to exercise indirect *control* over getting a silly haircut.

wouldn't be forming it through those activities of self-manipulation; it would be *necessary* to direct the self-manipulative process itself.

Even if someone resisted that the very same belief orients, and thus has to be preliminary to, the activities that constitute my exposure to certain evidence, there would still remain the risk of a regress.[47] In fact, suppose that I self-manipulate myself into believing that *p*. In order to do so, I would need to hold the belief that, minimally, I need to select this and this evidence. Even if we accept that this latter belief need not coincide with the belief that *p* itself, am I responsible for it? And how can we explain this? If we say that we need to exercise indirect control over *that* belief, then we fall into a regress. If we say that we are responsible for it despite lacking indirect control over it, then it's unclear why we need indirect control for other cases. And if we say that we are not responsible for it, then it's unclear how we can be responsible for a belief that depends on the activities that, in turn, are guided by a belief for which we are ultimately *not* responsible.

At this point, the advocate of the exposure and inclination accounts might suggest that we decide to which evidence to expose ourselves according to the *attractiveness* of a certain belief, rather than according to its presumptive truth or falsehood. I need not already believe that *p*, nor anything related to it. Rather, I manipulate myself into believing that *p* for reasons that are pragmatic rather than epistemic; I want to form beliefs that I wish to have.[48] Perhaps, sometimes we might be able to manipulate ourselves in this way. And here, we would count as having some sort of control over what theoretical reasons to follow. However, this cannot be considered the standard scenario. First, it would be incredibly costly always to monitor our belief-formation processes, select and manipulate our evidence. Second, we are hardly the main providers of the evidence that support the beliefs we end up believing. And if this were the standard case, the very nature of beliefs would be considerably different from how we typically think of it.

Along these lines, the indirect control views seem unable to explain how we control our beliefs in such a way that we are responsible for them. If we understand 'indirect control' in a thin way, it's unclear in what sense it is a form of control over our beliefs themselves. If we make sure that it is exercised over the beliefs themselves, then it becomes difficult to show it feasible.

---

[47] As Montmarquet puts it, "Almost without exception, such actions - or what is more common, failures to act - as might ground culpable belief-states are themselves only blameworthy inasmuch as they themselves are grounded in culpable belief states" (J. A. Montmarquet 2008, 383).

[48] Sometimes, we might want to believe something to the point that we modify the world accordingly, *e.g.* when we form the belief that the light is on by switching it on. See Feldman (2000, 671–672), and Kelly (2003, 169–170). But this can hardly capture the kind of beliefs we have in mind when we set out to investigate responsibility for beliefs.

4.2. Direct nonvoluntary control views

The second cluster of proposals I want to consider are the direct nonvoluntary control views. The nonvoluntary control views deny that indirect control is sufficient to ground full-blown responsibility for beliefs. However, they don't dismiss the control challenge. They address it by claiming that we exercise direct nonvoluntary control over our beliefs.[49] Even though I cannot believe at will, my beliefs are in my control in some other—more fundamental—way. Because of this, it's possible for me to be responsible for them.

> We enjoy this form of attitudinal control not only over our beliefs, but also over other attitudes, including intentions—that is, over the will itself. Closely tied to our capacity for reasoning, attitudinal control is in important respects more fundamental than the voluntary control that we exercise over our actions. (McHugh 2017, abstract)

> … we seem to be responsible for our beliefs in a rather direct way. We seem to be responsible for our beliefs, not merely as features of ourselves that we can manage or affect through our actions (not merely in the way we are responsible for, say, our health) (Hieronymi 2008, 357)

> … if we can be responsible for things because we can affect and control them through our actions, our responsibility for and control over our actions must be explained in some other way.[50] … There must be some more fundamental or original way in which we are responsible for our intentional actions. (Hieronymi 2014, 16)

Here I focus on McHugh's view, and I discuss Hieronymi's in section 5. McHugh understands nonvoluntary control as *attitudinal* or *guidance* control. The core idea is that we exercise guidance control over a belief if "we revise [it] directly for reasons", *i.e.* if it "is formed or sustained by a mechanism that is the believer's own, and that is receptive and reactive to epistemic reasons"

[49] See for instance Boyle (2009; 2011), Hieronymi (2006; 2008; 2009; 2014), McHugh (2011b; 2012; 2013; 2017); Raz (2011 esp. ch. 5 and part III), Smith (2005; 2008; 2015).

[50] Similarly, McHugh warns against the risk of regress here: "Indeed, if we were not responsible for our intentions, it's hard to see how we could be responsible for the actions that constitute their execution" (McHugh 2017, 2749).

(McHugh 2017, 2751). Someone is receptive[51] and reactive[52] to epistemic reasons if they are able to identify and respond to them appropriately, respectively. The notion of 'reason-responsiveness' is thus understood counterfactually: someone is responsive to reasons if and only if, sufficiently often in a relevant range of counterfactuals, they would respond appropriately to them. McHugh claims that the dispositional nature of guidance control can explain in what sense it is a form of *direct* control: it is direct because, as soon as our reasons change, so do our beliefs. We do not respond to modifications in our reasons by engaging in further intentional activities, *e.g.* by forming intentions to revise beliefs and then executing them. Rather, we modify them directly.

To begin with, it might be puzzling to understand reasons-*responsiveness* as a form of 'control' at all. McHugh motivates his commitment to the control requirement by reference to fairness and the importance of what we can and cannot help doing and believing. Responsiveness to reasons *has to* identify a form of direct *control* because, if we couldn't control and couldn't help believing them, it would not be possible for us to be responsible for them.

> Something that reflects your character might be grounds for evaluating you as a good or bad person, say, but that does not make you responsible for it. An agent who is unable to empathise is evaluable as a bad person. But she is not responsible for her failure to empathise. After all, *she can't help it*. (McHugh 2017, 2747) [my italics]

[51] McHugh borrows the notion of reasons-responsiveness, as consisting of reasons-receptivity and -reactivity, from the analysis of responsibility *for actions* put forth by Fischer and Ravizza (Fischer and Ravizza 1998). It's unclear whether the framework can apply as easily to beliefs. For instance, whilst it is perfectly plausible to imagine an agent failing to act upon the reasons they recognise to be sufficient to act (*e.g.* because of bodily constraints, akrasia or compulsion), it is less clear what it would mean for a believer to fail to react to reasons that they take to be sufficient to believe that *p*. The case of actions does not raise difficulties because there is a clear distinction between recognising (sufficient) reasons and performing an action. Acknowledging that you have sufficient moral or prudential reasons to act is not *thereby* to act. You have to do something over and above recognising reasons, *i.e.* intentionally act. By contrast, acknowledging that you have sufficient reasons to believe that *p* is just what it is to believe that *p*. There is nothing more you are supposed to do, or infer. The problem is that failure of reasons-reactivity is a failure of performance, but there is no room for performance errors in believing. For a formulation of the "No Gap Principle" see Raz (2011, 38). For some resistance see for instance Heil: "Admittedly, it's a short step from 'S believes that (the belief that) *p* is (or would be) warranted' to 'S believes that *p*', but it *is* a step" (Heil 1983, 359).

[52] There are at least two ways to make sense of the claim that paranoid delusions are not *receptive* to epistemic reasons. First, it could mean that what I take to support my delusion is not actually happening. For instance, I am convinced that someone follows me but nothing that I see could support my conviction: no one has been behind me for a prolonged period of time or is looking at me insistently. Second, it could be that what I take to support my delusion is in fact happening but does not offer sufficient support to my paranoia. Someone is indeed walking behind me, but that's not enough for me to believe that they are *following* me. It's unclear which of these two failures McHugh has in mind. One reading has to do with one's inability to distinguish truth from falsity, whereas the other one with one's ability to calibrate reasons. Strikingly, in both cases believers would form and maintain beliefs on false bases. But if this is why they fail to count as receptive to reasons, then it seems that all those who believe something on the basis of false considerations would fail to be receptive to reasons, and thereby not in the position to be responsible for their beliefs. McHugh resists this worry stressing the distinction between the lack and the failure of reasons-responsiveness. I expand on this distinction below in the text.

> It seems to me that it would *not* be legitimate to hold the psychopath responsible in this way. After all, moral demands not only do not, but *can* not, have any grip on them. The psychopath can't help it. (McHugh 2013, 138)

First, notice that McHugh is employing the phrase 'cannot help it' in the way I tried to show misleading in section 3. On this view, someone is excused from ascriptions of responsibility if *they cannot be moved by different reasons*. Insofar as they cannot see things differently, they cannot be deemed responsible for their conduct. Even setting this aside, the point here is that reference to control seems unable to play the explanatory role that it's meant to play. In fact, the importance of the phrase 'I cannot help it' is introduced to make sense of the difference between the paranoid person, who's not responsible for their beliefs because they *lack* the capacity for reasons-responsiveness, and us, who are criticisable or commendable insofar as we possess that capacity, though it *fails* us sometimes. The paranoid person cannot help it, whilst we could. The paranoid person cannot be deemed responsible for their beliefs, whilst we can.

But it's difficult to see how the fact that 'someone cannot help it' provides a criterion to distinguish between 'lack' and 'failure' of reasons-responsiveness; how it marks the difference between paranoid and non-pathological believers. As a matter of fact, also stubborn, close-minded, arrogant, dogmatic, hot-headed or dumb persons are typically unable to change their mind and see different reasons; they could easily say that they cannot help holding certain beliefs and, in most cases, it would be true. They are reluctant to revise their commitments even when presented with strong counterevidence. If so, then it either becomes too easy to escape responsibility, or this criterion is insufficient to show that ascriptions of responsibility are inappropriate.

At this point, someone might stress that the difference between the pathological and non-pathological believers lies in the fact that the latter could believe otherwise, *if only they wanted to*, or *if only they tried*. In effect, McHugh seems to suggest something along these lines whilst fleshing out the nature of guidance control, which we non-pathological believers exercise over our beliefs. The beliefs that we form automatically or "that are not being actively consciously controlled are nevertheless normally under a form of dispositional conscious control on the part of the agent" (McHugh 2013, 135).[53] To clarify the point, McHugh adds, "we *would* endorse [the beliefs that we form automatically] in judgement, if we considered their contents. And our beliefs are normally *susceptible* to control by acts at the level of judgement and inquiry" (McHugh 2013, 134). A belief that *p* formed below the level of consciousness is one that I *would have formed* if I had engaged in an

[53] He argues for the claim that "responding to a putative reason involves conscious agency, at least dispositionally" as follows. First, recognising a reason involves actual or potential consciousness; it is to be aware of it and of its probative force. Second, reacting to a reason involves agential control. One cannot count as reacting to a reason and as doing or believing something for a reason unless it is open to her the possibility of intervening if needed or considering its content and endorse it in judgement.

inquiry aimed at determining whether *p*. And when there is a discrepancy between the belief I formed automatically and the one I would consciously endorse, I would revise the former if only I reflected upon the matter.

If so, then when we say that the dogmatic person, but not the psychopath, failed to exercise their capacity properly and 'could' have believed otherwise, we are saying that they would revise their beliefs if they "considered their contents". Two thoughts here. First, this seems too charitable. Many stubborn, arrogant, or dogmatic persons would *not* revise their beliefs even if they attended to their content consciously—even if 'they tried'. And their beliefs would not be modified according to the change of evidence, in a wide range of situations. In effect, that is precisely what makes them stubborn, arrogant or dogmatic.

Second, and more importantly, this way of fleshing out the sense in which guidance control is dispositional might weaken the *directness* of it. Consider again McHugh's claim that "our beliefs are normally *susceptible* to control by acts at the level of judgment and inquiry". The idea here is that, although I do not control my beliefs "consciously", I *could* always do so by inquiring further into the issue and considering things again. Inquiring and judging are activities that I can control voluntarily. But then, whether or not I actually revise my belief depends on whether I perform well the activities that constitute belief-formation and revision. And if so, then exercising 'direct' nonvoluntary control would actually depend on exercising *indirect voluntary* control.

Finally, let me mention that views like this seem also to miss the core of the phenomenon we are trying to illuminate. For if responsibility emerges on *dispositional* control, then it doesn't really matter precisely what reasons I take to be conclusive as much as that I *would* consider other ones and revise my beliefs *if* the circumstances were different. However, in many cases precisely what we take to be true is extremely important, when we evaluate each other. And that is so even regardless of what we *could* take to be true, were things different. In other words, *both* indirect control views and this kind of nonvoluntary control views tend to reduce responsibility for beliefs onto responsibility for the activities that constitute their formation and revision—by either focusing on investigation or on a properly functioning 'mechanism' of belief-revision. In so doing, they fail to see the importance of the fact that believers are responsive to *certain* reasons and responsible for believing that *p* itself.

5. Evaluative control

In section 4, I focused on one way of understanding direct nonvoluntary control, *i.e.* in dispositional terms. I argued that this proposal, like the indirect voluntary control views, cannot provide a satisfactory answer to the control challenge. Both seem unable to explain how we can be said to

control our beliefs in such a way that the condition set by the control requirement is met. However, in introducing the direct nonvoluntary control views, I also quoted some passages from Pamela Hieronymi's work. Hieronymi introduces the notion of *evaluative control*, which is unequivocally direct and nonvoluntary. Reference to evaluative control offers a way to understand our relationship with our beliefs that can explain the fact that we are responsible for them. The question is whether this proposal actually counts as a way of addressing the control challenge. Let me start by spelling it out.

On Hieronymi's view, we are responsible only for what we are 'answerable', where 'to be answerable' is to be open to the relevant reason-seeking why-question. This question is given application as long as the subject has "settled for him- or herself (positively) the question of whether to φ" (Hieronymi 2014, 14). If you have *settled a question*, then you can be asked for your reasons to settle it in a certain way, even though you might have done so for no specific reasons. Crucially, each kind of attitude is related to a corresponding why-question. For instance, the question why someone *believes* that *p* seeks reasons that support the truth of *p*; the question why someone intends to φ seeks reasons to show φ-ing worth doing, and so forth for other attitudes. "[A]ny activity or attitude for which one is answerable…will be reasonably understood as…the settling of some question…on which such reasons bear" (Hieronymi 2008, 360).

In this framework, evaluative control is:

> the control you exercise over your mind by settling questions about whether something is the case or whether to act in some way. (Hieronymi 2014, 21)

Like McHugh's guidance control, evaluative control is understood with reference to the fact that we are responsive to reasons. However, it's not dispositional. We don't exercise direct control over our beliefs insofar as we *would* revise them appropriately in a certain range of situations. Rather, we exercise it in believing this particular belief in virtue of being responsive to *these* very reasons we take to be conclusive, or in virtue of settling the question whether *p* in this very way.

I said that evaluative control is *essentially* both nonvoluntary and direct. First, it is nonvoluntary because it lacks the two distinctive features of *voluntary* control, *i.e.* discretion and awareness. When someone does something voluntarily, they have to determine what is worth doing, be aware of their resolution and manipulate themselves and the world so that that action gets done.[54] Evaluative control lacks both features. First, when we exercise evaluative control over an intention or a belief, we enjoy no discretion over *what intention or belief* we would like to have or would be good or worth having. Instead, we determine what intention to have by exercising

[54] "An activity is *voluntary* just in case you decide to do it for reasons you take to settle the question of whether to do it, therein intend to do it, and, providing all goes well, do it by executing that intention" (Hieronymi 2008, 366).

discretion over what to do, *i.e.* by considering what action is worth doing, and we form the belief that *p* by considering whether *p*. Second, "once we set aside the ways in which awareness enhances self-management", it is unclear why being aware of one's settling a question would be necessary to settle that question.

To put it in other terms, acting is voluntary because, in acting, you are answerable for what you take to be *good* or *worth doing*. Voluntariness is distinctive of the form of answerability that is connected with acting. By contrast, believing[55] is not voluntary because, in believing, you are not answerable for what you take to be good or worth believing. Rather, when you believe, you can be asked for the reasons that you take to show the content of your belief true. Similarly, "[r]esenting, caring, fearing, being grateful, being angry, are none of them voluntary, in this sense—and all for the same reason. Each of these attitudes, like belief, expresses its subject's take on, opinion about, or orientation to, some object" (Hieronymi 2008, 367).

The fact that each attitude expresses its subject's take on the world also explains why evaluative control is essentially direct:

> Our control over our beliefs is immediate in a way that our control over ordinary objects is not. When you change your mind about what is true, you have therein, ipso facto, changed your beliefs. […] The immediacy of evaluative control is thus not temporal or causal but rather a consequence of the constitutive relation between the commitment to p as true and the belief. Since we believe p insofar as we are committed to p as true, when we take p to be true in the relevant way, our beliefs therein change. (Hieronymi 2006, 53–54)

> Because these attitudes embody our take on the world, on what is or is not true or important or worthwhile in it, we control them by thinking about the world, about what is or is not true or important or worthwhile in it. Because our minds change as our take on the world changes—because our minds change as we change our minds—we can be said to be "in control" of our commitment-constituted attitudes. (Hieronymi 2008, 370–71)

We exercise evaluative control over our beliefs insofar as "our minds change as we change our minds". When we change our minds about what is true or worth doing, our attitudes *thereby* change.[56] In answering questions, we bring about our answers and, thereby, our attitudes. We *bring our attitudes to be* by making up our minds on whether *p* is true, or φ worth doing, and so forth. Along these lines, *I am in control* when I exercise evaluative control in the sense that I am not passive

[55] And intending itself.

[56] In turn, this also explains why it is *not accidental* that evaluative control lacks discretion and awareness; if they were involved, this immediacy would get lost.

towards my answers and my attitudes: it is not by accident that I hold certain attitudes, I hold them because they reflect *my* sense of things.

Now, recall that the control challenge can be captured by the question what kind of control we exercise over our beliefs, given that we seem to be responsible for them and yet cannot believe voluntarily. The assumption underlying this question is that we need to exercise voluntary control over something to be responsible for it—otherwise, the fact that we are unable to control our beliefs voluntarily would not raise any difficulty. If my brief reconstruction of her view is fair, Hieronymi rejects this assumption, as I do; she denies that what I call Stoic-voluntariness is necessary for responsibility.

And yet, she does maintain that *some sort* of control is necessary. This is somehow puzzling. On the one hand, accepting the control requirement whilst rejecting the control challenge is not logically problematic, for the control requirement is assumed by the control challenge rather than entailed by it. However, in the first part of the chapter I argued that the control requirement emerges from taking the case of actions as a starting point and from the control reading of Aristotle's thesis, which typically brings with it a certain understanding of 'voluntariness' and 'choice'. And evaluative control is essentially at odds with both the assumption that the case of action is the paradigmatic instance of responsibility and with the Stoic-voluntariness that is typically linked to the control reading. If so, then it's unclear why we should retain the control requirement at all; where does the control reading of Aristotle's thesis gets its grip, once we free ourselves from the assumptions that back it up. It's unclear precisely what it is that control adds to our picture, which we would miss otherwise.[57]

Let's grant that we need to retain a control requirement, and let's accept that evaluative control is a form of control. Now, it emerges the important question whether evaluative control offers a way of meeting the control requirement. I think that this depends on how we understand the status of the *requirement* itself. Two seem to be the most popular understandings of it, amongst the advocates of an accountability framework. According to a weaker view, the control requirement identifies a necessary condition on responsibility so that, if someone is responsible for something, then they have to exercise control over it. According to a second stronger reading, it doesn't only identify a necessary condition but also the very *source* of responsibility. Someone is responsible only for what they control because they are responsible for it *insofar as* they control it—and they are

[57] One possibility is that control helps specify the sense in which we exercise 'agency', where agency is conceptually connected to responsibility. But again, it's unclear why agency (however we understand it) should necessarily involve reference to control.

criticisable or commendable depending on whether they violated or complied with norms.[58] That someone exercises control over something is not only needed for me to be responsible for it; it's the very thing that *explains why* I am responsible for it.

The two readings are compatible, and acceptance of the former does not commit one to the latter. Someone can maintain that control is necessary and yet insists that it's not what explains why someone is responsible for something. From what I have argued in this and previous chapters, the accountability view seems to accept the stronger reading. And also the views I discussed in sections 3 and 4 seem to take control to be required in the strongest sense. I am responsible for the beliefs that result from my controlled activities of belief-formation and revision, or from the controlled exercise of my responsiveness to reasons, *insofar* as I controlled them.

By contrast, I think that evaluative control is meant to meet the weaker, rather than the stronger reading of the control requirement. That is, I think that Hieronymi's proposal could be read along these lines. It is not that we are responsible for our beliefs insofar as we (evaluatively) control them. Rather, we are responsible for some of our beliefs and actions because they reflect our take on the world, and evaluative control is what guarantees that they do so. What's demanded is that we satisfy something like *a reflection requirement*—more than a control requirement—, which sometimes requires control as a condition on its satisfaction. If so, then evaluative control does not identify the source of responsibility; it plays a different explanatory role. Also, when we are not criticisable for an action we cannot avoid doing (because we are forced or ignorant in the relevant ways), that's not because we cannot avoid doing it, but rather because it doesn't display our evaluative orientation.[59] Hieronymi seems to suggest this move when she writes:

> Indeed, it seems natural to solve our original puzzle about why we are responsible for our intentional actions by claiming that we are responsible for them, not because they fall into our jurisdiction—not because we can affect and control them through our intentional actions—*but rather because they reflect our decisions or choices*. (Hieronymi 2014, 16–17) [my italics]

[58] In effect, I also think that the very notion of *compliance* with norms incorporates some reference to control. For if I don't exercise any control, then doing what a norm dictates doesn't amount to *adhering* to it and causing illicit outcomes does not amount to *violating* it. Consider for instance this passage by McHugh: "It's hard to see how we could be legitimately held to norms in a given domain if it wasn't under our control whether we satisfied them. If we lacked such control, how could it be our fault, and potentially blameworthy, if we violated these norms?" (McHugh 2017, 2747).

[59] Sher makes a similar point: "To see how there can be another explanation of our inability to blame people for what they cannot help doing, we may note, first, that a person's lack of control over what he does might figure in the explanation of our inability to blame him in either of two ways. It might enter because control itself is a necessary condition for blame-this is what (1) says-but it might enter, instead, because exercising control over one's acts is necessary for the satisfaction of some further condition that *in turn* is necessary for blame" (Sher 2001, 150).

We are responsible for our actions insofar as they reflect our choices. Along these lines, we shift the focus from the importance of controlling beliefs to the importance of guaranteeing that our attitudes and actions 'reflect our take in the world'. The 'conceptual connection' between a positive answer to a question and a certain attitude guarantees that each attitude *reveals* a different aspect of one's mind. "The differences between [attitudes] simply reflect the different "takes" that a subject can have on an object—the subject can decide whether it is true, or fearful, or worthy of resentment, or whether to do it" (Hieronymi 2008, 369). Our attitudes are the manifest counterparts of our settled questions. They reveal what we take to be true, fearful or worthy. And our take on the world is precisely what we are responsible for.

I believe that this way of understanding evaluative control—and its explanatory role—is compatible with my view. That is, if we take evaluative control to qualify the nature of the relationship between the way in which I see things and my attitudes, then in saying that I exercise evaluative control over my beliefs and intentions we are saying that my beliefs and intentions reflect my evaluative orientation. However, to reach this point, we had both to reject the understanding of voluntariness and choice that typically underlies the control challenge and to give a different explanatory role to control, if any is involved. Very little has remained of the original thrust that underlies the control challenge.

In this chapter, I hope to have made an initial case for the idea that we can vindicate responsibility for beliefs without resorting to a control requirement, and that grounding responsibility on someone's evaluative orientation offers a neat way of accounting for the continuity between responsibility for actions and responsibility for beliefs. In chapter V, I turn to the practices of *accountability* for beliefs.

V. Being accountable for one's beliefs

In previous chapters, I have investigated the merits of two different ways of understanding what it is to deem someone responsible for something, which I referred to as the 'accountability view' and the 'aretaic model'. According to one view, we hold someone responsible for their beliefs insofar as they have violated or complied with some relevant norms that we could expect and demand them not to violate, and we blame or praise them accordingly. According to the other view, when we deem someone responsible, we appraise them aretaically for their conduct insofar as it reflects the kind of person they are, *i.e.* for the values that orient their sensitivity to reasons. I have argued that a virtue-centred framework can provide a more comprehensive and nuanced account of responsibility for beliefs. Typically, we are interested in someone's beliefs because they tell us something about their believer, more than because they rate highly against some standards of rationality.

However, so far I have remained fairly neutral on the precise strength of my thesis, *i.e.* on the question whether the aretaic model is meant to *supplant* or *complement* the accountability view. In fact, I have committed myself only to the following claims. First, that the practices of attributability for beliefs have been overlooked in the literature, although they might be instead the core of the phenomenon of 'responsibility for beliefs'. Second, that an aretaic framework offers a viable and promising way of illuminating the practices of attributability for beliefs. Third, that the accountability view is moulded on the practices of accountability for actions. And finally, that it cannot explain certain key cases of responsibility for beliefs. But these commitments leave it open whether there are cases of responsibility for beliefs that the accountability framework is better (or uniquely) suited to explain.

One natural view is that the accountability view might be better suited to make sense of the practices of *accountability* for beliefs. My goal in this chapter is twofold. First, I clarify whether *there are* practices of accountability for beliefs at all. Second, I suggest that we need not buy into the accountability view to make sense of them. Put otherwise, I assess the following conditional:

> [HYPOTHESIS] If there is something like a practice of holding each other accountable for our beliefs, then we can illuminate it by a virtue-centred, person-focused and not norm-based account.

I start off by introducing Gary Watson's discussion of the practices of accountability for actions. In section 2, I turn to the case of beliefs and defend the antecedent of HYPOTHESIS. I then consider

two possible ways of spelling out the thought that beliefs can constitute wrongs. Believers might wrong others by believing that *p*, first, if believing that *p* disposes one to act immorally (section 3); second, if *p* is *about* someone and, in believing that *p*, the believer does not respect them (section 4). These accounts are not mutually incompatible. For my purposes though, I only need to convince the reader that at least *one* of them is plausible. My own inclination is that the second view is the best suited to explain why some beliefs constitute wrongs, as I find the former problematic in some key respects.

In section 5, I outline a virtue-centred explanation of the practices of accountability for beliefs. My suggestion is that there are *personal* virtues and vices, which are 'no one else's business', and other essentially *interpersonal* ones, like justice and honesty. I can appraise you aretaically for personal traits, whereas I can hold you accountable for interpersonal ones. In fact, crucially, part of what it is to have or lack an interpersonal trait is to be open to the demands and expectations that are characteristics of the practices of accountability. I conclude with some remarks on fairness and morality.

1. *The practices of accountability for actions*

The practices of accountability and the practices of attributability are informed by different kinds of concerns. As Watson puts it,

> Attributability has an importance to ethical life that is distinct from concerns about accountability. Responsibility is important to issues about what it is to lead a life, indeed about what it is to have a life in the biographical sense, and about the quality and character of that life. These issues reflect one face of responsibility (what I will call its aretaic face). Concerns about accountability reflect another. (Watson 1996, 229)

The dimension of attributability is concerned with how we lead our life and what kind of person we are. But I cannot *hold you accountable* for living your life in a certain way, or for being a person of a certain kind. For instance, Watson considers someone who "betrays her ideals by choosing a dull but secure occupation in favor of a riskier but potentially more enriching one" (Watson 1996, 231). According to Watson, we would assess her in deep ways from the aretaic perspective, which is "concerned with the question of what activities and ways of life are most choiceworthy" (*ibid.*). However,

> …by these assessments we are not thereby holding her responsible, as distinct from holding her to be responsible. (Watson 1996:235)

To hold someone accountable is not to commend or criticise them for being a person of a certain kind, or for living a certain kind of life. Rather, their conduct must be such that it affects you or

someone else personally, so to speak. The practices of accountability are informed by essentially *interpersonal* concerns.

> [To hold her responsible], we would have to think that she is accountable to us or to others, whereas in many cases we suppose that such behavior is "nobody's business". Unless we think she is responsible to us or to others to live the best life she can—and that is a moral question—we do not think she is accountable here. (Watson 1996, 231)

We are always accountable *to* someone. When we hold someone accountable, what they do is 'our business'; we have a say in their affairs.

In general, we can say that I am accountable *to* someone else insofar as I have wronged[1] them or done right by them, and I am open to criticism[2] or commendation accordingly. I wrong someone if my intentional action[3] causes (or attempts or threats to cause) some (undeserved)[4] harm to them,[5] or constitutes an instance of injustice committed against them, and I was (or could have been) aware of it,[6] and it was avoidable. Someone is harmed by my behaviour if they could have been better-off, psychologically or materially, if I hadn't so acted. Importantly, this means that for my action to count as *wronging* you, it's not sufficient that you are somehow harmed. The harm has to be brought about intentionally or it has to be a by-product (of which I was or could have been aware) of something done intentionally.

Admittedly, it's more natural to talk of accountability in the negative case, where a *wrongdoing* has occurred. The difficulties of finding an adequately general positive counterpart of 'wronging someone' might provide some further clue that the practices of accountability are primarily a negative phenomenon. However, I want to leave open the possibility of positive accountability. Given the specificity of concepts like 'benefitting', 'helping', or 'aiding', I suggest we say that I can be held accountable to someone insofar as I have *done right by* them. In general, I do right by someone if my intentional action causes (or purports to cause) some benefit to someone, such that

---

[1] I am concerned here with 'wronging others', rather than with 'doing something that is morally wrong'.

[2] T.M. Scanlon writes that blame (in a variety of forms) is made appropriate "by the fact that a person's attitudes are deficient as measured by the standards of morality or some relevant interpersonal relationship" (Scanlon 2013, 106). Angela Smith writes that "accountability blame" is appropriate "whenever an agent's conduct manifests unjustified ill-will or disregard toward others" (Smith 2019, 107).

[3] 'Holding accountable' can also refer to ascriptions of strict liability, and thus be linked with legal accountability. I won't discuss either in this chapter.

[4] That is, where there are no good reasons for me to bring about that harm. I don't wrong my offender if I defend myself from their assault.

[5] The person who holds me accountable and the one to whom I am held accountable need not be the same. I'll assume this from now on.

[6] The literature on the 'epistemic condition' on moral responsibility is extremely rich. For a collection of recent contributions, see Robichaud and Wieland (2017). For insightful discussion see Peels (2014), Rosen (2008) and Sher (2009). On whether the agent's ignorance has to trace back to some culpable act or omission see Smith (2011) and Peels (2011).

I was aware of it and it could have been avoided. Someone benefits from my conduct if they would have been worse off, psychologically or materially, otherwise.

To clarify, the fact that I am a person of a certain kind, who makes some choices rather than others, can also affect those around me in important ways. My being dishonest or kind might have negative or positive effects on them. But what's distinctive of wronging is that the fact that someone will be harmed by my acting is either a good reason for me to act in that way or at least not a sufficiently strong reason *not* to so act. Similarly, doing right by someone doesn't mean that what I do happens to have felicitous consequences for them. I need to act as I do at least partly *because* that will benefit someone. Others' well-being figures amongst my reasons.[7]

Because of this, when I wrong or do right by someone, what I do is also *their* business. Importantly, this makes sense of the fact that the practices of accountability are linked to certain kinds of expectations and demands[8] that others can impose on those who are held accountable. In effect, also the practices of attributability might involve certain expectations and demands, but these would only be *predictions* or *wishes* that someone does something. By contrast, when you hold me accountable, I am expected and demanded to do something in the sense that I *owe* it to you. More precisely, I am expected and demanded not to wrong you or, if I have already wronged you, to make up for it (by way of apologies, compensations, sanctions). Also, I can be expected and demanded to do right by you, if failing to do so would constitute a(n even very mild) wrong.[9]

A way to distinguish the expectations involved in the practices of attributability from the ones involved in the practices of accountability is to look at the reasons that support them. The expectations that belong to accountability, but not to attributability, draw on considerations that involve someone else besides the agent who's assessed. Consider an example. Suppose that you are disappointed at a friend who (without excuse) broke an important promise, and that you hold them accountable for it. And suppose that you are also disappointed at a friend who you think is wasting their talent. Strikingly, disappointment can figure in evaluations belonging to both the accountability and the attributability practices. However, it seems to be based on different kinds of

---

[7] In chapter III, I followed several authors in claiming that the importance of mutual regard is the core of what we call the 'moral' realm. Now, it seems that we can say that wronging others counts as an instance of disregard of mutual regard, and that doing right by them is to value their well-being and honour the relationships of reciprocal respect. If so, then when we hold others accountable for what they do, both their conduct and our assessments are a moral matter. In fact, I agree that the practices of accountability are always informed by concerns about others and, if that is the marker of morality, then I agree that they all belong to it. However, in the last part of chapter III, I suggested that several virtues and vices are concerned with mutual regard *in different ways*. If so, then the claim that the practices of accountability all belong to the moral realm might be underspecified. I elaborate on these ideas in section 5.3 below.

[8] For the idea that the concept of moral obligation is irreducibly second-personal see Darwall (1996). In chapter 6, Darwall extends his second-personal analysis to the concepts of dignity and respect, which are understood as respectively the authority to make claims and demands, and proper recognition of this standing.

[9] The relationships between the positive case (of doing right by someone) and expectations and demands are more difficult to pin down. At least in some cases, you do right by me *precisely because* I could have *not* expected and demanded you to do something. Yet, other times I thank you even though I could have expected you to act as you did.

reasons in the two contexts.[10] If asked why you are disappointed in the first scenario, you would say things like: 'I trusted them'; 'they hurt me'; 'they should have not done so'; 'they gave me their word; 'I thought we were friends'. By contrast, with respect to the second case, you would say things like 'such a lack of ambition, what a pity!'; 'I think they could do more with their life'; 'they could be braver'; 'they are so timid and fearful!'; 'so shallow, it's just so sad'. The first set of considerations include reference to the fact that your friend's failure to keep their promise infringes *your* integrity, autonomy or well-being,[11] and they cast doubts on your relationship with them. The second set of considerations come from a place of (some) detachment. Of course you are invested in your friend life, or else you would not be disappointed at all, but their life choices don't wrong you, though they can make you sad, angry, disappointed and so forth. Even though 'it's such a pity', it's not your business after all. In general then, when you hold someone accountable, your expectation that they do (or stop doing) something necessarily draws on considerations that involve *someone else* as well. By contrast, the expectations you might have in appraising what kind of person someone is are not supported by reasons that necessarily involve you and your wellbeing. And this difference makes sense of why you can expect and demand friends to keep promises, in the relevant sense, but not to exploit their talent.

I said that aretaic appraisals come from a place of some 'detachment'. By this, I only mean that you don't owe it to me to modify your behaviour. What kind of person you are or the nature of your evaluative orientation is not my business. However, this doesn't mean that, in appraising you aretaically, I need be disinterested. In fact, I might well be more invested in what kind of person my partner is than in the wrongdoings of a stranger. But the point remains that I cannot impose on the former the same kind of expectations and demands that I can impose on the latter. For instance, suppose that I am angry at my partner for wasting time playing videogames every weekend and angry at my neighbour for not recycling. Of course, I am closer to my partner than to my neighbour. And this might seem to imply that I could *always* expect and demand *more* from him than from the latter. However, my partner doesn't wrong me[12] or anyone else. I can think that they are wasting their time, that it's a pity and a poor way of leading a life. I'm angry because they could be better, and they don't take up that opportunity. But strictly speaking, I cannot demand that they

[10] I remain neutral as to the metaphysical status of disappointment, resentment, anger, or pride, *i.e.* on whether they are emotions, beliefs or combinations of the two. For instance, Wallace (1994) argues that they all involve the *belief* that the person's commission of a moral wrong would make a reactive attitude appropriate (Wallace 1994, 83). By contrast, Deigh insists that resentment does not imply such judgments or indeed any propositional thought (Deigh 2011). However we understand their nature, in any case, what matters for my purposes is that we grant that they can be supported by reasons.

[11] Henceforth, I will use 'well-being' as a placeholder for happiness, autonomy, integrity and the like.

[12] For simplicity, let's assume that he plays when I am already busy, we don't have children, frequent exposure to videogames does not otherwise affect his personality, his playing does not worsen our financial situation, and so forth.

use their time in a more nurturing way. By contrast, I can expect and demand that my neighbour behaves differently even if I am not personally close to them. The expectations, demands and annoyance involved when I hold my neighbour accountable are based on the fact that they owe it to me, their community, state or planet. In justifying my anger towards my neighbour, I would draw on the fact that their action wrongs others.

In fact, I would demand and expect that my neighbour, for instance, starts to recycle. By contrast, if I expect that my partner changes his behaviour, that is not because I think he *owes* it to me, let alone that he should change *for me*.[13] If anything, he should change his behaviour for himself. And if he did start to fight his laziness or childishness, that change would not stand for a way of *compensating* me as much as for a sign of personal growth.

Lastly, let me stress that the distinctive interpersonal element that characterises the practices of accountability explains why accountability introduces concerns about fairness and control. As Watson puts it,

> Because some of the practices - and notably the practice of moral accountability - involve the imposition of demands on people, I shall argue, they raise issues of fairness that do not arise for aretaic appraisal. It is these concerns about fairness that underlie the requirement of control (or avoidability) as a condition of moral accountability. (Watson 1996, 235)

Given that accountability is responsibility *to* someone and that its demands and expectations require something from the person who's held accountable, *e.g.* a compensation, it introduces concerns about fairness because it might be unfair to impose those demands and expectations. In turn, these concerns about fairness introduce the importance of control, since it seems unfair to demand or expect from me something that doesn't fall under what I can possibly control and affect. That fairness and control play an important role in the practice of accountability is confirmed by the fact that we ordinarily protest if someone holds us accountable when we could have not done otherwise, if they claim that we did something wrong that we believe we haven't done, or if they expect unduly compensations. I discussed these issues in more detail in chapter IV.

To sum up, the practices of accountability and attributability are distinct phenomena insofar as they are informed by different concerns. When we hold someone accountable, we do so because they have somehow wronged us or did right by us, whereas this particular sort of interpersonality is not a necessary ingredient of aretaic appraisals. Relatedly, the person who commits a wrong can

[13] And if he does, it would be only *indirectly* so, *i.e.* because I deserve to have someone who's better for themselves as a life companion.

be expected and demanded to make up for it. Finally, because of this underlying structure, the practices of accountability are concerned also with fairness and control.

If this is a fair sketch of the nature of the practices of accountability, then it is quite easy to see why they have favoured a *norm*-based explanation. For it is quite natural to make sense of the kind of 'should' we find in 'my friend should have not broken their promise' or in 'my neighbour should recycle' in terms of compliance with norms. Quite intuitively, the explanation goes: my friend should have not broken the promise because, in doing so, they violated the norm of promise-keeping to which they were subjected. However, it is not obvious that this is the only possible way in which we can make sense of the practices of accountability. Before exploring an alternative in section 5, let me address the question *whether* we can hold someone accountable for their *beliefs*. In fact, following Watson, I have drawn the distinction between the practices of holding accountable and appraising aretaically by looking exclusively at the case of actions. But this does not suffice to defend the antecedent of my HYPOTHESIS.

2. Beliefs that wrong

The question is whether we can hold others accountable for their *beliefs*. In light of section 1, holding someone accountable for their beliefs would require us to think that their *believing* is the kind of thing that can wrong or do right by others. This cannot simply mean that your beliefs make me sad, happy or angry, sometimes. Nor, given the nature of beliefs, can it mean that you believe that *p with a view* to harm or benefit me. The fact that a belief will have certain (harmful or beneficial) effects on others doesn't support the truth of its content.[14] Thus, we cannot believe upon considerations of that sort, nor can we be culpable if we don't. Finally, if I can wrong you by believing, then you should be able to place some expectations and demands on me. And if I can do right by you by believing something, then you might thank me for believing as I do. This much sets out our desideratum.

I think that the key challenge is to explain how believing can somehow fail to manifest a sufficient degree of respect for others, or else manifest it, even though considerations *about* others' well-being cannot support the truth of beliefs. That beliefs have at least to manifest some degree of respect for others' well-being follows from the distinctive nature of the practices of accountability. If they didn't, then it would be hard to see how we could be held accountable for them. That is, if we accepted the idea that there are practices of accountability that are free from concerns about

[14] Besides beliefs *about* others' well-being, that is.

wronging and doing right by others, then it would be difficult to see the practices of accountability as constituting a distinctive set at all.

Notice that the conceptual connection between accountability and the relevant kind of interpersonality implies also that believers cannot be held accountable for a belief simply *insofar as it is false*.[15] Being false does not (in and by itself) constitutes a wrong to anyone. For instance, think about the belief that whales are fish. Of course, believing truly is important for us and, in believing falsely, I might well be open to a wide range of (mostly negative) evaluations. However, I cannot be held *accountable* for it; there's no injured party that can expect and demand anything from me.[16]

If beliefs cannot wrong someone by being false, they have to do so in other ways. In sections 3 and 4 I consider two possibilities.[17] According to one view, I can hold you accountable for holding a belief if, by acting upon it, you would commit a wrong. Here, the wrongness of the belief is derivative on the wrongness of the action that could be carried out in its light. According to a second view, in certain circumstances, beliefs can wrong by being unjustified. In particular, believers might wrong someone by believing things about them for certain (insufficient) reasons.

3. Derivative wrongness[18]

In a recent interview to The Irish Times, Tim Williamson has suggested that it's not okay to have beliefs that could lead their believers to act wrongly.[19] Let me report the passage at length:

> A totalitarian state where it's illegal to have unorthodox opinions is a nightmare. But think of a Holocaust-denier who just shrugs his shoulders and says 'I'm entitled to my opinion' when presented with overwhelming evidence that the Holocaust really happened. What he's doing is not okay. In both a rational and an ethical sense, he's not entitled to ignore the evidence on something so important. Not to be coy about my political opinions, the Brexit vote and the Trump election resulted from many people

[15] Strikingly, although most accounts of responsibility for beliefs seem to be variations on the accountability view, most of the examples they consider are ones where the beliefs are false, irrational or unjustified, so, crucially, where no wrong has occurred. In other words, these accounts include concerns about norms, fairness and control, which they inherit from explanations of the practices of holding someone accountable for their actions, whilst overlooking that these concerns generate from the very fact that someone *wronged* someone else. But without a potential for a wrong, concerns about norms, fairness and control lose their *raison d'être*.

[16] In effect, there might be cases where you can rightly hold me accountable for the belief that whales are fish. But we will need a richer story to explain why that is so, one that goes beyond the fact that whales are not fish. The falsity of my belief does not provide a bedrock explanation for why you would hold me accountable for it, if you do.

[17] The topic has received increasing attention in recent years. For an overview see Basu (2018).

[18] See Montmarquet (1992) for the converse view that certain morally culpable acts are culpable insofar as they are grounded in beliefs which are themselves culpable. See also Montmarquet (2008); Adams (1984; 1985).

[19] Clifford says: "But forasmuch as no belief held by one man, however seemingly trivial the belief, and however obscure the believer, is ever actually insignificant or without its effect on the fate of mankind, we have no choice but to extend our judgment to all cases of belief whatever", and *on this basis* he concludes "It is wrong always, everywhere, and for any one, to believe anything on insufficient evidence" (Clifford 1877, 24). *Cf.* Code (2004) on social epistemology and how ignorance can contribute to social-political arrangements of dominance and subordination.

> voting on the basis of bigoted, ill-informed opinions they felt entitled to. Brexit is likely to have bad consequences in these islands, Trump in the whole world. It's in the nature of opinions that people act on them, in ways that affect others as well as themselves, so it can't be morally indifferent what opinions they hold. But not all moral matters are suitable for legislation.[20]

Given that this is an extract from a newspaper article, it would be unfair to take it as an accurate representation of Williamson's view. Regardless of exegetical concerns, however, it offers a thesis worth considering. The core idea is that my belief is 'not okay' if I risk carrying out morally wrong actions by "acting upon it". If so, then the belief itself constitutes a wrong. Notice that the actions that result from the beliefs in question are *morally* wrong, whilst the beliefs are both *rationally* and *ethically* problematic. They are rationally problematic because there are insufficiently supported. But if they were merely rationally problematic, it would be difficult to explain why they issue in morally problematic actions. There would be a gap between kinds of normativity. So, Williamson suggests, they are also *ethically* problematic. The idea is fairly intuitive, but several qualifications are due.

First, consider the claim that a belief constitutes a wrong if, were the agent to *act upon it*, they would carry out an immoral action. So formulated, the thought is underspecified, for there are several ways in which we can understand what it is to 'act upon' a belief. On one, I think misleading, view, 'to act upon a belief' is a matter of acting for the reason that one has the belief. For instance, say that I'm quite aware of being biased against Asian people, and that I know that it's awful. Because of this, I act on my beliefs by going to bias training and moving to an area of town where there are many Asian people in the attempt to get to know many of them and modify my attitude. In cases like this, I do act upon my belief. And yet, my actions are *not* wrong, though my belief is. If so, then this cannot be the right way of fleshing out Williamson's remarks.

Rather, I suggest we say that Williamson is concerned with cases where someone who believes that *p* is *disposed* to take the putative fact that *p* as contributing to their reasons for action. 'To act upon a belief' is not a matter of acting for the reason that one has the belief' (as in the case of a reflective racist), but for the (presumed) reason that *p*. Talking of dispositions in this context has several advantages. First, it allows me to specify the relation between the belief and the action in the relevant way. My racists belief do dispose me to take certain (putative) facts to support acting in a biased way, though I can make sure I avoid following my inclinations. Furthermore, it allows me to single out the beliefs that we would consider wrongs and leave out the ones that would be required for me to commit a wrong action without constituting wrongs themselves. For instance, say that I intend to terminate your life and plan to set your house on fire. In a strong sense, I do

[20] https://www.irishtimes.com/culture/unthinkable-how-do-we-know-anything-1.2992520.

'act upon the belief' that you are in the house. In effect, I wouldn't proceed without holding it. And yet, the belief that you are in the house doesn't constitute a wrong in and by itself, for it's not sufficient to dispose me to set your house on fire. Finally, identifying the beliefs that dispose us to act in certain ways makes sense of the fact that a belief might constitute a wrong even if I *never* acted upon it.

A believer might have a belief that disposes them to carry out wrong actions in three kinds of situations. In the first scenario, the believer both knows that they believe that *p* and that their belief might dispose them to carry out an action that they themselves recognise as immoral. These are the simplest cases to assess: on these conditions, the belief in question will always constitute a wrong. The believer is fully aware of its wrongness and there are no mitigating circumstances.

The other two scenarios are philosophically richer. Either the believer doesn't think that the belief that *p* is likely to yield that action, or they don't think that that action is morally wrong. Let me start from the former case. Here, depending on *why* the believer failed to see that their belief would issue in a certain action (that is immoral), their belief might or might not constitute a wrong. It will *not* constitute a wrong if the believer is (non-culpably) ignorant of some key facts, lacks elements to assess causal nexuses, or the relationship between the belief and action is particularly winding. These seem to be more akin to performance errors than to normative disfunctions. The believer simply doesn't realise that their belief risks bringing about a certain action.

In other cases, it might be fairly easy for the believer to see that their belief is likely to bring about a certain action, and yet they get things wrong. If their mistake is due to carelessness, then the belief constitutes a wrong. The point is not that they are culpable if they *could have* believed that their belief was risky; if it was in their power to so believe. This would bring the control requirement back into the picture. Rather, they are culpable insofar as they are *careless*, if they are someone who didn't put effort in considering appropriately the importance of a certain topic, or who underestimated the risks of believing certain things. As Williamson puts it, "he's not entitled to ignore the evidence on something *so important*" [my italics]. In this context, a subject matter is 'important' precisely because a belief about it can issue in actions that are located in the realm of morality.[21] Understanding the importance of the subject matter is to understand the importance of believing things about it truly, given the impact that false beliefs might have on the world around us. And to be careless is to fail to give some subject matter due importance, thereby enhancing the possibility that certain immoral actions occur.

[21] For the view that the epistemic value of having a particular true belief should actually be weighted by how interested we are in the topic see Fallis (2006), who builds on Goldman (1999).

For instance, suppose that you know that raping is morally wrong. Yet, you also believe that the feminist movement has gone too far and that activists are exaggerating. Many are pressing charges of rape without having been harassed. At the end of the day, catcalling is quite flattering and rape-jokes are funny. You hold these beliefs and think that they don't issue in immoral actions. You wouldn't rape anyone, in light of them. Therefore, you'd insist, it's not problematic to hold any of them. Now, it seems to me that situations like this identify cases of beliefs that wrong, insofar as they manifest that the believer is too quick in disregarding the possibility that they might lead to immoral conduct. The believer downplays the risk that these opinions might dispose one to harass someone, or that they might normalize a certain way of thinking about rape.

In a second cluster of cases, someone might know that their belief will yield a certain action but fail to see that action as immoral. They are convinced that it is not. For instance, consider the Holocaust denier. Here the problem is not that they fail to see importance of the topic. Nor do they belittle the risks of bringing about certain actions. Rather, the Holocaust denier deems it *very* important to know what happened during the WWII, precisely given that the resulting beliefs will issue in actions that (they would agree) are morally relevant. However, because of *what* they believe, they think that insulting survivors is not morally wrong. Rather, from their perspective, it's morally wrong to resent Germans unduly. The problem here is not that they are careless. Rather, it lies with their understanding of right and wrong. They hold a certain belief because they don't see the actions that it leads to as immoral. Also in these cases, it seems to me, the belief constitutes a wrong. The fact that the believer *cannot* see things rightly does not excuse them.[22]

Someone might resist this claim on the basis that, if the belief held by the Holocaust denier can constitute a wrong despite its believer's inability to see the (possibly) resulting action as immoral, then also some beliefs held by psychopaths can count as wrongs. However, the objection would go, *precisely because* the psychopath cannot see the corresponding action as immoral, we might want to say that their belief does *not* constitute a wrong. Holding the psychopath accountable for their belief seems unfair.

In effect, I do think that there are some differences between the Holocaust denier's and the psychopath's ways of (mis)understanding morality, such that we could maintain that the former's beliefs constitute wrongs whilst the latter's don't. For instance, the Holoucast denier mistakes right for wrong, whereas the psychopath fails to understand morality *tout court*. If the psychopath fails to see something as morally wrong, that is not because they are convinced that it is morally right; they simply don't see it as a moral matter at all. However, it remains unclear why

[22] Rosen has argued for the opposite view that moral ignorance can exculpate (Rosen 2004). See Harman (2011) and Talbert (2013) for criticism.

this (or any) difference between them should entail that only the beliefs held by the Holocaust denier can count as wrongs. For one thing, given that they get things the other way round, the Holocaust denier hardly *understands* morality any better than the psychopath can. Thus, it's not obvious that they can play the game of interpersonality any better than them. Second, more importantly, the fact that the psychopath doesn't see a certain action as immoral does not trivially make their belief less of a wrong towards those who are harmed by it.[23]

More generally, it seems to me that if we accept the idea that someone's belief doesn't count as a wrong because *they* couldn't see it as favouring a disposition to act immorally, then very few beliefs will count as a wrong. Instead, that someone does not see a certain action as morally wrong does not excuse them from the practices of accountability. However, and importantly, this doesn't mean that someone's inability to understand morality doesn't carry any weight, when we assess them. On the aretaic model, precisely why someone doesn't see an action as morally wrong does affect in important ways the way in which we think of and treat them. That they have a certain pathology or have had a particular upbringing will work as a mitigating circumstance, and it will determine that we treat them differently from how we would treat someone else. But this, in and by itself, doesn't entail that their beliefs don't wrong, or that we are outside the game of accountability *tout court*.

To sum up, on this view my belief constitutes a wrong in case I know that it is likely to dispose me to φ, where φ-ing is morally wrong, whether I know that it is morally wrong or not; also, it constitutes a wrong in case I fail to know that it is likely to dispose me to φ because I fail to see the importance and significance of the subject matter, due to carelessness or other vices.[24] I think that this view captures an intuitive way in which we think about beliefs that wrong. Yet, some questions remain unanswered.

First, it's difficult to understand whether it is actually the belief *itself* that wrongs. In fact, the idea seems to be that, although the wrongness of the belief is derivative on the wrongness of the actions that it disposes me to carry out, also the belief itself constitutes a wrong. The decision

[23] The question about the psychopath's moral status has received considerable attention. For some, psychopaths cannot be held accountable and are not blameworthy because we cannot make demands on someone who cannot understanding their force, for instance see Levy (2007b), McHugh (2013), Watson (2011). Nevertheless, in discussing the case of Robert Harris, Watson claims that we can indeed evaluate him in attributability-terms. For others, at least sometimes, psychopaths are blameworthy, see Talbert (2008). Similarly, Scanlon argues that "in the sense relevant to questions of blame a person can be blind to reasons that he really does have" (T. M. Scanlon 2012, 93).

[24] Due to space limitations, I cannot elaborate on how this approach can accommodate the positive case, *i.e.* the idea that believers can do right by someone by believing. Plausibly, the rightness of my belief would derive from the rightness of the actions that I would be disposed to carry out by having that belief. However, it seems to me that the criteria would have to be more stringent: my belief would count as one that does right by someone only if I know both that having a belief would dispose me to φ and that φ-ing is morally right. For it wouldn't be a belief that does right by someone if I thought that φ-ing is morally wrong, or if I didn't think that having it would dispose me to φ.

to phrase things in terms of dispositions to act was an attempt to address worries of this sort. And yet, it's still not obvious why *being disposed* to act in a certain way should matter as much as (or in the same sense as) acting in that very way.

Second, this framework encounters some difficulties in accommodating moral disagreement. For in some cases, it's not obvious whether the relevant action is immoral and, in turn, whether the corresponding belief constitutes a wrong. Consider the debate on abortion. Both those who fight for the right to abort and those who fight against the possibility of getting an abortion consider themselves as having strong arguments in favour of their views. The former draw on persons' right to decide over their bodies; the latter insist that terminations of pregnancy constitute murder. For *both*, the other ones' belief would issue immoral actions and thus constitutes a wrong. But it's hard to adjudicate the matter.

Finally, perhaps most importantly, I think that this proposal fails to place sufficient emphasis on the distinctive kind of interpersonality that underlies the practices of accountability. In fact, in deriving the wrongness of the beliefs from the wrongness of the actions, the view doesn't explain why those beliefs themselves count as instances of disregard of mutual respect; *precisely how* the believer fails to respect someone else *in believing* certain things. Put otherwise, the view remains overly silent on what characterises the relevant beliefs; on how they themselves play a role within the relationships of mutual regard. In section 4, I turn to consider an alternative explanation that I think does better in identifying the source of *the belief*'s own wrongness.

4. Unjustified beliefs

In section 2, I suggested that beliefs can constitute wrongs also when they are unjustified.[25] However, not all unjustified beliefs wrong others. My proposal is that, if I hold a belief about someone,[26] it constitutes a wrong if it is not sufficiently supported by reasons, for that betrays that I don't care about getting things right[27] about another person. The threshold of what counts as 'sufficiently' supported depends on the nature of my relationship with the person about which I

[25] To clarify, most of the beliefs I considered in section 3 are also unjustified: no one can be *justified* in believing that rape-jokes are okay, or that the Holocaust did not happen. However, on that view, beliefs weren't wrongs *because of that* but because of the immoral actions they could back up.

[26] In section 3, I suggested that the importance of a subject matter can depend on whether forming a belief about it could dispose me to act immorally. Here, I want to add that a subject matter is important—so that it's important to form a true belief about it—also if it concerns others, especially if they are our loved ones.

[27] When I talk of 'getting things right' I mean something more than 'believing truly whether *p* is the case'. Sometimes, I might disrespect a friend in forming and holding a *true* belief about them, if I formed it by luck. Of course, it might be easier for them *to get to know* of my carelessness if my belief is false but this does not mean that I fail to care about getting things right only when my belief ends up being false. Also, I might wrong others by believing something true about them if, for instance, they had asked me not to investigate into a certain subject matter, which concerns them and not me and about which I didn't need to form any opinion.

hold a belief.[28] Moreover, if we stand in certain relationships, even failing to form a belief about them can constitute a wrong.

Some preliminary remarks. In saying that we can wrong someone by forming and holding unjustified beliefs *about them*, the meaning of 'about' is (notoriously) difficult to pin down. I use it in the ordinary sense. Beliefs about someone else are beliefs whose content involves reference to someone other than the speaker. They can concern others' present, past or future actions, attitudes, character. They are about others' plans, desires, commitments. I also want to remain as neutral as possible on what we mean by 'unjustified'. In general, someone is unjustified in forming and holding a belief about someone else if they believe upon insufficient evidence, or for the wrong kind of reason, or hastily and sloppily.

In believing something unjustified about you, I wrong you because I show disrespect towards you. In fact, to honour the relationships of mutual regard, I don't simply have to behave in certain ways with you; also, I need to be fair in assessing propositions that concern you. I need to form my beliefs *respectfully*, as it were. Part of what it is to respect you is to care about getting things right about you. Of course, this doesn't imply that your well-being should figure amongst my reasons to assess the truth of *p*; that would provide the wrong kind of reason. For instance, I cannot believe that you are a great friend, when you're not, just because my believing so would make you happy. Rather, assessing the truth of a proposition *p* about you *respectfully* is a matter of understanding what kind of information is relevant to assess whether *p* and how much information is sufficient to settle the question whether *p*.[29] Put otherwise, I need both to discern what kind of considerations matter to form an accurate belief about you, and to understand when I can settle the question whether *p* and when instead I need to dig further into the issue. When I do believe something about you respectfully, I do right by you.

I think that this view is more attractive than the one I discussed in section 3 because it makes better sense of the fact that my believing *itself* constitutes an instance of disrespect of mutual regard. If I form or hold an unjustified belief about you, then I wrong you by believing. I fail to respect you *in believing* something about you. The wrongness of my belief is not derivative on

[28] Along these lines, I agree that non-evidential factors can influence the *threshold* of when a belief is epistemically justified or sufficiently supported. To reiterate, this doesn't entail that practical considerations indicating a belief's value can be reasons for belief. Throughout this thesis, I've been concerned with one's sensitivity to truth-related reasons for beliefs, though I think *which* truth-related reasons someone considers can be affected by certain interpersonal relationships. For a discussion of different accounts of how our personal relationships might ground reasons for doxastic partiality see Crawford (2020).

[29] To see that it cannot only be a matter of understanding *how much* information we need to settle a question, but also the *quality* of it, consider cases like this. Suppose that you are considering whether to believe a woman who says she's been sexually assaulted. You may need more information to make up your mind, and you (legitimately) ask more questions. However, say that you want to know whether she was wearing a skirt. This kind of information is of the wrong kind. In requiring it to make up your mind, you are wronging her.

anything else, *e.g.* the actions that it disposes me to carry out. Thus, this view ties my belief to the distinctive interpersonality that underlies the practices of accountability.

Yet, it's impossible to codify what it takes to assess your reasons respectfully. We cannot give a formula that delivers precise guidelines for each and every situation. That will have to be assessed on a case-by-case basis. However, we can investigate what affects that variability, *i.e.* why sometimes we can settle the question fairly quickly, whereas sometimes doing so constitutes a wrong. My suggestion is that assessing the truth of a proposition about someone else respectfully depends on the relationship you have with the person about which your belief is. In particular, you owe more care in inquiring into the truth of *p* to your friends and loved ones, when *p* is about them. And you owe at least a *minimal* amount of care to *every* human being, insofar as they are fellow persons. I expand on both ideas one at a time.

4.1. Beliefs within relationships

In chapter I, I mentioned Jessica Brown's example of Maud the clairvoyant. Maud thinks that she is a clairvoyant and, using her powers, she comes to believe that the President is in New York. Yet, she has strong evidence that the President is not in New York and no evidence that she really is a clairvoyant. Brown argues that Maud is blameworthy for her belief, even though we wouldn't resent her. In a footnote Brown adds:

> There may be other cases in which blameworthy belief does lead to resentment and demands for apology. E.g. if a close friend or lover believes on insufficient evidence that I've had an affair or deceived them, I might be resentful and ask for an apology. (Brown 2020, 17, f. 5)

In their recent *Doxastic wronging*, Rima Basu and Mark Schroeder make a similar point: suppose you've been sober for eight months after a long struggle with alcoholism. One night, you return home from a reception and your spouse smells the wine someone spilled on your jacket. Your spouse seems to think that you "have fallen off of the wagon". Basu and Schroeder gloss by saying:

> If you are like us, then you will be prone to feel wounded by this. Yes, you have a long history of falling off of the wagon, and yes, there is some evidence that this time is another. You can see how it could be reasonable for someone to draw this conclusion. But it still hurts—not least because in your eyes, tonight was an achievement to stay on the wagon despite adverse circumstances. (Basu and Schroeder 2019, 182)

And along similar lines, Sarah Stroud writes:

> Suppose, for instance, that a third party reports that your friend Sam recently slept with someone and then cruelly never returned any of that person's calls, knowingly breaking

> that person's heart. The question I now want to pose is, how, as a good friend, should you respond to this new information? I don't mean what actions you should take in this situation, but rather what you should make of this new information. What conclusions should you draw from it? How should you update your set of beliefs about your friend's actions and character in light of it? ... Indeed, … I would go so far as to say that you owe your friends something other than an impartial and disinterested review of the evidence where they are concerned. (Stroud 2006, 503–4)

All these passages seem to point to the same phenomenon, *i.e.* that when a close friend[30] or lover believes on insufficient evidence that I've deceived them or done something wrong, I might not only be angry at them but also resentful and wounded. I might be owed an apology and legitimately demand it. I might hold them accountable for forming and holding their belief.

My suggestion is that the relationship that holds between a friend or lover and myself is such that, in inquiring into the truth of a proposition about them, I owe them to be more careful than I might be in assessing the truth of other propositions. Let me stress once again that this doesn't mean that I should not believe the evidence because doing so would make a friend unhappy. I accept the core of the classic evidentialist thesis that only truth-related consideration should bear on my forming and holding a belief that *p*. The point is that, if the proposition at issue is about a friend, I might need to take extra[31] care to ensure that I get things right; I might raise the threshold of what counts as 'sufficient' evidence.

There are at least two ways of understanding the point. According to the advocates of pragmatic encroachment,[32] the considerations that set the bar of what counts as sufficient evidence will be moral.[33] Others say instead that that bar is set by the very nature of friendship. For instance, Sarah Stroud argues that, if we are friends, there are certain demands on me about what to do and to believe with respect to you, and that these demands of friendship are constitutive of what it is to be friends. However, she comments, "by calling constitutive elements of friendship 'demands' of friendship, or things you owe your friend, I do not wish to imply that they are necessarily moral demands" (Stroud 2006, 502). I don't aim to settle the issue here. For one thing, it seems to me uncontroversial that we wouldn't possibly be friends if I always thought the worst of you, trusted whoever reports awful deeds that you allegedly carried out, and never gave you the benefit of the

[30] On doxastic expectations and friendship see also Keller (2004) and Aikin (2008).
[31] As I will say shortly in the text, if we are friends, I might also be expected to believe upon *less* evidence than I would require normally. If I asked for more evidence, that might be taken as lack of trust or of knowledge—both of which would cast doubt on the idea that we are friends at all.
[32] *e.g.* Basu and Schroeder.
[33] *Cf.*: "But we infer, in contrast, that the bar for sufficiency of evidence does depend on moral factors. This picture is not inconsistent with core evidentialism. On our view, any belief is epistemically rational just in case it is supported by sufficient evidence. But the bar for sufficiency on evidence is sensitive to moral considerations" (Basu and Schroeder 2019, 201–2).

doubt.[34] Part of what it is to be friends is that, in forming beliefs about you, I wouldn't be satisfied with less than a certain amount (and quality) of information. And you would expect and demand that I were not. However, whether demands and expectations of this kind necessarily belong to the moral realm is another question—one that depends on whether being a good friend is itself a moral issue. Intuitively, I'd agree with Stroud that it *need* not be. At least in certain cases, and crucially if I don't disrespect you *as a person*, I might indeed fail to be a good friend but only in the sense that I haven't treated you any *better* than I would have treated a stranger. But I would have not necessarily behaved as a *morally* bad person.

It's worth stressing that some expectations and demands are constitutive also of other kinds of relationships, besides friendship and kinship. For instance, getting things right about someone might be part of what it is to have a job or institutional role. If I am a jury, I am expected to care about getting things right about the defendant. Someone else might not be expected so to care, and I am not expected to care about getting things right about others, *e.g.* the defendant's mother.[35] Similarly, if I am a referee or a teacher, I need to make sure that I get it right whether someone cheated, and so forth for different roles. Thus, if I am a jury, a referee or a teacher, I can wrong the defendant, the player and the student by believing certain things about them, or by believing things about them in a certain way. In all these cases, I fail to live up to the expectations to which I was rightly subjected.[36] Yet, once again it's not obvious that, if I fail to live up to the expectations that are constitutive of certain relationships, I am *thereby* an immoral person.

However we decide on these issues, the key point is that it is possible to wrong those with whom we are in a certain relationship by misunderstanding how much or what kind of information we need to assess the truth of a proposition about them. The mistake can be by defect or by excess: I might wrong a friend both in settling the question whether *p* on *too little* information and in requiring *too much* information. For instance, recall the case outlined by Stroud. Suppose that someone tells me that my close "friend Sam recently slept with someone and then cruelly never returned any of that person's calls, knowingly breaking that person's heart". If this were sufficient to settle the question whether Sam did it, then I would fail to be a good friend to Sam. The fact

[34] I think that this stands even if I've always *behaved* as a great friend.

[35] Provided that she's not involved in the case.

[36] The claim that believers can be expected and demanded to hold different beliefs according to their roles in society sets the ground for a related point, *i.e.* that also expressing a certain value can take different forms according to the believer's role in society. Different persons can express a similar evaluative orientation by means of holding different beliefs. For instance, being honest will require different persons to have different beliefs and act in different ways. As Hursthouse writes: "Where relevant, we expect [honest people] to uphold the ideals of truth and honesty in their jobs; if they are academics, to be resistant to fashion and scrupulous in their research; if they are teachers, to resist pressure to teach what they do not believe, or if doctors to defend the importance of trust between doctor and patient, or if in business, to resist sharp practice and argue for honesty as the best policy" (Hursthouse 1999, 11). In effect, it is unclear to me whether the accountability view can accommodate these situations, *i.e.* whether it can explain cases where someone is criticisable for holding the belief that *p* even though someone else, who also believes that *p*, is not.

that we are friends implies that to simply trust a third party's testimony, not even asking Sam for their version of the story, counts as wronging them. However, suppose that Sam told me that it's all a lie, and that they are actually the ones who've been ghosted. If I asked them to show me their texts to support their case, Sam wouldn't be pleased at my reaction, and legitimately so. I would be asking for too much evidence. If we are close friends, I don't need a proof that Sam's telling me the truth.

Crucially, the fact that we are in a certain kind of relationship does not only place a demand on me to believe things about you *in a certain way*. Sometimes, it's also expected of me *that* I believe something about you. For instance, if I know that Sam felt unwell yesterday, I might be expected to form a (true) belief about their state of health today. Or if I'm your mother and you are three, I am expected to hold true beliefs about your whereabouts. In other words, not only do I owe it to my friends and kin to believe truly, *when* I believe something about them. Also, I owe it to them to hold certain (true) beliefs about them to begin with. The demands placed on me in these cases are not conditional.

This sheds some light on why I am not accountable for cases like ABSENCE, which I discussed in chapter III. In ABSENCE, I fail to form the belief that you didn't come to work today. We are colleagues but not friends, you are missing an important meeting, and you are usually an organised person. In chapter III, I said that you couldn't resent me for failing to form a belief about you. Now I can explain why that's so: if I formed a belief about you, I would owe it to you to get things right. But I don't owe it to you to form any belief about you, because we are only acquaintances. I'd owe it to you only if we stood in certain kinds of relationships.

4.2. Beliefs outside relationships

Along these lines, we are expected to get it right whether *p* especially if *p* is about a loved one. However, there is a certain threshold of attention that we are expected and demanded to pay when assessing the truth of a proposition about *any* person. More precisely, if we hold a belief about someone, we are supposed *not* to believe *falsely* about them. For instance, we are expected not to think the worst of anyone too easily. This seems to be part of what it is to be in a relationship of mutual respect with fellow members of our society.

I think this might offer a way of understanding what goes wrong with prejudice. When you think that someone is dishonest or dangerous because they belong to a certain ethnicity, you are wronging them as a person insofar as you are not treating them as an *individual*.[37] You are effectively

[37] On this view, it's also problematic when people assumed that someone is a fast runner *just* because they're black, or that someone else must be good at maths *just* because they are Asian. These beliefs are wrong even if they are true and, in terms of their content, they are positive. I take this to be a felicitous implication of my proposal.

treating them as a specimen of a certain class, over which it is possible to generalise from a small (and arbitrary) pool of examples. The individual differences are levelled. Even more importantly, the problem is that the risk of being wrong, *i.e.* of believing something about someone who belongs to a certain ethnicity and yet is not dishonest or dangerous, is not sufficiently weighty for you to stop and assess whether *this particular person* is. Prejudices display that someone does not care about getting things right about another person's conduct, attitudes and character. Those who are prejudiced think that something is likely to be true, and that likelihood suffices for them to decide on any particular case.

Again, it's impossible to codify what it would be good for us to believe in each and every situation. We cannot make a list of evidence we should consider before settling the question whether *p*, where *p* is about someone. That will have to be decided on a case-by-case basis. The point is that, if we are convinced that someone did something *because they are black*, or *because they are a woman*, then we are clearly failing to treat them as a person. For any person is much more complex than their gender and race, and, if we are ready to generalise over those traits, then we are essentially treating them as objects. We are showing that we do not care about getting things right about them.

I think we owe it to other persons to care about getting things right about them. Or rather, we owe it to them not to get things wrong. That is, when we hold beliefs about someone with whom we don't have any particular relationship, we are subjected to less demanding expectations. *If* I believe something about them, then I need to be accurate about it. However, I might also not form a belief about them at all. We have limited cognitive and emotional capacities and it would be impossible for us to believe things about everyone else, let alone truly. In effect, I don't really care about believing truly that either Angelina Jolie or Brad Pitt is the most blameworthy for the failure of their glamorous marriage. However, insofar as they are persons in the flesh as much as I am, I owe it to them to suspend my judgment. I should do so precisely because it is highly probable that I won't get it right, if I try to settle the question. And I owe it to them even if they will never know whether I do so or not.

Whilst treating the case of friendship, I wanted to remain neutral on whether the expectations and demands that are constitutive of friendship (or certain jobs) are necessarily a moral matter. For instance, if I fail to give you the benefit of the doubt, I might fail to be a good friend but not necessarily be an immoral person. Here instead, it seems to me that there is a minimal amount of care that we need to pay in assessing truths about *any* other human being. So, every time we don't care not to get things wrong when forming a belief about someone, we wrong them morally. We do not respect them as persons. We might say that these demands and expectations are constitutive of what it is to be in a relationship of mutual respect with other human beings. But

unlike other kinds of relationships, this one is inescapable. We are *always* subjected to the demand not to get things wrong about others, when we believe something about them.[38]

5. An aretaic account

In section 1, I introduced HYPOTHESIS:

> [HYPOTHESIS] If there is something like a practice of holding each other accountable for our beliefs, then we can illuminate it by a virtue-centred, person-focused and not norm-based account.

In sections 3 and 4, I offered two ways of understanding what it could be to hold someone accountable for their beliefs, *i.e.* two ways of understanding the idea that sometimes[39] we wrong or do right by others by believing. I expressed my preference for the second view, which I think makes better sense of the fact that the belief *itself* constitutes an instance of disrespect of mutual regard. The aim of this section is to gesture at a possible way of illuminating the practices of accountability for beliefs within the aretaic model, without buying into the accountability view. Here, I focus on beliefs *about* others but the explanation I offer should be able to illuminate beliefs that wrong or do right by others in general, however one understands their nature.

Suppose that the advocate of the accountability view agrees that sometimes I wrong you by holding unjustified beliefs about you. Plausibly, they would explain this by reference to the fact that, in holding unjustified beliefs about you, I violate norms with which I could and should have complied. First, I violate norms of rationality in forming beliefs that are not conclusively supported. Second, I violate moral norms;[40] I fail to pay you the respect that I should (morally) pay you. You can resent me on the basis that I didn't abide by norms to which both you and I are subjected.

---

[38] Besides beliefs that dispose someone to act immorally and unjustified beliefs about others, there's also a third way of understanding how beliefs can wrong: we might wrong others also whilst we engage *with* them in the activities of belief formation. For instance, we can wrong others when we fail to consider them as capable players in the game of testimony. I cannot do justice to the richness of the topic here, which has received increasing attention especially after Miranda Fricker's work on testimonial injustice (Fricker 2007). As Marusic and White puts it, "in shared reasoning doxastic wronging could be understood in terms of a failure to meet a standard of mutual answerability" (Marušić and White 2018, 99). The core idea is that these standards are met when we treat others as *informants* rather than merely as *sources of information*. Treating someone as an informant is something that we can and *should* do by default: we shouldn't believe others insofar as we have evidence (or we lack evidence against the fact) that they are reliable informants. Rather, we should take what they say as bearing on the truth of *p* insofar as it is the exercise of their rational powers. In effect, I suspect that this kind of approach could be seen as a species of the view I discuss in section 3: in the end, when we fail to treat others as informants, we do so in light of some beliefs that we hold *about* them.

[39] It's perhaps worth stressing once more that the practices of accountability for beliefs don't seem to pin down the central instances of responsibility for beliefs after all.

[40] The advocate of the accountability view would have to explain precisely how we violate *moral* norms by believing, which is not trivial. Even granted that they can do so, the point is that the resulting explanation would still fail to be sufficiently deep.

The main problem with this approach is that it seems unable to provide the *bedrock explanation* as to why your belief counts as a wrong. The fact that you didn't respect me 'as you should have' does not provide a sufficiently deep story; it doesn't clarify why you didn't and why that matters. Furthermore, this view makes all instances of wronging look similar. If I wrong you (by believing something unjustified about you) insofar as I violate certain norms, then I wrong you in the same way and I am open to the same kind of criticism *whatever* I believe about you, as long as it manifests my failure to believe as I should have. So, for instance, I'd wrong my housemate in believing that they damaged my plant as I'd wrong my sister in believing that she doesn't deserve the popularity she gets, and I'd be open to the same criticism in both cases. Yet, it seems that I would be assessed in substantially different ways, in the two situations.

Of course, the aretaic model is not committed to rejecting the importance of norms. It might well be that believers sometimes violate norms, in believing as they do. The point is that we should ask *why* someone violated any, if they did. We should shift the focus from the violation of norms to the *person* who believes certain things in a certain way. The aretaic model attempts to provide a deeper and more nuanced explanation of the practices of accountability for beliefs by exploring the promises of this shift.

More precisely, I suggest that we can explain the practices of accountability for beliefs from a virtue-centred perspective by distinguishing between *personal*[41] and *interpersonal* virtues, or between personal and interpersonal values more generally.[42] The practices of attributability for beliefs I explored in the first three chapters are concerned mostly with personal traits. These are the ones that make everyone's evaluative orientation distinctively their own. My sensitivity to reasons is oriented by values that are likely to be different from the ones that orient yours. I might be someone who has curiosity and kindness amongst their values, whilst you might have optimism and determination amongst yours, and this explains why we see things in different ways. Importantly, it's 'no one else's business' whether you or I have a personal trait. Of course, aretaic appraisals are evaluative judgements and, on their basis, you can decide whether to interact with me or not. You may find my curiosity attractive or annoying. And I can prefer to engage with kind people and avoid pessimistic persons. But I don't *owe* it to you to stop being curious, and no one owes it to me

[41] As already noted, of course also the traits belonging to this set will affect other people's lives, especially of one's loved ones. The distinction is not drawn at the level of the consequences that having a trait brings about. Rather, it concerns the reasons why someone could be asked to modify them. Only interpersonal traits are such that they could (and should) be modified because the person owes it to someone else to do so.

[42] Recall that in chapter II, I argued that values (and lack of values and anti-values) orient our evaluative orientation, and that virtues (and vices) are *amongst* those. In being funny, optimistic, bigoted, or old-fashioned, we display values that are neither virtues nor vices. I take the distinction between personal and interpersonal virtues that I discuss here to apply also to values, broadly construed. Thus, in the text I talk of 'trait' as a placeholder for virtue, value, vice, anti-value and lack of value.

to be kind—I cannot demand or expect that they are. In being unkind, they don't wrong me. I don't have a say in how kind they are.

By contrast, I want to say, the practices of accountability for beliefs are concerned with *interpersonal* virtues, such as honesty or justice. When I hold you accountable for believing something, I take your belief to express that you possess or lack one of the interpersonal traits that are essential to being in the relationships of mutual regard that underlie the practices of accountability. I assess you for being a person who values and respect mutual regard, or who fails to do so. Because of this, interpersonal virtues are such that we can expect and demand other people to have them. For instance, we can expect that everyone sees the fact that something is unjust as a bad reason for them to do something. If they don't master the value concepts 'just' or 'honest', we might still expect them to see that something gathered by hearsay is insufficient to believe the worst of another person. Or we might expect that everyone sees the fact that I accidentally dropped my purse as a good reason to believe that I lost it and should have it returned, rather than as a good reason to believe that it's their chance to make some extra cash.

Crucially, the kind of expectations we can impose on others when interpersonal traits are involved are not available to us when it comes to personal traits. If I insist that what you believe 'is so unjust!', that's meant to provide you with a reason to revise your way of seeing things, a reason that you should be able to see. In a strong sense, *it is* my business if you are unjust because we are part of the same community, and your injustice prevents you from honouring the mutual regard that is essential to it. By contrast, I can *invite* you to modify a belief on the basis that 'it's so pessimistic of you!', but that wouldn't suffice to justify why you should revise it. I'd need to explain also why being so pessimistic is problematic, or why it's *my* business to begin with.

5.1. Relationships and demands

The distinction between personal and interpersonal traits is not clear-cut. For one thing, only few traits seem unequivocally interpersonal, so that I can always expect that others have or value them. Besides justice and honesty, there are perhaps few more. Given that these traits are essentially interpersonal, we are in the realm of accountability as soon as one of them is concerned. At the other end of the spectrum, just few traits are unequivocally personal, so that no one ever owes it to me to have or not have one of them. For example, it's hard to imagine a case where I hold you accountable for being overzealous; that can hardly be an interpersonal matter. Being too harsh with oneself, or being insecure, are not traits that someone should ever modify *for me*. No modification of them can be *owed* to me.

Instead, most traits count as personal or interpersonal according to the circumstances. More precisely, according to, first, how my trait is displayed and, second, what relationship holds between those who assess me and myself. As for the former point, consider curiosity. Sometimes I can appraise you aretaically and some other times I can hold you accountable for being *curious*, according to *what* you are curious about; according to what expresses your curiosity. For instance, if you are curious about peculiarities of Australian insects, I cannot demand or expect that you stop. At most, I can praise your interests or deem them ridiculous. But if you are extremely nosy about me and my life, I can take your curiosity to wrong me and violate my personal sphere, and I can demand you stop being so curious.

In effect, pinning down precisely what I can expect and demand from you in the latter case is a hard question. It's not obvious what you owe me. For one thing, if the claim is that you wrong me by believing (and not only by doing) something, then I need to be able to demand more than simply that you stop *acting* in a certain way, *e.g.* spying on me, inquiring into my business, etc. Of course, I can also expect that you don't act in those ways, but these expectations fall within the practices of accountability *for actions*. Rather, if you wrong me *by believing*, insofar as your beliefs display certain traits or lack of respect towards me, then the relevant expectation and demands will have to be concerned either with your beliefs or with the traits expressed by them. According to a weaker view, I can expect and demand that you stop holding the unkind belief that wrongs me. I can make a case for the idea that you should revise it; you owe it to me as a person[43] to get things right about me. According to a stronger view, I can also expect and demand that you stop being so unkind to begin with. It's not sufficient that you modify your belief. You owe it to me that you modify your attitude towards the world (and me), and that you change your belief because you change your way of seeing things. On this stronger view, the demands and expectations we place on others are deeper than simple requests for due compensations. We don't just expect and demand others to make up for their wrongdoing, *e.g.* by rectify their beliefs; we expect and demand them to revise their values.

Whether I can impose demands on you in this stronger sense depends mostly, I think, on the nature of our relationship. To clarify, above I suggested that I can always demand that you modify unequivocally interpersonal traits and never demand that you modify unequivocally personal traits. Here, the point is that sometimes I can demand that you modify a trait that is *typically* personal, if we are in certain kinds of relationship. For instance, I can demand you stop being so unkind if you are my partner or son. You owe it to me to be kinder and to have kinder beliefs

[43] I will qualify this thought in a moment, whilst discussing the importance of the relationship that holds between us.

about me. I can expect that you do. However, I cannot demand that you are kinder or more generous if you are my colleague or neighbour.

This sheds new light also on some of the cases I discussed in chapter I and III. For instance, recall PLANT: I return home and notice that my plant has been damaged, and I come to believe that you, my careful friend, caused the damage. You are angry at me and resent the fact that I came to believe something false about you so easily; that I settled for an explanation that puts you in a bad light. The key point here is that the fact that we are friends *both* raises the threshold of what counts as a 'sufficiently' supported belief *and* allows you to expect and demand that I change my evaluative outlook. Given our relationship, not only should I have paid extra attention to make sure not to form a false belief about you, which you can expect me to drop. Also, you can expect me to stop being so uncharitable or distrustful. You might request that I modify my way of seeing things, which is what brought about my unjustified belief. If we weren't friends, at most you could demand that I stop believing falsities about you but not that I modify my values.

The extent to which we can expect someone to modify their values is proportionate to the role they play in our lives. This is so because different kinds of relationships require that the values of the ones involved are more or less aligned. For instance, recall Bigotgate from chapter II. After a strenuous conversation with Mrs. Duffy, Gordon Brown comes to believe that she is a bigoted woman. I discussed the case at length as a central instance of the practices of attributability. However, from the point of view of Mrs. Duffy herself, Mr. Brown is to be held *accountable* for his belief insofar as it's a belief that wrongs her. In effect, we might think she is wrong in thinking that he is accountable for his belief. But the point here is that, if Mrs. Duffy resents Mr. Brown and expects his apology, it's especially because he's a candidate for Prime Minister. As such, he should be respectful of his voters, since he purports to represent them at the governmental level. Because of this, Mrs. Duffy and many others want their values to be reflected by him. By contrast, neither Mrs. Duffy nor her allies would expect her life coach to apologize for believing she's bigoted, whilst filing the preliminary assessment of her case.

5.2. Fairness

I think that also concerns about fairness can be illuminated by looking at the relationships between believers. In effect, fairness figures in the practices of accountability in two different ways. First, it could be unfair for me to believe something about you; second, it could be unfair for you to expect and demand that I believe otherwise. In both directions, you and I somehow fail in the game of mutual regard: I owe you to believe more respectfully; you owe me not to place unreasonable expectations and demands on me. Both kinds of failure, I want to say, depend on the fact that we

haven't tailored appropriately our beliefs and expectations to the kind of relationship that holds between us. In fact, both what counts as 'believing respectfully' and what you can expect of me will vary according to how we are related to each other. For instance, you can insist that it is unfair for me to believe that you cheated *because we are friends*, or to hold a certain other belief about you *because I don't know you*. You would draw attention to the fact that you deserve a better, more respectful treatment. And you would ground your complaints on the fact that we are in a certain kind of relationship. Likewise, it would be unfair (and ridiculous) for me to expect that Angelina Jolie believes that I am doing a PhD in Philosophy. I cannot demand that she holds that belief, she doesn't owe it to me.

Above, I have expanded on what it is to believe respectfully and fairly, and I already said something on fair expectations and demands. Let me just emphasise a few points here. First, regardless of our relationship, it will always be unfair for you to expect that I modify an unequivocally personal trait, *e.g.* that I am simple-minded or funny. I don't owe it to you to be otherwise. If you expect it from me, you'd be committing a category mistake: you'd mistake a personal trait for an interpersonal one. You'd fail to understand that being simple-minded or funny is not a kind of trait that can be modified because a modification is owed to someone else.

Second, it would be unfair for you to hold me accountable for having an interpersonal trait that I don't have. This can happen if you misunderstand me. In effect, often we interpret others' beliefs as displaying attitudes towards us that are worse than the ones actually displayed. For instance, you might take my belief to express a tendency to shame or belittle you, when I'm simply gullible or curious. Or you might take my laziness and lack of self-confidence as disregard. In these cases, *you* would be wronging me. For you wouldn't be paying sufficient care in getting things right about me. You would be the one who judges me hastily and fail to treat me as a well-rounded person who deserves a fair assessment.[44]

Third, recall my discussion of the one-instance worry in chapter II. Suppose that I am a kind person who's unkind sometimes. I think it wouldn't be unfair to hold me accountable for being unkind towards you sometimes, if we are in a relationship that allows for you to have expectations over my kindness. However, it would be unfair for you to infer that I am *an unkind person overall*, and to hold me accountable *for that*. In practice, the difference between holding me accountable for being unkind sometimes and for being an unkind person overall lies in the kind of judgements you'd have about me and in the kind of treatment you'd reserve for me.[45] The

[44] In chapter IV, I suggested that unfairness has to do especially with those ascriptions of traits that return an unjust portrait of someone's evaluative orientation. In some cases, those ascriptions might be mistaken on a certain occasion and yet not necessarily unfair, for they don't ascribe to the person values that they don't have.

[45] I lean towards the idea that something similar may occur when we hold accountable someone who's paranoid, or a child. It's not that we wouldn't hold them accountable because they cannot help it. We would hold them accountable,

difference would both be both quantitative and qualitative. For instance, you could be annoyed in both cases, but *more* annoyed in the latter. And you could be sick of me and willing to detach yourself from me in the latter scenario, but not in the former. In general, if you take me to be someone who's vicious in some circumstances, or about certain topics, your assessment of me will be milder than your assessment of those you think are *always* vicious.

5.3. Interpersonal traits and morality

I want to conclude by highlighting one last point about accountability and morality. I said that, when I hold you accountable for believing something, I take your belief to express that you possess or lack one of the traits that I can expect and demand you to have. Some of these traits are essentially interpersonal, so that I can always expect you to have them. Others are such that I can expect you to have them only insofar as we are in a certain relationship. Given that the core of morality lies in the relationships of mutual regard, then it seems that all the traits involved in the practices of accountability belong to the moral realm. That is, it might seem that interpersonal traits are all moral.

This might be so. However, even if we accept this claim, it seems to me crucial to emphasise that they will belong to the moral realm *in different ways*. 'Regard' comes in different shapes, and we need a fine-grained account of it, if we want to provide a detailed picture of the practices of accountability. Whenever I hold you accountable, I take your belief (or conduct) to indicate that you respect or fail to respect others appropriately. I deem you to be a person of such a kind that you value or fail to value sufficiently someone else's well-being. Nevertheless, the way in which you wrong me when you are dishonest is different from the way in which you wrong me when you are arrogant, or gullible, or unjust. In each case, I might be angry, sad, disappointed, or afraid. In wronging me, you might disrespect my intelligence, cast doubts on my capacities, underestimate my value. If my pride is damaged, I might feel embarrassment and anger. But if my trust is affected, I might feel insecure and lonely. The fact that you wronged me will change the way I see you; sometimes I'll take you to be mean, others to be envious, yet others to be insensitive or careless. The ways in which you can fail to have regard for interpersonal relationships are many and diverse.

The possible scenarios are countless, and simply grouping them together on the basis that they are all concerned with the relationships of mutual regard risks levelling their distinctive features. My hope is to have shown that a virtue-centred approach can provide us with the refined tools required to illuminate some of the differences between them.

---

but *in a different way* than we'd hold someone else accountable. That is, the fact that someone does not see a certain belief as potentially wrong does not excuse them from the practices of accountability but it does affect (and mitigate) the way in which we think of and treat them.

Concluding remarks

Gary Watson is often mentioned for having distinguished between two *faces* of responsibility, 'accountability' and 'attributability'. Yet, many questions surrounding this distinction are still a matter of controversy. There's no agreement as to what precisely grounds it; what 'reactive attitudes' are appropriate in each context, if they are; whether the two faces are irreducible. Furthermore, there seems to be a strong tendency in the debate to consider accountability to be the primary, most central, way in which we are (and deem others) responsible for something. Attributability, if we accept it as a genuine mode of responsibility at all, is a thinner phenomenon, which belongs to the realm of responsibility only tangentially. When I am attributability-responsible for an action or a belief, that action or belief can be predicated of me, and it can tell you something important about me, but I am not responsible for it in *as strong* a sense. In this thesis, I have attempted to redress this asymmetry by making a case for the idea that the practices of attributability for beliefs are far from being a marginal phenomenon.

To begin with, I clarified the reference of 'accountability' and 'attributability' (chapter I). In fact, the distinction can be taken to individuate either two practices, *i.e.* of holding someone accountable and appraising someone aretaically, or two theoretical frameworks, *i.e.* two ways in which philosophers think about what it is to be responsible. I take the first distinction to be a genuine and important one, and the very one with which Watson was concerned. 'Accountability' and 'attributability' capture two ways of being responsible for something that are irreducible to each other insofar as they are underwritten by different concerns. We hold someone accountable for something when it wrongs or does right by someone. We are always accountable *to* someone else. By contrast, we appraise someone aretaically for something insofar as it displays what kind of person they are. And importantly, this doesn't apply to someone's actions only: we deem others responsible also for their beliefs according to what they manifest of them, regardless of whether they wrong us or violate norms in believing as they do.

Yet, the practices of attributability for beliefs cannot be easily explained by the most common theory of what it is to be responsible, *i.e.* 'the accountability view'. The accountability view comes in different flavours but typically has it that someone is responsible for something insofar as they have abided by or violated the relevant norms that govern it, whose compliance was in their control. I argue that the accountability view is modelled on the practices of accountability for actions and thus it might provide a satisfactory explanation of *those*, but it falls short when employed to illuminate attributability for beliefs.

In light of this, I have articulated an aretaic model of responsibility as an alternative theoretical framework to the accountability view (chapter II). The core idea is to shift the focus from the particular action or attitude for which someone is responsible to *the person* who acts and believes in certain ways. You are not responsible for something insofar as it counts as a violation or an observance of some relevant norms but rather insofar as it displays that you are a certain kind of person; insofar as it displays your evaluative orientation. Your evaluative orientation is your distinctive sensitivity to practical and theoretical reasons, oriented by your values.

In chapter III, I develop the notion of 'evaluative orientation' in aretaic terms. Amongst the values that orient our sensitivity to reasons, there can be virtues like honesty, kindness and generosity, which call for commendation when displayed by our beliefs. Conversely, we are open to criticism for the beliefs that display a vice, which I understand as either a *lack* of value or as an *anti*-value. Finally, we also have values that are neither virtues nor vices, and sometimes we are open to criticism and sometimes to commendation for the beliefs that display them. The resulting virtue-centred framework opens up rich and varied normative dimensions of assessment. Believers are not just *epistemically* responsible, and epistemically criticisable or commendable. In fact, the accountability view tends to depict the phenomenon this way because it explains responsibility by reference to compliance with and violation of relevant norms that, in the case of beliefs, are thought to be *epistemic* in nature. However, once we resist a norm-based account, we have no reason to reduce *doxastic* responsibility onto *epistemic* responsibility.

To strengthen this line, I openly reject a taxonomy of virtues and vices that mimics the one into which norms are typically divided: it's not straightforward why we should accept that there are epistemic and moral (or prudential and aesthetic) virtues and vices. For one thing, it does not seem possible to draw such distinctions, given that each of the traits that we individuate as a 'virtue' or 'vice' is far more complex that any formulation of a particular norm. But for another, more interesting, thing, it might not even be a *desirable* route. In fact, one of the charms of reasoning in terms of virtues and vices is precisely that they allow for a plurality of normative dimensions of assessments, which may overlap with each other and yet remain distinctively unique. We are responsible for our beliefs in as many different ways as there are normative dimensions of assessments. And we seem to lack any good reason to reduce this plurality to the *moral* and *epistemic* realms only.

In chapter IV, I discuss the control requirement, according to which someone has to be able to exercise a certain amount of control over something in order to be responsible for it. The control requirement is usually taken to be a necessary ingredient of any account of responsibility. The aretaic model I have outlined casts some doubt on its force. In fact, the control requirement seems

to follow from the acceptance of certain, not inescapable, assumptions, *e.g.* that the case of action is the paradigmatic one, that we should understand the notion of 'voluntariness' (and its conceptual connection with responsibility) in a certain way, and that it would be unfair to criticise someone for something over which they lacked control. By drawing on Michael Frede's reading of Aristotle's notion of '*hekontes*', I suggest that we understand the conditions of responsibility and fairness without appealing to control.

Finally, in chapter V I argue for the idea that, whilst the accountability view encounters some difficulties in explaining attributability for beliefs, the aretaic model can make sense of *both* the practices of attributability and of accountability for beliefs. I develop the idea by drawing a distinction between personal and interpersonal virtues and vices. It's essential to interpersonal virtues that we can demand and expect that each of us has them amongst their values, and that we can complain in the relevant way if they do not.

In this thesis, I have attempted to draw attention to a phenomenon that is as familiar in everyday life as it is fairly neglected in the literature: the practices of attributability for beliefs. Given the vastness of the subject and its intersections with multiple debates, I have not aimed at completeness. Much more should be said on each of the issues I touch upon. My hope is to have indicated where I believe it would be worth exploring much more.

Bibliography

Adams, Robert. 1984. 'The Virtue of Faith'. *Faith and Philosophy* 1 (1): 3–15. https://doi.org/10.5840/faithphil1984115.

———. 1985. 'Involuntary Sins'. *The Philosophical Review* 94 (1): 3. https://doi.org/10.2307/2184713.

Adler, Jonathan E. 2002. *Belief's Own Ethics*. Cambridge: MIT Press. http://cognet.mit.edu/book/belief%E2%80%99s-own-ethics.

Ahlstrom-Vij, Kristoffer. 2013. 'In Defense of Veritistic Value Monism'. *Pacific Philosophical Quarterly* 94 (1): 19–40. https://doi.org/10.1111/j.1468-0114.2012.01442.x.

Aikin, Scott F. 2008. 'Evidentialism and James' Argument From Friendship'. *Southwest Philosophy Review* 24 (1): 173–80. https://doi.org/10.5840/swphilreview200824133.

Alston, William P. 1988. 'The Deontological Conception of Epistemic Justification'. *Philosophical Perspectives* 2: 257–99. https://doi.org/10.2307/2214077.

Alvarez, Maria. 2010. *Kinds of Reasons: An Essay in the Philosophy of Action*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199550005.001.0001.

———. 2018. 'Reasons for Action, Acting for Reasons, and Rationality'. *Synthese* 195 (8): 3293–3310. https://doi.org/10.1007/s11229-015-1005-9.

Aristotle. 2009. *The Nicomachean Ethics*, translated by W. D. Ross, Oxford: Oxford University Press.

Anscombe, G. E. M. 1958. 'Modern Moral Philosophy'. *Philosophy* 33 (124): 1–19. doi:10.1017/S0031819100037943

———. 1963. *Intention*. 2nd ed. Oxford: Blackwell.

Appiah, Kwame Anthony. 2008. *Experiments in Ethics*. Cambridge, Harvard University Press. https://doi.org/10.2307/j.ctv19m61c9.

Aquinas, Thomas. 2010. *Summa Theologiae*. Translated by Fathers of the English Dominican Province. Claremont: Coyote Canyon Press.

Arpaly, Nomy. 2003. *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford; New York: Oxford University Press.

Audi, Robert. 2001. 'Doxastic Voluntarism and the Ethics of Belief'. In *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, edited by Matthias Steup, 93-114. New York: Oxford University Press. https://doi.org/10.1093/0195128923.003.0007.

———. 2008. 'The Ethics of Belief: Doxastic Self-Control and Intellectual Virtue'. *Synthese* 161 (3): 403–18. https://doi.org/10.1007/s11229-006-9092-2.

Axtell, Guy. 2008. 'Expanding Epistemology: A Responsibilist Approach'. *Philosophical Papers* 37 (1): 51–87. https://doi.org/10.1080/05568640809485214.

Baehr, Jason. 2008. 'Four Varieties of Character-Based Virtue Epistemology'. *The Southern Journal of Philosophy* 46 (4): 469–502. https://doi.org/10.1111/j.2041-6962.2008.tb00081.x.

Basu, Rima. 2018. 'Can Beliefs Wrong?' *Philosophical Topics* 46 (1): 1–17. https://doi.org/10.5840/philtopics20184611.

Basu, Rima, and Mark Schroeder. 2019. 'Doxastic Wronging'. In *Pragmatic Encroachment in Epistemology*, edited by Brian Kim and Matthew McGrath, 181–205. Abingdon: Routledge.

Battaly, Heather. 2008. 'Virtue Epistemology'. *Philosophy Compass* 3 (4): 639–63. https://doi.org/10.1111/j.1747-9991.2008.00146.x.

Bird, Alexander. 2007. 'Justified Judging'. *Philosophy and Phenomenological Research* 74 (1): 81–110. https://doi.org/10.1111/j.1933-1592.2007.00004.x.

Bond, E. J. 1983. *Reason and Value*. Cambridge: Cambridge University Press.

Booth, Anthony, and Rik Peels. 2010. 'Why Responsible Belief Is Blameless Belief'. *Journal of Philosophy* 107 (5): 257–65. https://doi.org/10.5840/jphil2010107519.

Booth, Anthony Robert. 2014. 'On Some Recent Moves in Defence of Doxastic Compatibilism'. *Synthese* 191 (8): 1867–80. https://doi.org/10.1007/s11229-013-0378-x.

Boyle, Matthew. 2009. 'Active Belief'. *Canadian Journal of Philosophy Supplementary Volume* 35: 119–47. https://doi.org/10.1080/00455091.2009.10717646.

———. 2011. 'Making up Your Mind' and the Activity of Reason'. *Philosopher's Imprint* 11 (17): 1–24. http://hdl.handle.net/2027/spo.3521354.0011.017

Brink, David O., and Dana K. Nelkin. 2013. 'Fairness and the Architecture of Responsibility'. In *Oxford Studies in Agency and Responsibility Volume 1*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199694853.003.0013.

Broadie, Sarah. 1991. *Ethics with Aristotle*. Oxford University Press.

Brogaard, Berit. 2014. 'Wide-Scope Requirements and the Ethics of Belief'. In *The Ethics of Belief*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199686520.003.0008.

Broome, John. 1997. 'Reasons and Motivation'. *Aristotelian Society Supplementary Volume* 71 (1): 131–46. https://doi.org/10.1111/1467-8349.00022.

Brown, Jessica. 2020. 'What Is Epistemic Blame?' *Noûs* 54 (2): 389–407. https://doi.org/10.1111/nous.12270.

Brueckner, Anthony, and Alex Bundy. 2012. 'On "Epistemic Permissiveness"'. *Synthese* 188 (2): 165–77. https://doi.org/10.1007/s11229-011-9921-9.

Buckareff, Andrei A. 2014. 'Deciding to Believe Redux'. In *The Ethics of Belief*, edited by Jonathan Matheson and Rico Vitz, 33–50. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199686520.003.0003.

Chisholm, Roderick. 1966. *Theory of Knowledge*. Englewood Cliffs: Prentice-Hall International.

Chrisman, Matthew. 2018. 'Epistemic Normativity and Cognitive Agency'. *Noûs* 52 (3): 508–29. https://doi.org/10.1111/nous.12184.

Clarke, Bridget. 2017. *Virtue as a Sensitivity*, edited by Nancy E. Snow. Vol. 1. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199385195.013.12.

Clifford, William K. 1877. 'The Ethics of Belief'. In *The Ethics of Belief and Other Essays*, 70–97. Amherst, New York: Prometheus Books.

Code, Lorraine. 1984. 'Toward a 'Responsibilist' Epistemology'. *Philosophy and Phenomenological Research* 45 (1): 29. https://doi.org/10.2307/2107325.

———. 1987. *Epistemic Responsibility*. Providence: Brown University Press. http://www.sunypress.edu/p-6971-epistemic-responsibility.aspx.

———. 2004. 'The Power Of Ignorance'. *Philosophical Papers* 33 (3): 291–308. https://doi.org/10.1080/05568640409485144.

Cohen, Alix. 2013. 'Kant on Doxastic Voluntarism and Its Implications for Epistemic Responsibility'. *Kant Yearbook* 5 (1). https://doi.org/10.1515/kantyb.2013.5.1.33.

Comesaña, Juan, and Matthew McGrath. 2014. 'Having False Reasons'. In *Epistemic Norms*, edited by Clayton Littlejohn & John Turri, 59-80. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199660025.003.0004.

Côté-Bouchard, Charles. 2016. 'Can the Aim of Belief Ground Epistemic Normativity?' *Philosophical Studies* 173 (12): 3181–98. https://doi.org/10.1007/s11098-016-0657-8.

———. 2017. 'Is Epistemic Normativity Value-Based?' *Dialogue* 56 (3): 407–30. https://doi.org/10.1017/S0012217317000506.

Cowie, Christopher, and Alexander Greenberg. 2018. 'Constitutivism about Epistemic Normativity.' In *Metaepistemology: Realism and Anti-Realism*, edited by Christos Kyriacou and Robin McKenna, 173–96. London: Palgrave Macmillan. https://doi.org/10.1007/978-3-319-93369-6_8.

Crawford, Lindsay. 2020. 'Relationships and Reasons for Belief'. In *The Ethics of Belief and Beyond. Understanding Mental Normativity*, edited by Sebastian Schmidt and Gerhard Ernst, 87–108. Abingdon: Routledge.

Dancy, Jonathan. 2000. *Practical Reality*. Oxford: Oxford University Press.

———. 2014. 'On Knowing One's Reason'. In *Epistemic Norms*, edited by Clayton Littlejohn & John Turri, 81-96. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199660025.003.0005.

Darwall, Stephen. 1996. *The Second Person Standpoint: Morality, Respect, and Accountability*. Cambridge: Harvard University Press.

Darwall, Stephen L. 1983. *Impartial Reason*. Ithaca, New York: Cornell University Press.

David, Marian. 2001. 'Truth as the Epistemic Goal'. In *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, edited by Matthias Steup, 151–70. New York: Oxford University Press.

Davidson, Donald. 1963. 'Actions, Reasons, and Causes'. *The Journal of Philosophy* 60 (23): 685. https://doi.org/10.2307/2023177.

———. 1969. 'How Is Weakness of the Will Possible?' In *Moral Concepts*, edited by Joel Feinberg. Oxford: Oxford University Press.

Deigh, John. 2011. 'Reactive Attitudes Revisited'. In *Morality and the Emotions*, edited by Carla Bagnoli, 196–216. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199577507.003.0010.

DePaul, Michael R. 2001. 'Value Monism in Epistemology'. In *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, edited by Matthias Steup, 170-86. New York: Oxford University Press. https://doi.org/10.1093/0195128923.003.0011.

Dougherty, Trent, ed. 2011. *Evidentialism and Its Discontents*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199563500.001.0001.

Douven, Igor. 2009. 'Uniqueness Revisited'. *American Philosophical Quarterly* 46 (4): 347–61. http://www.jstor.org/stable/40606911.

Driver, Julia. 2003. 'The Conflation of Moral and Epistemic Virtue'. *Metaphilosophy* 34 (3): 367–83. https://doi.org/10.1111/1467-9973.00279.

Engel, Pascal. 2005. 'Truth and the Aim of Belief'. In *Laws and Models in Science*, 79–99. London: King's College Publications.

Fairweather, Abrol, and Mark Alfano, eds. 2017. *Epistemic Situationism*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780199688234.001.0001.

Fallis, Don. 2006. 'Epistemic Value Theory and Social Epistemology'. *Episteme* 2 (3): 177–88. https://doi.org/10.3366/epi.2005.2.3.177.

Feldman, Richard. 1988. 'Epistemic Obligations'. *Philosophical Perspectives* 2: 235–56. https://doi.org/10.2307/2214076.

———. 2000. 'The Ethics of Belief'. *Philosophy and Phenomenological Research* 60 (3): 667–95. https://doi.org/10.2307/2653823.
———. 2005. 'Epistemological Duties'. The Oxford Handbook of Epistemology. 2005. https://doi.org/10.1093/oxfordhb/9780195301700.003.0013.
———. 2006. 'Reasonable Religious Disagreements'. In *Philosophers Without Gods: Meditations on Atheism and the Secular Life*, edited by Louise Antony, 194–214. Oxford: Oxford University Press.
———. 2019. 'Voluntary Belief and Epistemic Evaluation'. In *Contemporary Epistemology*, 17–27. Hoboken: John Wiley & Sons. https://doi.org/10.1002/9781119420828.ch2.
Feldman, Richard, and Earl Conee. 1985. 'Evidentialism'. *Philosophical Studies* 48 (1): 15–34. https://doi.org/10.1007/bf00372404.
Fieser, James. 1998. 'Hume's Wide View of the Virtues: An Analysis of His Early Critics'. *Hume Studies* 24 (2): 295–311.
Fischer, John Martin, and Mark Ravizza. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.
Foley, Richard. 1992. *Working Without a Net: A Study of Egocentric Epistemology*. New York: Oxford University Press.
Frankfurt, Harry. 1982. 'The Importance of What We Care About'. *Synthese* 53 (2): 257–72. https://doi.org/10.1007/BF00484902.
———. 1969. 'Alternate Possibilities and Moral Responsibility'. *The Journal of Philosophy* 66 (23): 829–39. https://doi.org/10.2307/2023833.
———. 1988. *The Importance of What We Care About: Philosophical Essays*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511818172.
Frede, Michael. 2012. *A Free Will: Origins of the Notion in Ancient Thought*, edited by A. A. Long. Oakland: University of California Press.
Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford; New York: Oxford University Press.
———. 2009. 'The Value of Knowledge and The Test of Time'. *Royal Institute of Philosophy Supplement* 64: 121–38. https://doi.org/10.1017/s1358246109000034.
Friedman, Jane. 2018. 'Junk Beliefs and Interest-Driven Epistemology'. *Philosophy and Phenomenological Research* 97 (3): 568–83. https://doi.org/10.1111/phpr.12381.
———. 2020. 'The Epistemic and the Zetetic'. *The Philosophical Review* 129 (4): 501–36. https://doi.org/10.1215/00318108-8540918.
Gardiner, Georgi. 2012. 'Understanding, Integration, and Epistemic Value'. *Acta Analytica* 27 (2): 163–81. https://doi.org/10.1007/s12136-012-0152-6.
Gertken, Jan, and Benjamin Kiesewetter. 2017. 'The Right and the Wrong Kind of Reasons'. *Philosophy Compass* 12 (5): e12412. https://doi.org/10.1111/phc3.12412.
Gibbard, Allan. 2005. 'Truth and Correct Belief'. *Philosophical Issues* 15 (1): 338–50. https://doi.org/10.1111/j.1533-6077.2005.00070.x.
Gibbons, John. 2013. *The Norm of Belief*. Oxford; New York: Oxford University Press.
Ginet, Carl. 2001. 'Deciding to Believe'. In *Knowledge, Truth and Duty*, edited by Matthias Steup, 63–76. Oxford: Oxford University Press.
Glover, Jonathan. 1970. *Responsibility*. New York: Humanities Press.

Goldberg, Sanford. 2014. 'Assertion and the Ethics of Belief'. In *The Ethics of Belief*, edited by Jonathan Matheson and Rico Vitz, 261-283. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199686520.003.0015.

Goldman, Alvin I. 1978. 'Epistemics: The Regulative Theory of Cognition'. *Journal of Philosophy* 75 (10): 509–23. https://doi.org/10.2307/2025838.

———. 1999. *Knowledge in a Social World*. Oxford: Oxford University Press. https://doi.org/10.1093/0198238207.001.0001.

———. 2012. *Reliabilism and Contemporary Epistemology: Essays*. New York: Oxford University Press.

Greco, John. 1999. 'Agent Reliabilism'. *Philosophical Perspectives* 13: 273–96. https://doi.org/10.1111/0029-4624.33.s13.13.

Greco, John, and Jonathan Reibsamen. 2018. 'Reliabilist Virtue Epistemology'. In *The Oxford Handbook of Virtue*, edited by Nancy Snow, 725–46. New York: Oxford University Press.

Haack, Susan. 2001. '"The Ethics of Belief" Reconsidered'. In *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, edited by Matthias Steup, 21-33. New York: Oxford University Press. https://doi.org/10.1093/0195128923.003.0002.

Hare, R. M. 1963. *The Language of Morals*. Oxford: Oxford University Press. https://doi.org/10.1093/0198810776.001.0001.

Harman, Elizabeth. 2011. 'Does Moral Ignorance Exculpate?' *Ratio* 24 (4): 443–68. https://doi.org/10.1111/j.1467-9329.2011.00511.x.

Harman, Gilbert. 1986. *Change in View: Principles of Reasoning*. Cambridge: MIT Press.

Hart, H. L. A. 1968. *Punishment and Responsibility: Essays in the Philosophy of Law*. Oxford: Oxford University Press.

Heil, John. 1983. 'Doxastic Agency'. *Philosophical Studies* 43 (3): 355–64. https://10.1007/BF00372372.

Heller, Mark. 2000. 'Hobartian Voluntarism: Grounding a Deontological Conception of Epistemic Justification'. *Pacific Philosophical Quarterly* 81 (2): 130–41. https://doi.org/10.1111/1468-0114.00099.

Heuer, Ulrike. 2018. 'The Relevance of the Wrong Kind of Reasons'. In *Normativity: Epistemic & Practical*, edited by Conor McHugh, Jonathan Way, and Daniel Whiting, 47–67. Oxford: Oxford University Press. https://global.oup.com/academic/product/normativity-9780198758709?q=Epistemic%20and%20Practical%20Normativity&lang=en&cc=gb#.

Hieronymi, Pamela. 2004. 'The Force and Fairness of Blame'. *Philosophical Perspectives* 18 (1): 115–48. https://doi.org/10.1111/j.1520-8583.2004.00023.x.

———. 2005. 'The Wrong Kind of Reason'. *The Journal of Philosophy* 102 (9): 437–57. https://doi.org/10.1002/9781119420828.ch3

———. 2006. 'Controlling attitudes'. *Pacific Philosophical Quarterly* 87 (1): 45–74. https://doi.org/10.1111/j.1468-0114.2006.00247.x.

———. 2008. 'Responsibility for Believing'. *Synthese* 161 (3): 357–73. https://doi.org/10.1007/s11229-006-9089-x.

———. 2009. 'Two Kinds of Agency'. In *Mental Actions*, edited by Lucy O'Brien and Matthew Soteriou, 138–62. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199225989.003.0007.

———. 2014. 'Reflection and Responsibility'. *Philosophy & Public Affairs* 42 (1): 3–41. https://doi.org/10.1111/papa.12024.

Hobart, R. E. 1934. 'Free Will as Involving Determination and Inconceivable Without It'. *Mind* 43 (169): 1–27. https://doi.org/10.1093/mind/XLIII.169.1.

Hookway, Christopher. 1994. 'Cognitive Virtues and Epistemic Evaluations'. *International Journal of Philosophical Studies* 2 (2): 211–27. https://doi.org/10.1080/09672559408570791.

Hornsby, Jennifer. 2008. 'A Disjunctive Conception of Acting for Reasons'. In *Disjunctivism: Perception, Action, Knowledge*, edited by Adrian Haddock and Fiona Macpherson, 244-61. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199231546.003.0011.

Howard, Christopher. 2020. 'Weighing Epistemic and Practical Reasons for Belief'. *Philosophical Studies* 177 (8): 2227–43. https://doi.org/10.1007/s11098-019-01307-y.

Hursthouse, Rosalind. 1991. 'Arational Actions'. *Journal of Philosophy* 88 (2): 57–68. https://doi.org/10.2307/2026906.

———. 1999. *On Virtue Ethics*. Oxford: Oxford University Press. https://doi.org/10.1093/0199247994.001.0001.

Hyman, John. 2015. *Action, Knowledge, and Will*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198735779.001.0001.

James, William. 1897. 'The Will to Believe, and Other Essays in Popular Philosophy'. *Philosophical Review* 6 (3): 331. https://doi.org/10.2307/2175898.

Kahn, Charles H. 1988. 'Discovering Will: From Aristotle to Augustine'. In *The Question of 'Eclecticism': Studies in Later Greek Philosophy*, edited by John M. Dillon and A. A. Long, 235–60. Berkeley: University of California Press.

Kane, Robert. 2002. 'Responsibility, Reactive Attitudes and Free Will: Reflections on Wallace's Theory'. *Philosophy and Phenomenological Research* 64 (3): 693–98. https://doi.org/ppr200264355.

Kauppinen, Antti. 2018. 'Epistemic Norms and Epistemic Accountability'. *Philosopher's Imprint* 18 (8): 16.

Keller, Simon. 2004. 'Friendship and Belief'. *Philosophical Papers* 33 (3): 329–51. https://doi.org/10.1080/05568640409485146.

Kelly, Thomas. 2003. 'Epistemic Rationality as Instrumental Rationality'. *Philosophy and Phenomenological Research* 66 (3): 612–40.

Kenny, Anthony J. P. 1979. *Aristotle's Theory of the Will*. New Haven: Yale University Press. http://archive.org/details/aristotlestheory0000kenn.

Kim, Brian, and Matthew McGrath, eds. 2018. *Pragmatic Encroachment in Epistemology*. Routledge Studies in Epistemology 1. New York: Taylor & Francis.

Kornblith, Hilary. 1983. 'Justified Belief and Epistemically Responsible Action'. *The Philosophical Review* 92 (1): 33–48. https://doi.org/10.2307/2184520.

Korsgaard, Christine. 1997. 'The Normativity of Instrumental Reason'. In *Ethics and Practical Reason*, edited by Garrett Cullity and Berys Gaut, 26-68. Oxford: Clarendon Press.

Kvanvig, Jonathan L. 2003. *The Value of Knowledge and the Pursuit of Understanding*. Cambridge: Cambridge University Press.

———. 2005. 'Truth Is Not the Primary Epistemic Goal'. In *Contemporary Debates in Epistemology*, edited by Matthias Steup and Ernest Sosa, 285–95. Malden: Blackwell Publishing Inc.

Leary, Stephanie. 2017. 'In Defense of Practical Reasons for Belief'. *Australasian Journal of Philosophy* 95 (3): 529–42. https://doi.org/10.1080/00048402.2016.1237532.

Levy, Neil. 2005. 'The Good, the Bad, and the Blameworthy'. *Journal of Ethics and Social Philosophy* 1 (2): 1–16. https://doi.org/10.26556/jesp.v1i2.6.

———. 2007a. 'Doxastic Responsibility'. *Synthese* 155 (1): 127–55. https://doi.org/10.1007/s11229-005-3983-5.

———. 2007b. 'The Responsibility of the Psychopath Revisited'. *Philosophy, Psychiatry, and Psychology* 14 (2): 129–38. https://doi.org/10.1353/ppp.0.0003.

Levy, Neil, and Eric Mandelbaum. 2014. 'The Powers That Bind'. In *The Ethics of Belief*, edited by Jonathan Matheson and Rico Vitz, 15–32. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199686520.003.0002.

Littlejohn, Clayton. 2013. 'The Russellian Retreat'. *Proceedings of the Aristotelian Society* 113: 293–320. http://www.jstor.org/stable/42705836.

———. 2014. 'The Unity of Reason'. In *Epistemic Norms*, edited by Clayton Littlejohn & John Turri, 135-54. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199660025.003.0008.

Lockard, Matthew. 2013. 'Epistemic Instrumentalism'. *Synthese* 190 (9): 1701–18. https://doi.org/10.1007/s11229-011-9932-6.

Marušić, Berislav, and Stephen White. 2018. 'How Can Beliefs Wrong?: A Strawsonian Epistemology'. *Philosophical Topics* 46 (1): 97–114. https://doi.org/10.5840/philtopics20184616.

Matheson, Jonathan. 2011. 'The Case for Rational Uniqueness'. *Logic and Episteme* 2 (3): 359–73. https://doi.org/10.5840/logos-episteme20112319.

McCormick, Miriam. 2011. 'Taking Control of Belief'. *Philosophical Explorations* 14 (2): 169–83. https://doi.org/10.1080/13869795.2011.569745.

———. 2015. *Believing Against the Evidence*. London: Routledge.

McCormick, Miriam, and Michael Schleifer. 2006. 'Responsibility for Beliefs and Emotions'. *Paideusis* 15 (1): 75–85. https://doi.org/10.7202/1072695ar.

McDowell, John. 1979. 'Virtue and Reason'. *The Monist* 62 (3): 331–50.

McHugh, Conor. 2011a. 'What Do We Aim At When We Believe?' *Dialectica* 65 (3): 369–92. https://doi.org/10.1007/s11098-011-9728-z.

———. 2011b. 'Judging as a Non-Voluntary Action'. *Philosophical Studies* 152 (2): 245–69. https://doi.org/10.1007/s11098-009-9478-3.

———. 2012. 'Epistemic Deontology and Voluntariness'. *Erkenntnis* 77 (1): 65–94. https://doi.org/10.1007/s10670-011-9299-6.

———. 2013. 'Epistemic Responsibility and Doxastic Agency'. *Philosophical Issues* 23: 132–57. https://doi.org/10.1111/phis.12007

———. 2017. 'Attitudinal Control'. *Synthese* 194 (8): 2745–62. https://doi.org/10.1007/s11229-014-0643-7.

McHugh, Conor, and Daniel Whiting. 2014. 'The Normativity of Belief'. *Analysis* 74 (4): 698–713. https://doi.org/10.1093/analys/anu079.

Meylan, Anne. 2015. 'The Legitimacy of Intellectual Praise and Blame': *Journal of Philosophical Research* 40: 189–203. https://doi.org/10.5840/jpr201511537.

———. 2017. 'The Consequential Conception of Doxastic Responsibility'. *Theoria* 83 (1): 4–28. https://doi.org/10.1111/theo.12110.

Millar, Alan. 2004. *Understanding People: Normativity and Rationalizing Explanation*. Oxford: Oxford University Press.

Mills, Eugene. 1998. 'The Unity of Justification'. *Philosophy and Phenomenological Research* 58 (1): 27–50. https://doi.org/10.2307/2653629.

Montmarquet, James. 1986. 'The Voluntariness of Belief'. *Analysis* 46 (1): 49–53. https://doi.org/10.2307/3328747.

———. 1987. 'Epistemic Virtue'. *Mind* XCVI (384): 482–97. https://doi.org/10.1093/mind/XCVI.384.482.

———. 1992. 'Epistemic Virtue and Doxastic Responsibility'. *American Philosophical Quarterly* 29 (4): 331–41.

———. 2008. 'The Voluntariness of Virtue and Belief'. *Philosophy* 83 (3): 373–90. https://doi.org/10.1017/S0031819108000739.

Murdoch, Iris. 1971. *The Sovereignty of Good*. Abingdon: Routledge.

Naylor, Margery Bedford. 1985. 'Voluntary Belief'. *Philosophy and Phenomenological Research* 45 (3): 427–36. https://doi.org/10.2307/2107274.

Nelkin, Dana Kay. 2020. 'Equal Opportunity: A Unifying Framework for Moral, Aesthetic, and Epistemic Responsibility'. *Proceedings of the Aristotelian Society* 120 (2): 203–35. https://doi.org/10.1093/arisoc/aoaa010.

Nottelmann, Nikolaj. 2006. 'The Analogy Argument for Doxastic Voluntarism'. *Philosophical Studies* 131 (3): 559–82. https://doi.org/10.1007/s11098-004-7489-7.

———. 2007. *Blameworthy Belief: A Study in Epistemic Deontologism*. Synthese Library, Vol. 338. Dordrecht: Springer.

Owens, David. 2000. *Reason Without Freedom: The Problem of Epistemic Normativity*. Abingdon: Routledge.

———. 2002. 'Epistemic Akrasia'. *The Monist* 85 (3): 381–97. https://doi.org/10.5840/monist200285316.

———. 2012. *Shaping the Normative Landscape*. Oxford: Oxford University Press.

———. 2013. 'Value and Epistemic Normativity'. *Teorema: Revista Internacional de Filosofía* 32 (3): 35–58. http://www.jstor.org/stable/43047010.

———. 2017. *Normativity and Control*. Oxford, New York: Oxford University Press.

Pascal, Blaise. 1905. *The Thought of Blaise Pascal, Translated from the Text of M. Auguste Molinier by C. Kegan Paul...* London: George Bell & Sons. https://catalog.hathitrust.org/Record/100879429.

Peels, Rik. 2011. 'Tracing Culpable Ignorance'. *Logos and Episteme* 2 (4): 575–82. https://doi.org/10.5840/logos-episteme2011246.

———. 2014. 'What Kind of Ignorance Excuses? Two Neglected Issues'. *The Philosophical Quarterly* 64 (256): 478–96. https://doi.org/10.1093/pq/pqu013.

———. 2017. *Responsible Belief: A Theory in Ethics and Epistemology*. New York: Oxford University Press.

Pojman, Louis P. 1986. *Religious Belief and the Will*. Abingdon: Routledge.

Price, H. H. 1954. 'Belief and Will'. *Aristotelian Society Supplementary Volume* 28 (1): 1–26. https://doi.org/10.1093/aristoteliansupp/28.1.1.

Pritchard, Duncan. 2007. 'Recent Work on Epistemic Value'. *American Philosophical Quarterly* 44 (2): 85–110. http://www.jstor.org/stable/20464361.

———. 2011. 'What Is the Swamping Problem?' In *Reasons for Belief*, edited by Andrew Reisner and Asbjørn Steglich-Petersen, 244–59. Cambridge: Cambridge University Press.

———. 2014. 'Truth as the Fundamental Epistemic Good'. In *The Ethics of Belief*, edited by Jonathan Matheson and Rico Vitz, 112-29. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199686520.003.0007.

Proust, Marcel. 1918. *In the Shadow of Young Girls in Flower*, edited by William C. Carter, translated by Charles Kenneth Scott-Moncrieff. 2nd edition. London; New York: Penguin Classics.

Railton, Peter. 1994. 'Truth, Reason, and the Regulation of Belief'. *Philosophical Issues* 5: 71–93. https://doi.org/10.2307/1522874.

Raz, Joseph. 1999. *Engaging Reason: On the Theory of Value and Action*. Oxford; New York: Oxford University Press.

———. 2011. *From Normativity to Responsibility*. Oxford; New York: Oxford University Press.

Reisner, Andrew. 2009. 'The Possibility of Pragmatic Reasons for Belief and the Wrong Kind of Reasons Problem'. *Philosophical Studies* 145 (2): 257–72. https://doi.org/10.1007/s11098-008-9222-4.

———. 2018. 'Pragmatic Reasons for Belief'. In *The Oxford Handbook of Reasons and Normativity*, edited by Daniel Star, 705-729. Oxford: Oxford University Press.

Reisner, Andrew, and Asbjørn Steglich-Petersen, eds. 2011. *Reasons for Belief*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511977206.

Rettler, Lindsay. 2018. 'In Defense of Doxastic Blame'. *Synthese* 195 (5): 2205–26. https://doi.org/10.1007/s11229-017-1332-0.

Rinard, Susanna. 2015a. 'Against the New Evidentialists'. *Philosophical Issues* 25 (1): 208–23. https://doi.org/10.1111/phis.12061.

———. 2015b. 'Believing for Practical Reasons'. *Philosophical Issues* 25 (1): 208–23. https://doi.org/10.1111/nous.12253

———. 2017. 'No Exception for Belief'. *Philosophy and Phenomenological Research* 94 (1): 121–43. https://doi.org/10.1111/phpr.12229.

———. 2019. 'Equal Treatment for Belief'. *Philosophical Studies* 176 (7): 1923–50. https://doi.org/10.1007/s11098-018-1104-9.

Robichaud, Philip, and Jan Willem Wieland, eds. 2017. *Responsibility: The Epistemic Condition*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198779667.001.0001.

Rosen, Gideon. 2001. 'Nominalism, Naturalism, Epistemic Relativism'. *Noûs* 35 (s15): 69–91. https://doi.org/10.1111/0029-4624.35.s15.4.

———. 2004. 'Skepticism about Moral Responsibility'. *Philosophical Perspectives* 18 (1): 295–313. https://doi.org/10.1111/j.1520-8583.2004.00030.x.

———. 2008. 'Kleinbart the Oblivious and Other Tales of Ignorance and Responsibility': *Journal of Philosophy* 105 (10): 591–610. https://doi.org/10.5840/jphil20081051023.

Rott, Hans. 2017. 'Negative Doxastic Voluntarism and the Concept of Belief'. *Synthese* 194 (8): 2695–2720. https://doi.org/10.1007/s11229-016-1032-1.

Russell, Bruce. 2001. 'Epistemic and Moral Duty'. In *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, edited by Matthias Steup, 34-48. New York: Oxford University Press. https://doi.org/10.1093/0195128923.003.0003.

Ryan, Sharon. 2003. 'Doxastic Compatibilism and the Ethics of Belief'. *Philosophical Studies* 114 (1/2): 47–79. https://doi.org/10.1023/A:1024409201289.

Scanlon, T. M. 2012. 'Interpreting Blame'. In *Blame*, edited by D. Justin Coates and Neal A. Tognazzini, 84–99. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199860821.003.0005.

———. 1998. *What We Owe to Each Other*. Cambridge: Harvard University Press.
———. 2013. 'Giving Desert Its Due'. *Philosophical Explorations* 16 (2): 101–16. https://doi.org/10.1080/13869795.2013.787437.
Schmidt, Sebastian. 2020. 'Responsibility for Attitudes, Object-Given Reasons, and Blame'. In *The Ethics of Belief and Beyond. Understanding Mental Normativity*, edited by Gerhard Ernst and Sebastian Schmidt, 149–75. Abingdon: Routledge.
Schoenfield, Miriam. 2014. 'Permission to Believe: Why Permissivism Is True and What It Tells Us About Irrelevant Influences on Belief'. *Noûs* 48 (2): 193–218. https://doi.org/10.1111/nous.12006.
———. 2019. 'Permissivism and the Value of Rationality: A Challenge to the Uniqueness Thesis'. *Philosophy and Phenomenological Research* 99 (2): 286–97. https://doi.org/10.1111/phpr.12490.
Scott-Kakures, Dion. 1994. 'On Belief and the Captivity of the Will'. *Philosophy and Phenomenological Research* 54 (1): 77–103. https://doi.org/10.2307/2108356.
Setiya, Kieran. 2007. *Reasons without Rationalism*. Princeton: Princeton University Press.
Shah, Nishi. 2002. 'Clearing Space For Doxastic Voluntarism'. *The Monist* 85 (3): 436–45. https://doi.org/10.5840/monist200285326.
———. 2003. 'How Truth Governs Belief'. *The Philosophical Review* 112 (4): 447–82. https://doi.org/10.1215/00318108-112-4-447.
———. 2006. 'A New Argument for Evidentialism'. *The Philosophical Quarterly (1950-)* 56 (225): 481–98. https://doi.org/10.1111/j.1467-9213.2006.454.x.
Sher, George. 2001. 'Blame for Traits'. *Noûs* 35 (1): 146–61. https://doi.org/10.1111/0029-4624.00291.
———. 2009. *Who Knew?: Responsibility Without Awareness*. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195389197.001.0001.
Smith, Angela M. 2005. 'Responsibility for Attitudes: Activity and Passivity in Mental Life'. *Ethics* 115 (2): 236–71. https://doi.org/10.1086/426957.
———. 2008. 'Control, Responsibility, and Moral Assessment'. *Philosophical Studies* 138 (3): 367–92. https://doi.org/10.1007/s11098-006-9048-x.
———. 2015. 'Attitudes, Tracing, and Control: Attitudes, Tracing, and Control'. *Journal of Applied Philosophy* 32 (2): 115–32. https://doi.org/10.1111/japp.12107.
———. 2019. 'Who's Afraid of a Little Resentment?' In *Oxford Studies in Agency and Responsibility Volume 6*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198845539.003.0004.
Smith, Holly M. 2011. 'Non-Tracing Cases of Culpable Ignorance'. *Criminal Law and Philosophy* 5 (2): 115–46. https://doi.org/10.1007/s11572-011-9113-1.
Sosa, Ernest. 2000. 'Reliabilism and Intellectual Virtue'. In *Knowledge, Belief, and Character: Readings in Virtue Epistemology*, edited by Guy Axtell, 33–40. Washington: Rowman & Littlefield Publishers.
Statman, Daniel. 1993. *Moral Luck*. New York: SUNY Press.
Steup, Matthias. 2000. 'Doxastic Voluntarism and Epistemic Deontology'. *Acta Analytica* 15 (1): 25–56. https://doi.org/10.1093/0195128923.001.0001
———. 2008. 'Doxastic Freedom'. *Synthese* 161 (3): 375–92. https://doi.org/10.1007/s11229-006-9090-4.
———. 2011. 'Belief, Voluntariness and Intentionality'. *Dialectica* 65 (4): 537–59.

———. 2012. 'Belief Control and Intentionality'. *Synthese* 188 (2): 145–63. https://doi.org/10.1007/s11229-011-9919-3.

Stroud, Sarah. 2006. 'Epistemic Partiality in Friendship'. *Ethics* 116 (3): 498–524. https://doi.org/10.1086/500337.

Sutton, Jonathan. 2007. *Without Justification*. Cambridge: A Bradford Book.

Talbert, Matthew. 2008. 'Blame and Responsiveness to Moral Reasons: Are Psychopaths Blameworthy?' *Pacific Philosophical Quarterly* 89 (4): 516–35. https://doi.org/10.1111/j.1468-0114.2008.00334.x.

———. 2013. 'Unwitting Wrongdoers and the Role of Moral Disagreement in Blame'. In *Oxford Studies in Agency and Responsibility Volume 1*, edited by David Shoemaker, 225–45. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199694853.003.0010.

Velleman, David. 2000. 'On the Aim of Belief'. In *The Possibility of Practical Reason*, edited by David Velleman, 244–81. Oxford: Oxford University Press.

———. 1992. 'The Guise of the Good'. *Noûs* 26 (1): 3–26. https://doi.org/10.2307/2215684.

Vendler, Zeno. 1957. 'Verbs and Times'. *The Philosophical Review* 66 (2): 143. https://doi.org/10.2307/2182371.

Walker, Arthur F. 1989. 'The Problem of Weakness of Will'. *Noûs* 23 (5): 653–76. https://doi.org/10.2307/2216006.

Wallace, R. Jay. 1994. *Responsibility and the Moral Sentiments*. Cambridge: Harvard University Press.

———. 1999. 'Three Conceptions of Rational Agency'. *Ethical Theory and Moral Practice* 2 (3): 217–42. https://doi.org/10.1023/A:1009946911117.

Watson, Gary. 1975. 'Free Agency'. *Journal of Philosophy* 72 (April): 205–20. https://doi.org/10.2307/2024703.

———. 1996. 'Two Faces of Responsibility'. *Philosophical Topics* 24 (2): 227–48. https://doi.org/10.5840/philtopics199624222.

———. 2004. 'Volitional Necessities'. In *Agency and Answerability*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199272273.003.0005.

———. 2011. 'The Trouble with Psychopaths'. In *Reasons and Recognition: Essays on the Philosophy of T.M. Scanlon*, edited by R. Jay Wallace, Rahul Kumar, and Samuel Freeman, 307–23. https://doi.org/10.1093/acprof:oso/9780199753673.003.0013.

Way, Jonathan. 2016. 'Two Arguments for Evidentialism'. *The Philosophical Quarterly* 66 (265): 805–18. https://doi.org/10.1093/pq/pqw026.

Weatherson, Brian. 2008. 'Deontology and Descartes's Demon'. *Journal of Philosophy* 105 (9): 540–69. https://doi.org/10.2307/20620125

Wedgwood, Ralph. 2002. 'The Aim of Belief'. *Philosophical Perspectives* 16: 267–97. https://doi.org/10.1111/1468-0068.36.s16.10.

———. 2017. *The Value of Rationality*. Oxford: Oxford University Press.

White, Roger. 2005. 'Epistemic Permissiveness'. *Philosophical Perspectives* 19 (1): 445–59. https://doi.org/10.1111/j.1520-8583.2005.00069.x.

———. 2007. 'Epistemic Subjectivism'. *Episteme* 4 (1): 115–29. https://doi.org/10.3366/epi.2007.4.1.115.

Whiting, Daniel. 2010. 'Should I Believe the Truth?' *Dialectica* 64 (2): 213–24. https://doi.org/10.1111/j.1746-8361.2009.01204.x.

———. 2012. 'Does Belief Aim (Only) at the Truth?' *Pacific Philosophical Quarterly* 93 (2): 279–300. https://doi.org/10.1111/j.1468-0114.2012.01421.x.

———. 2014. ‘Reasons for Belief, Reasons for Action, the Aim of Belief, and the Aim of Action’. In *Epistemic Norms: New Essays on Action, Belief, and Assertion*, edited by Clayton Littlejohn and John Turri, 219-38. Oxford: Oxford University Press.

Williams, Bernard, ed. 1970. ‘Deciding to Believe’. In *Problems of the Self: Philosophical Papers 1956–1972*, 136–51. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511621253.011.

———. 1976. ‘Persons, Character, and Morality’. In *Moral Luck: Philosophical Papers 1973–1980*, edited by James Rachels. Cambridge: Cambridge University Press.

———. 1981. ‘Practical Necessity’. In *Moral Luck: Philosophical Papers 1973–1980*, 124–31. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139165860.011.

———. 1990. ‘Voluntary Acts and Responsible Agents’. *Oxford Journal of Legal Studies* 10 (1): 1–10. https://doi.org/10.1093/ojls/10.1.1.

———. 1993. ‘Moral Incapacity’. *Proceedings of the Aristotelian Society* 93: 59–70. https://doi.org/10.1017/CBO9780511621246.005.

———. 2008. *Ethics and the Limits of Philosophy*. Transferred to digital print. London: Routledge.

Williamson, Timothy. 2000. *Knowledge and Its Limits*. Oxford: Oxford University Press. https://doi.org/10.1093/019925656X.001.0001.

Zagzebski, Linda. 2004. ‘Epistemic Value and the Primacy of What We Care About’. *Philosophical Papers* 33 (3): 353–77. https://doi.org/10.1080/05568640409485147.

———. 1996. *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139174763.