Reasons and Blame

On the conflict between reasons internalism and blameworthiness

Louise Margaret Clover

Thesis presented for the degree of:
MASTER IN PHILOSOPHY

Supervised by Professor Caj Strandberg

Department of Philosophy, Classics, History of Art and Ideas
Faculty of Humanities

UNIVERSITY OF OSLO

Spring 2018
Reasons and Blame: On the conflict
between reasons internalism and
blameworthiness
Copyright Louise Clover

2018

Reasons and Blame: On the conflict between reasons internalism and blameworthiness

Louise Clover

http://www.duo.uio.no

Trykk: Weberg's Printshop, Oslo
Abstract

This thesis is concerned with the meta-ethical debate between reasons internalism and externalism, and more specifically with an apparent conflict between reasons internalism and the practice of moral blame. Broadly understood, reasons internalism means that a person has a normative reason to perform an action if and only if she wants to perform it, or would want to under certain conditions. Yet, when a person is blamed, it seems to involve a claim that she had most reason to do other than she did – irrespective of what she wanted. I argue that, as the problem is presented above, blame provides objections against reasons internalism. I then investigate the possibility that blameworthiness is not contingent upon reasons in the way the problem supposes. I cannot find grounds for an account of blameworthiness that both fulfils the conditions that are required under internalism to avoid the problem and yet still entails that all the characters who should be are suitably blameworthy. The blame problem persists.
Acknowledgements

First and foremost I would like to thank Caj Strandberg for his support and insightful comments. I further thank him for introducing me to the topic and for pointing me in the right direction when my writing went astray. I would also like to thank Maria Seim, Eirin Meinich, Aimée Povel, Mathias Helseth and Christel Fricke for commenting on various texts at various times as part of the ‘Blame’ discussion group. Finally I would like to thank my family and friends for their support throughout this endeavour, particularly Avery MacNeish, Lydia Fong, Melanie Engelfriet and Jonas Saabel – who have endured me whining about various philosophical difficulties for some time.
# Contents

1. Reasons: Internal and External................................................................. 1  
   1.1 The Broader Picture ............................................................................. 2  
   1.1.1 For Internalism .................................................................................. 2  
   1.1.2 The Central Problem ........................................................................ 3  
   1.1.3 The Blame Problem ......................................................................... 4  
1.2 Blame ....................................................................................................... 6  
1.3 Reasons Internalism ............................................................................... 7  

2 The Classical Argument and the Problem of Blame............................ 11  
   2.1 The Argument ....................................................................................... 11  
   2.2 Bernard Williams on Blame ................................................................. 19  
   2.3 Many culprits are blameworthy............................................................ 20  
   2.4 The Proleptic Mechanism .................................................................... 25  
   2.4.1 Problems around reasons .................................................................. 28  
   2.4.2 A Problem of Authority ................................................................... 31  
   2.4.3 A Logical Problem ............................................................................ 32  
   2.4.4 Summarising .................................................................................... 0  
2.5 A broader problem with the internalist argument .............................. 0  
2.6 Moving on............................................................................................... 2  

3 Blameworthiness and Reasons................................................................. 4  
   3.1 Moral Responsibility, Choice and Rationality ....................................... 7  
   3.1.1 Backward-looking vs. Forward-looking Accounts ............................. 8  
   3.1.2 Acting for reasons ............................................................................ 9  
   3.2 Attributability, answerability, accountability ....................................... 13  
   3.3 Attributability and Accountability as a Bipartite Theory ..................... 25  

4 Bad Characters ............................................................................................ 32  
   4.1 Are Psychopaths Blameworthy? ............................................................ 34  
   4.2 Should psychopaths be attributability blameworthy? .......................... 43  
   4.3 The Mob Gangster and The Egoist ....................................................... 44  
   4.4 Concluding Remarks .......................................................................... 50
1. Reasons: Internal and External

What gives us reasons to perform actions? Perhaps the most obvious response is to suggest that you have a reason to do something if and only if it fulfils some goal or desire; if it is in some way related to your pre-existing motivations. To see why is easy: a desire to learn French might be a reason to take a class, or to travel to France. However, the example need not always be so basic. For instance, one can have a long-term desire to own a house by the sea and, with such houses being expensive, surmise that this is a good reason to become a doctor (since they make a lot of money). This might then be a reason to work hard in school, or to pay for additional tuition in the sciences. The desire to achieve something in the long-term might be very far down the line indeed, but still provide reasons to perform actions in the present. What matters is that, at base, the reason for action is internal in that it is related to desires (or some other psychological feature of the agent) which are internal to the agent. A reasons internalist holds that all reasons for action are internal in this way while a reasons externalist denies this.

The division between reasons internalism and externalism is interesting and important because both views face substantial objections, and siding with either has important consequences for blameworthiness and moral responsibility (as I will later show).

The structure of this thesis is as follows:

Chapter 1: I set out the debate between reasons internalism, explain the problem from blame and how it is situated within the broader picture.

Chapter 2: I then give a thorough overview of the particular account of internalism I will measure against the problem: the internalist account put forward by Bernard Williams. I then put forward an argument as to why I think Williams’ attempts to mitigate the blame problem do not succeed.

---

1 I recognize that the use of ‘if’ and ‘only if’ has consequences with regard to whether the desire to do something is a merely a necessary condition for having a reason to act, or a necessary and sufficient condition. I assume throughout that the Internalist assumes that it is at least a necessary condition. They may also hold that it is a sufficient condition (as Bernard Williams, for instance, does), but I do not think that this particular distinction has much consequence for the problem I address in this paper.

2 Williams, ‘Internal and External Reasons’.
Chapter 3: This chapter examines the conditions that underlie the blame problem, namely conditions of choice and of rational criticism, and I attempt to see if there is an account of moral responsibility that can avoid the problems posed by these conditions. Specifically, I look at strong attributability – answerability – as an account of moral responsibility; and weak attributability as part of a bipartite account for moral responsibility as opportunities for the externalist to avoid the problem.

Chapter 4: I will look at the difficult characters who do not warrant blame proper according on the basis of the blame problem. I will attempt to see whether it is acceptable for these characters to be excluded from the sphere of moral accountability and whether it is sufficient that these characters are blamed only upon the conditions of attributability.

1.1 The Broader Picture

Before moving to the more specific problem of the conflict between blame and internalism, I want to briefly discuss the more general arguments for and against internalism and externalism. The reason for this is that I think it is helpful to understand why one might be motivated to be an internalist (or an externalist) in the first place. I also want to locate the blame problem within the realm of a much bigger, more general objection to internalism commonly called ‘The Central Problem’\(^3\). Understood in relation to the Central Problem, one can see how forms of internalism which attempt to avoid the CP might also avoid the problem from blame, and how – at the same time – the blame problem seems to be a case where our actual moral practise gives weight to the CP and points to it really being a substantial argument against internalism.

1.1.1 For Internalism

One reason to be an internalist about reasons comes from an argument that internalism has a kind of commonsensical plausibility. Some reasons seem to be clearly driven by desires, and if that is the case why should it not be true of all reasons?\(^4\) It seems fair to say that I have a

\(^3\) Finlay and Schroeder, ‘Reasons for Action’.
\(^4\) For an argument of this kind see Schroeder, *Slaves of the Passions*.
reason to study for my science exam if I want to be a veterinarian, or I have a reason to go to the shops if I will want dinner later. The internalist then just needs to expand this to explain how it is our desires might give us moral reasons, but the point is that the basic process – the move from desires to reasons - is already in place. The externalist, however, is left with a dilemma. They can accept that some reasons are internal, but then they must explain why it is that some reasons – namely moral reasons – are different. Or, the externalist can deny that any reasons are internal, but then they face the challenge of explaining why it is that in many cases reasons do seem to be driven by desires.5 Anyone who has difficulty with the latter two options might find themselves swayed towards internalism.

More problematic still for the externalist is in explaining how normative reasons might ever successfully motivate an agent if they are not internal to them. Desires seem often very clearly to motivate us, and to be used as the explanation for our actions. When asked why they stole a jacket, a thief is likely to exclaim that they wanted it. Their desire was for them a reason to act, and a sufficiently strong reason that they were actually motivated to act.6 But what kind of explanation for motivation can be given such that the explanation is not linked to an agent’s psychology? That question is difficult to answer. And if an answer cannot be found then the externalist has a problem. For in order to be a reason at all, it would seem that normative reasons have to be capable of motivating us at least some of the time7. There is a significant burden on externalists to explain what external reasons are metaphysically, and then to show how they are capable of motivating us. To be a reasons externalist seems to require more metaphysical legwork than is required of reasons internalism.

1.1.2 The Central Problem

As aforementioned, reasons internalism faces a substantial obstacle in ‘The Central Problem’. This might be explained simply as the recognition that there are cases where what we want to do clearly conflicts with what it is morally right to do. The moral rightness of an action appears to entail that there is a normative reason to perform the action. It appears that we

5 Parfit, On What Matters.
6 I use desires in a very flippant way here and acknowledge that whether desires are motivating is contentious. However, one can swap desires for some other psychological element and still maintain the internalist thesis. I do not wish to get into a debate about which kind of motivation is the right one, merely to point out that it is easier to explain how an internal, psychological element might be motivating as opposed to some external entity.
7 This argument is taken from Williams, ‘Internal Reasons and the Obscurity of Blame’.
always have these moral reasons irrespective of our desires or any other element of our psychology. Moreover, these moral reasons appear to take precedence over other sorts of reasons. For example, I might really want an A on my science essay and the best chance of my getting this is to get my friend to write it for me. One can assume that my means-end reasoning is not wrong; that I am not particularly good at this class and that my friend really is a prodigy who will certainly get me an ‘A’. Yet, we would still say that I have most reason not to cheat on the paper. I have most reason to do what is morally right, and there is something inherently unfair and immoral about getting a friend to write my essay for me. It seems plainly true that some actions, like not cheating, are morally right. And it also seems to be the case that the rightness of these actions entail reasons to act. Reasons internalism has a problem in cases where the rightness of an action does not coincide with what an agent wants.

Of course, it is possible to deny either of the underlying premises. One can just flatly deny that there is such a thing as morally right or wrong, though I take it this would be an unpopular view. One can also accept the moral wrongness of an action but deny that it follows that there is a reason not to perform the action. For instance, an agent may have been horribly treated his whole life, as a result there is nothing in his current motivations which would make him want to behave well. He beats his poor wife horribly. On this line of thinking he has no reason not to beat his wife, but it is still wrong that he does so. His reasons to perform or not perform the beating, do not affect the rightness or wrongness of the action itself or vice versa. There are several ways in which this view could be attacked, but one of the more compelling – and the subject of this thesis – is evidence from reactive attitudes, specifically from blame.

1.1.3 The Blame Problem

If an agent does something morally wrong then she is blameworthy. If she is blameworthy then it is justified to blame. This blame does not seem to be merely a condemnation of the act – when you blame someone you are not only saying ‘it is bad that you did that’. Rather there is also a claim that I had a choice about what I did and that I had a reason not to do it. This can be seen when you consider the circumstances that exempt or excuse an agent from blame. Those who are out of their mind, children and the severely mentally impaired are often
exempted from blame precisely because they lack the capacity to recognise and respond to the right sort of reasons. When we say “He didn’t know what he was doing!” in the context of such a person, what we often mean is “He didn’t understand the reasons he had not to do that!” Similarly, where an agent is rational but is placed in a position where they have a reason not to behave morally, they are excused from blame. If someone is accosted with “Give me the money or I’ll shoot” then we understand their reasons for giving away the money – even if it was not theirs to give. As such, they are exempted from blame for losing the money. Blame is only fair and appropriate where an agent had most reason (in this case a moral reason) to do something, and yet proceeded to do otherwise.\(^8\)

The problem of blame can be seen to cast serious doubt on the internalist thesis. This is because under reasons internalism it is possible to conceive of people who have nothing in their motivations to make them want to behave morally – Hitler and Caligula being classic examples. These are the sorts of people who truly want only to inflict suffering. Without any sort of desire or motivation to behave otherwise, on the internalist picture they have no reason to behave morally. On the conception of blame I have drawn above they cannot be blameworthy because, when confronted with a choice, they have no reason to behave in the expected way. Yet, these are exactly the sort of people we blame most – the people who really enjoy inflicting pain without a shred of remorse, people who truly have no scruples about lying or cheating. If the practise of blaming is right, then these people must have had reasons. And these reasons have to be external reasons.

In the pages that follow, I discuss whether blame so construed really is a problem for reasons internalism as I have presented it. I then go on to discuss whether blame is indeed contingent on reasons, or at least sufficiently contingent on reasons to maintain the objection to internalism. If blame is not contingent on reasons as presupposed then the tension with reasons internalism is dissolved. Before that, I look to define two of the key concepts involved in the discussion: blame and reasons internalism.

---

\(^8\) The claim about most reason is defended in Chapter 2. For the time being I take it to be true.
1.2 Blame

I throw the term ‘blame’ around abundantly throughout this paper, but what is actually at issue in the blame problem is an agent’s liability to be blamed– their ‘blameworthiness’. I think it is necessary to elucidate a little before continuing on what I mean by ‘blame’, ‘blameworthiness’ and their connection with reasons. This brief account will serve us for the first two chapters of this thesis.

Much of the debate around blame is about the sort of thing blaming is: whether it is an emotional state or a judgment, for instance. Further debate concerns how blame manifests itself – is it an attitude or something else? I am not concerned much with what actually constitutes blame in this paper. I am further unconcerned with the scope of who is able to blame a wrongdoer. Rather, I accept in a very general sense that blame is the negative response of the wronged towards the wrongdoer. I take the following as a paradigm case: An agent (A) does some wrong to another agent (B). B then blames A for the wrongful action. This might be seen as an overly simplified model, but I take it as the paradigm because I doubt many would disagree that this is an instance of justified blame, assuming that A meets the conditions for moral agency and there are no excusatory factors. I am as such not interested in cases where blaming is unjustified, that is either because the agent was not responsible for the action or because the action was not really wrong. I am also always concerned with moral blame, such that the wrongdoer has breached some moral standard.

Rather than being concerned with what constitutes blame or who has the right to blame, I am concerned with the conditions that make an agent liable to be blamed. What it is that renders an agent blameworthy. I suggest that blame is only justified when an agent had a reason not to do what they did. Evidence for this is in part taken from what we actually do when we blame. When B says to A ‘I blame you for X-ing’ it appears to also involve the claim ‘You had a reason not to X’. But it can also just be seen as a condition of fairness. It seems grossly unfair to blame someone for an action when they had no reason at all to behave otherwise. And this condition of fairness also appears in our actual practices of blame. This is because, as aforementioned, we exempt the people who are incapable of understanding their own reasons from blame (the mentally ill for instance) and because we excuse those who had

---

9 These distinctions are taken from Tognazzini and Coates, ‘Blame’.

more reason to do the non-moral thing, or who are unaware of their reasons due to epistemic limitations. If I open a door and unknowingly knock over a vase then it is true that I had a reason not to open the door – but I could not have known that I had a reason and as such am likely to be excused from blame. In the first half of this thesis I just take it as true that justified blame is contingent upon the wrongful agent’s blameworthiness which is contingent upon them having had reasons for not doing what they did.

In the second half of this thesis, I examine competing accounts of moral responsibility which require different degrees of responsiveness to reasons in order for an agent to be blameworthy. However, for the time being it should be taken that what I mean by ‘blame’ and its contingence upon reasons is as construed above.

1.3 Reasons Internalism

The way in which reasons internalism has been portrayed thus far is extremely crude, amounting to little more than the following:

a) A person has a reason for action if and only if she is motivated to perform the action

Though this is a form of reasons internalism, it is not a stance that many people would adopt without either modification or a qualification of terms. This is because, if what it means for an agent to be motivated to do something is that she was in fact actually motivated and did perform the action, then a) is patently untrue since it disallows the possibility of an agent having a reason to do something that she is not actually motivated to do. And it is uncontroversial that we are all actually motivated to do a great many things that it seems we have a reason not to do, often simply due to a lack of awareness of the relevant facts or some fault in our rational processing. I might be motivated to jump into shark-infested waters, but clearly I have good reason not to do that. Maybe I do not know the waters are shark-infested and as such do not know that I have a reason not to jump, or maybe I jump under the false belief that these are the sort of sharks that cannot eat me. But there are clearly possible circumstances in which I might actually jump, and as such it just cannot be that all the things that we are actually motivated to do correspond perfectly to the things we have normative reasons to do.
Due to this, some interesting forms of internalism do not commit the agent to actually being motivated, but rather require that the agent have some sort of motivational state. For instance, a very simple sort of state would be one in which the agent has the relevant desire. The desire entails that the agent could be moved by her own volition to perform the action, but it does not entail that she actually will perform it. Hence it is a weaker but more effective form of internalism than that given in a). Other interesting forms of internalism maintain the claim that reasons are borne out of what an agent can actually be motivated to do (rather than merely some motivational state) but then limit reasons to what an agent would be motivated to do under certain conditions. For instance, one can avoid the problem of the shark-infested waters by claiming that I only have reasons to do things that I would be motivated to do if I were I were free from false beliefs: the belief that the water is free from sharks or the belief that the sharks will not eat me, for instance. If I am free from false beliefs then I will be aware that there are sharks and that they will eat me – and as such I will no longer be motivated to jump in the water. A third sort of internalism combines the latter two sorts and claims that a person has a reason for action if and only if she would have the relevant sort of motivational state under certain conditions. This is a different claim still as it requires that an agent has a reason if she would have, for instance, a certain desire following, for instance, a period of rational deliberation.

These distinctions are relevant because they affect the susceptibility of internalism to the blame problem. For example, any sort of internalism that claims that an agent has a reason if and only if she actually has some motivational state will necessarily fall foul of the blame problem. This is due to the fact that it is always possible to conceive of a person who actually possesses no motivational state that could give them a reason to behave morally. To see this, let us take a classic example of this sort of internalism: the Humean Theory of Reasons (HTR). The HTR states that a person has a reason for action if they have a desire that would be served by that action. So an agent’s reasons are contingent upon her actual desires. But it is entirely possible that there is an agent with no desire whatsoever to behave morally, and thence no reason to behave morally. This behaviour means others blame them. And this blame carries the assumption that the bad agent always had a reason to behave morally. Hence the conflict between this sort of internalism and the practise of blame rears its head. So

---

10 The example of the HTR here is taken from Finlay and Schroeder, ‘Reasons for Action’.
any sort of internalism which subscribes to the HTR and, indeed, any form of internalism which makes an agent’s reasons contingent upon her actual motivational state is subject to this challenge.

On the other hand, not all counterfactual versions of internalism are necessarily susceptible to the problem. For instance, theoretically the problem of blame can be avoided by making the counterfactual such that everyone has the same reasons in the same situation (or at least in morally relevant situations). A potential problem with this is that the counterfactual becomes so weak that it is just trivially true and as such loses its appeal as an interesting form of internalism. For example, some people uphold the claim that if an agent has a reason to do something then necessarily she will be motivated to do it if she is fully rational. This is a sort of internalist thesis as it places a necessary link between an agent’s reasons and her motivations. However, if all it means to be rational is that you are motivated by your reasons then the claim is a tautology; it says: ‘if an agent has a reason to do something then she will be motivated to do it if she is a person who is motivated by her reasons’. On its own, it places no constraint on what can or cannot be a reason and hence, though it could avoid the blame problem, it would also be compatible with the claims of reasons externalism.

A further issue for the internalist is the plausibility of assuming convergence. Michael Smith, for instance, puts forward a counterfactual which assumes everyone would converge on the same desires if they were fully rational. But, one might query how likely this is. Were everyone in the same position would we really all want the same things? And a significant downside to such a view is that it risks losing one of the aspects that makes internalism appealing. A part of what the internalist picture does well is to adequately account for our differences - it explains why we have different reasons because we have different motivations or motivational states.

Perhaps as a result of these difficulties, many counterfactual versions of internalism do maintain a connection between an agent’s motivations and her psychology and as such do not try to avoid the problem from blame. For instance, one can accept that an agent has a reason

---

11 For instance Korsgaard, *The Sources of Normativity*.  
12 Smith, *The Moral Problem*.  
13 This is perhaps an unfair construal of Smith who lets in certain preferences in order to maintain the interesting variations in agents’ reasons under internalism. However, I take it that either way such a view is nevertheless possible and would avoid the blame problem.
for action if and only if she has a desire that would be fulfilled by that action in so far as she is rational, but deny that this would lead to any sort of convergence\textsuperscript{14}. Rationality in this case would not mean the capacity to be motivated by one’s reasons, but rather the capacity to rationally achieve what one desires, or some other non-trivial account. Moreover, though I have used rationality repeatedly as an example of the sort of condition put forward under counterfactual forms of internalism, what is important is that any sort of counterfactual which permits that agents could have different reasons under the same conditions (or at least when confronted with moral dilemmas) will fall foul of the criticism from blame.

There are many different accounts of reasons internalism – far too many to successfully address them all in this paper. However, as per the brief outline drawn, the problem from blame will affect any sort of internalism which makes a claim that an agent has a reason to do something if and only if she \textit{actually} has the requisite motivational state, and further that it will also affect many sorts of counterfactual versions of internalism. As such, though in the next section I focus particularly on the internalist account from Bernard Williams \textsuperscript{15}, the arguments I make from the blame problem often apply more broadly. I hope to show how the arguments employed by Williams in an attempt to mitigate the problem from blame are ultimately unsuccessful. In the subsequent section I then look more closely at some internalist theses which might theoretically avoid the problem, with the intention not of denying that they are unaffected but of highlighting the lengths to which the philosopher must go to avoid the issue.

---

\textsuperscript{14} see, for instance Joyce, \textit{The Myth of Morality}. Williams’ account is of course another example as we shall see.  
\textsuperscript{15} Williams, ‘Internal and External Reasons’.
2 The Classical Argument and the Problem of Blame

In this chapter I look at the account and arguments put forward for internalism from Bernard Williams. In some ways this is an odd choice, as Williams’ argument was introduced in 1981 and has several much discussed drawbacks. Why revisit what seems to be a very well trodden path? To answer this I would counter that, although Williams’ argument is discussed in great length generally, the literature around the blame problem specifically is relatively sparse. It is far more common for the argument to be criticised on theoretical grounds than on practical ones, yet it must be able to meet the demands of the latter too if it is to succeed. Morever, an advantage to looking at Williams’ argument is that he directly addresses the problem from blame in the essay ‘Internal Reasons and the Obscurity of Blame’. An important aspect of this is that he does not deny that blame poses a theoretical problem to internalism, but offers several ways in which it might be able to meet the challenge. As I have tried to show, the blame problem does not necessarily affect all forms of internalism, but it will affect many forms. As it definitely does affect Williams’ internalism this seems a reasonable starting point.

2.1 The Argument

Williams’ internalist account is introduced in the 1981 essay ‘Internal and External Reasons’. The division is further discussed in subsequent essays but Williams never substantially revises his initial offering. There is some controversy as to how the account should be interpreted, however in many cases these intricacies do not affect the effectiveness of the criticism from blame. This is because susceptibility to the criticism from blame rests on whether an agent’s reasons are the product of some subjective aspect of their psychology, and as far as I am aware there is no interpretation of Williams where this is not the case. It is,

---

16 The weaknesses most often brought up are that the account relies on the Humean Theory of Motivation which some philosophers flatly deny Darwall, Impartial Reason or that there is no possible interpretation of Williams which does not beg the question against internalism (see Russ. Shafer-Landau, Moral Realism., Finlay, ‘The Obscurity of Internal Reasons’.
17 Williams, ‘Internal Reasons and the Obscurity of Blame’.
18 See Finlay, ‘The Obscurity of Internal Reasons’.for an excellent explanation as to the various traditional interpretations
however, necessary to understand the elements of the account in order to grasp Williams’ subsequent claims about the blame problem. Here I put forward what I consider to be a fairly standard reading of Williams’ argument.

Williams starts from the point of the Humean Theory of Reasons. To reiterate, the HTR states the following:

A has a reason to Q if A has some desire the satisfaction of which would be served by his Q-ing.19

Understandably, Williams finds this reasons statement to be too simple and proposes four amendments to work it up into something more usable. As a starting point he suggests the following:

i) An internal reasons statement is falsified by the absence of some appropriate element from the agent’s motivational set.20

A reasons statement is any statement which has the form ‘A has a reason to Q’. An agent’s motivational set includes desires, but it further includes dispositions of evaluation, patterns of emotional reaction and personal loyalties. From hereon, I tend to use the term desires as Williams does, but it should be read as encompassing all the motivational states listed unless otherwise specified.

However, Williams argues, there are some elements in an agent’s motivational set which should be discounted: those based on false belief21.

ii) A member of an agent’s subjective motivational set will not constitute a reason for action if it is contingent on false belief, or if the agent’s belief that the action will bring about their desired goal is false.

20 Williams, 102.
21 Williams, 103.
This claim comes from a difference between normative reasons (reasons to do things) and explanatory reasons (reasons that causally explain why an agent acted). Williams uses the example of an agent who believes a glass of petrol is a glass of gin. Though the agent desires a gin and tonic, Williams queries whether this desire constitutes a reason for the agent to drink the petrol. On one hand, the answer is clearly no because the desire is one to drink gin, not petrol. But, on the other hand, were he to drink the petrol, in his explanation as to the reason why, the agent would clearly point to his belief that it was gin. This ‘explanatory dimension’ is not sufficient on its own to constitute a normative reason, says Williams. This is because were the agent fully rational and free of the relevant false beliefs then he would not drink the petrol.

Williams notes that it is entirely possible that:

iii) a. an agent may believe a false internal reasons statement about himself
b. an agent may not know a true reasons statement about himself

So, an agent behaving oddly, drinking petrol or acting against their own best interest may do so from false belief, either about the world or themselves, or from ignorance of their own motivational set. For example, I take it that an agent may falsely believe that they desire strawberries, only to taste one and discover that they dislike them. Conversely, an agent may not be aware that they would very much enjoy raspberries, and consequently avoid eating them. In both cases the agent has reasons, explanatory reasons, for having acted as they did – and in the first case their thinking was to some degree rational in that they were attempting to satisfy what they thought was a pre-existing desire. But it is clear that the agent has normative reasons to eat the raspberries and not the strawberries. Their explanatory and normative reasons are not equivalent in this case – but they would be free from the relevant false beliefs.

Williams’ final adjunct is that

iv) internal reasons statements can be discovered in deliberative reasoning

---

22 Williams, 103.
23 Williams, 104.
Williams writes as though this point follows naturally from the previous two, however what he really means here is the subject of some debate. The lack of clarity comes from two issues: the first is that there is some ambiguity regarding the definition Williams intends for ‘deliberative reasoning’; the second is that it is further unclear as to what he intends the outcome of this deliberative reasoning to be. In regards to the first problem, it can be seen that Williams at least intends for deliberative reasoning to encompass means-end rationality. On a very basic level, this requires that an agent can identify the right sorts of actions to fulfil their desires. However, I would put it that the sort of instrumental rationality he intends amounts to more than merely being able to identify the actions that will causally lead to the right sorts of outcomes. Williams writes:

A clear example of practical reasoning is that leading to the conclusion that one has reason to φ because φ-ing would be the most convenient, economical, pleasant etc. way of satisfying some element in S, and this of course is controlled by other elements in S. 24

So there is another dimension to reasons statements in that it seems they must often be capable of fulfilling several desires at once, and part of the purpose of deliberative reasoning is to identify how to best achieve your desires given the content of your emotional set as a whole. For example, if you are desperate to lose weight but of a lazy disposition and with a hatred of the outdoors, you probably have a reason to cut your calorie intake rather than to sign up for cross country running - this being the most practical way of achieving the desired outcome given your disposition. ‘Practical Reasoning’ or ‘deliberative reasoning’ in Williams’ sense is a kind of extended instrumental rationality, it is about the capacity to identify correct reasons statements and then rationally achieve them (given the content of an agent’s motivational set generally).

Regarding the second problem, it seems as though sometimes Williams expects the outcome of practical reasoning to be a belief. So following a period of deliberative reasoning an agent will come to believe that they have a reason to Q in virtue of the existence of some desire which will be satisfied by their Q-ing. This idea seems probable when one considers that the purpose of deliberative reasoning is to correct false beliefs about reasons statements. However, at other times it seems as though the result of deliberative reasoning is a change in

---

24 Williams, 104.
an agent’s S, e.g. a change in the desires that they have. It can then be argued that the outcome of practical reasoning is a change in the agent’s actual motivational state rather than a change in her belief. However, I think dividing the outcome of practical reasoning in this way is misleading. It seems to me rather that where an agent deliberates rationally, both her beliefs and her desires are subject to change. He writes for instance that an agent following rational deliberation ‘may come to have some more sense of what would be involved, and lose his desire for it’. In this case it seems that agent’s belief about his reasons and his desires both alter following rational deliberation.

The outcome of this is a reasons statement that is definitely true of the agent: She does in truth have a reason to Q because Q-ing would in truth serve her genuine desires. This is as opposed to the sort of reasons statement examined in the gin case, which might explain the agent’s action but which is a false statement because it is based on false belief. Practical reasoning entails that the agent has both the right sorts of desires and the right sorts of beliefs about those desires to entail reasons. Beliefs about what reasons one has are beliefs about one’s desires because ‘I have a reason to φ’ just means ‘I would have a desire to φ were I to go through a deliberative process’.

However, the claim that desires can change following rational deliberation needs some clarification. It is not the case that an agent can acquire a completely new desire – unrelated to anything presently in an agent’s motivational set – as a result of rational deliberation. This is because the agent must have something to deliberate from in order to be able to deliberate at all. And the only tools that an agent has to deliberate from are his existing motivational states. This point can clearly be seen in Williams’ discussion of Owen Wingrave. Owen Wingrave is a character in a Britten opera who is told by his father that he has a reason to join the army because his family have historically all been in the army. Owen has no motivation whatsoever to join the army. Can his belief that he does indeed have a reason to join the army motivate him to do it? Williams acknowledges that it can but, he argues, this point is only plausible if we agree that Owen already has a disposition to please his family (or some other relevant aspect in his motivational set). It is as a result not only of his changed belief about the reasons he has, but its combination with his current dispositions, that he might then find that he does in fact have a normative reason to join the army. However, this reason is an

25 Williams, 105.
26 Williams, 106–11.
internal reason, as it’s the result of practical deliberation from Owen’s pre-existing motivational set.

Williams uses this argument to question the plausibility of external reasons. He claims that in order for something to be a reason at all it must be capable of motivating an agent and must figure at least sometimes in an explanation of why they acted. To return to the gin/petrol problem, the explanatory and normative reasons in that case are not the same because the agent falsely believes the petrol is gin. If the petrol were indeed gin, however, then the agent’s explanatory and normative reasons would be the same. This seems to indicate some sort of conceptual connection between normative and explanatory reasons — normative reasons have an explanatory dimension in that it must at least be possible that people sometimes act for the reasons that are also their normative reasons. But, Williams argues, there is no possible way that an external reason could ever come to motivate an agent and as such, he queries, how could it ever have this explanatory dimension? The external reasons theorist has to claim that an agent will be able to be motivated by external reasons irrespective of their pre-existing motivations. But what could ever motivate an agent to accept a belief about external reasons if not already some pre-existing motivation?

I want to stress that in no way does this imply that Williams thinks that all external reasons statements will be internalized. That is, there is no way in which the belief that something is a reason will necessarily entail an agent is motivated by it or that they would be motivate by it following rational deliberation. An agent can only acquire new motivations and, consequently, new reasons following rational deliberation from their actual, existing motivational set — though the acquisition and application of new beliefs plays a role in rational deliberation and as such affects the reasons an agent has. This point is important because the way in which I have interpreted Williams is considerably wider than some others who take his use of ‘desires’ as meaning only desires in the narrowest sense or who deny that he thinks beliefs have any role to play at all. However, though my interpretation is wider, it is still the case that reasons are subjective to the agent because at root they are contingent upon the agent’s motivational state. At its most extreme, an agent who only has bad desires, dispositions of evaluation and preferences will only have reasons to do bad things. This can be seen clearly in Williams’ acceptance that ‘regrettable, imprudent or deviant elements in

27 Williams, 108.
28 Williams, 109.
S'\textsuperscript{29} can give rise to reasons. So it is certainly possible on this picture that there can be blameworthy characters who have no reason to do anything other than bad things\textsuperscript{30}.

A final aspect of the argument I want to highlight is that it can seem sometimes as though the process of rational deliberation is intended not only as a necessary condition for what an agent has reasons to do, but also as resulting in what an agent has most reason to do. I think this is an assumption that could be drawn from the extended sort of instrumental rationality that Williams advocates, in that rationality of this type is not only means-end rationality, it also requires an agent to weigh up their desires more generally to ascertain their reasons. To return to my previous example: the truly lazy agent has a reason to cut her calories rather than to take up running, say. If the deliberative process involves some aspects of rational judgement like this then one might assume that rational deliberation not only makes clear what an agent’s reasons are but also what they have most reason to do. This would then have some bearing on the blame problem, if one goes along with me in thinking that you are usually blamed for not acting in accordance with what you had most reason to do.

Williams’ own comments on this are somewhat vague. He writes that ‘an internal reasons statement does not apply only to that action which is the uniquely preferred result of the deliberation’. This suggests that rational deliberation will (or at least may) have a uniquely preferred result; that is there will be a reason which clearly carries the most weight as a result of rational deliberation. At the same time he makes a point of saying that an agent can also have many internal reasons. It is possible that rational deliberation could involve both judgements as to what an agents reasons are and judgements about what they have most reason to do, the difficulty is that Williams does not demarcate the rational process which leads only to a pro-tanto reason from the process which leads to a judgment of what you have most reason to do. It is, on the other hand, clear throughout that he intends that it is possible that an agent has multiple reasons for action at any one time. Therefore I suggest that the agent who rationally deliberated in the sense required only to know one’s pro-tanto reasons is not an agent who will also necessarily do what he has most reason to do. This would stretch the scope of rational deliberation too far. I draw attention to this distinction because

\textsuperscript{29} Williams, 102.

\textsuperscript{30} I use the most extreme example here purely because I think it makes the case most clearly. But there is no requirement that the agent is of the worst sort of character, only that the internal reasons the agent has do not correspond to the reason for which the agent is blamed.
Williams’ vagueness on this point is problematic for his own defence against the blame problem, as we will see in the next chapter.

To summarise, Williams puts forward an argument for internalism in which an agent has an internal reason for action if it would be true that the agent had some desire (in a broad sense) that would be served by his so acting following a period of rational deliberation. The process of deliberation can mean that either an agent’s desire or the means to achieve the desire are changed (or both). However, it cannot be that an agent is ever able to acquire a new desire without there already being something in his motivational set which could lead him to his new desire. This is because desires are necessary for motivation\(^{31}\). If desires are necessary for motivation then it is not the case that any action could ever be explained without reference to desires. Since reasons are the sorts of things that must be at least capable of providing this explanation, then it must be that reasons require desires. As such, internal reasons statements can be true while external reasons statements are necessarily false. The process of deliberation also allows that an agent can have many reasons at once, and how an agent is supposed to decide between those reasons is left unclear by Williams. From this explication of the Williams’ argument, I now move to address the problem at hand: the problem of blame.

\(^{31}\) This claim is commonly known as the Humean Theory of Motivation though, again, there is some evidence that it is not what Hume thought. I have tried to avoid too much reference to Hume in my explanation because I understand such terms to entail a much narrower reading of ‘desire’ than I think is intended by Williams.
2.2 Bernard Williams on Blame

A substantial portion of the literature on internal reasons and blame comes from an essay Williams himself wrote titled ‘Internal reasons and the obscurity of blame’. In it, Williams attempts to defend an internalist theory of reasons against the criticism from blame and, indeed, to argue that an internal theory of reasons in fact offers a better explanation of the practise of blame than can be offered by reasons externalism. His defence of internalism has three strands. One argument pertains to a distinction between moral appraisal, claiming that an action was bad, and blaming, claiming that the guilty party had a reason not to behave as they did. This issue is the subject of a Chapter 3. For now, I turn to the other arguments of Williams’ essay. The first of these claims simply that most of the people we think are blameworthy are in fact blameworthy, even on internalism, because they did have internal reason to behave in the appropriate manner, but failed to respond to these internal reasons in the right way. The second, related, argument puts it that even if the agent lacks the reason we think they ought to have had to behave appropriately, they had other reasons which via the ‘proleptic mechanism’ of blame renders them blameworthy. The former argument is far less controversial, and it is to this I turn first, examining a criticism put forward by David Sobel. After, I examine the more controversial, and far less convincing ‘proleptic mechanism’.

To briefly recapitulate, Williams’ view is that a person only has a reason for action if he could reach the conclusion that he has a reason by a process of deliberation from his actual motivational set. This allows for the possibility of persons who have nothing within their motivational set to make them to behave morally and who would not be motivated to act morally even if they went through a process of rational deliberation. If it is not possible that they could be motivated to behave morally then the agent has no reason to behave morally. And yet, as Williams notes in the essay, when we blame people we say ‘you ought not to have done that, you ought to have behaved differently’. This ‘ought’ appears to entail a claim that the agent had reasons to behave other than they did. If an agent has nothing within their motivational set to make them behave in the appropriate way then it would seem that the reasons we point to when blaming them must be external reasons or we blame them unjustly.

32 Williams, ‘Internal Reasons and the Obscurity of Blame’.
33 Sobel, ‘Subjectivism and Blame’.
34 Williams, ‘Internal Reasons and the Obscurity of Blame’, 41.
The task Williams sets himself in the essay is to mitigate the threat to internalism posed by this problem.

As is the subject of Chapter 3, one way around the problem posed by blame is to deny its connection with reasons. One can put it that blaming someone is not about, or at least not only about, the agent having good reason to behave in a certain way. It might instead, for instance, be the communication of a negative assessment of the agent’s character. However, it is worth noting that Williams accepts that blame is contingent upon reasons in the way the problem supposes, albeit his definition of blame is not strictly the same as I outlined in the first chapter. Williams is talking about ‘focused blame’—blame directed at specific agents for specific acts and omissions. His definition differs from mine in that I assume justified blame requires the breach of some moral obligation, while Williams regards blame only as having a moral dimension. For instance, he notes that one prisoner may blame another for screwing up a bank robbery. Clearly the bad robber has not breached a moral obligation to the other prisoner, but a much stronger moral obligation to the owner of the bank—yet he is blamed nonetheless. Williams claims there is a moral dimension to blame in that there ‘is some generally reprehensible characteristic involved in the explanation of the action’ like carelessness, laziness, selfishness etc. I would query whether this is a fair definition of warranted blame. It seems that the robber is blameworthy for trying to rob a bank and it is strange to suggest that he can be justifiably blamed for being a useless robber. However, Williams’ account of blame with its acceptance that an agent’s blameworthiness is contingent upon reasons is close enough to the definition of blame given in the introductory chapter for his arguments to be relevant. I take it that what I mean by blame is a subset of what Williams means by blame.

2.3 Many culprits are blameworthy

The first defence given against externalism is that much of the time an agent does have the appropriate item in their motivational set to bring about the right sort of reason, they just failed to bring it about in the right way. Recall that what makes a reasons statement true, Williams claims, is that an agent would properly have been able to understand themselves as having a reason following rational deliberation from their motivational set. There is no

---

35 Williams, 40.
36 Williams, 41.
37 Williams, 41.
expectation that in real life you will rationally deliberate - your reasons exist irrespective of your knowing them - but were you to rationally deliberate then you would know them. As such, an agent can have reasons that they do not know they have, or can go about trying to achieve desires that they know they want in the wrong way. The claim then is just that we can grant that in most cases if the agent had deliberated correctly then they would have found that they had a reason to act morally. Blaming them is entirely appropriate because they really did have a reason which they failed to act upon, either through ignorance or irrationality.

This explanation seems very plausible in some respects, in part because of the breadth Williams allows in the motivational set. Because an agent’s motivational set not only includes specific desires but includes broader dispositions, there are likely to be very few people who have no disposition at all to ‘do the right thing’, and therefore very few people after rational deliberation who would find that they had no desire, and hence no reason, whatsoever to do the right thing. This idea does not seem hugely interesting and Williams treats it with relative brevity. However, the point is more complicated than it first appears, in virtue of the claim about most reason. As I have already expressed, the sort of rational deliberation Williams’ endorses does not produce one overriding reason for action, rather it produces many pro tanto reasons. But what the problem from blame requires is that the agent always has most reason to do what is morally right. So, David Sobel makes the argument that even if we go along with Williams and agree that his argument shows that many people have a reason to behave morally, this is insufficient as a solution to the blame problem because Williams cannot account for why an agent has most reason to behave morally.38

This argument does not absolutely quash Williams’ case against the blame problem, but there must be some additional element added to his theory in order to answer the problem from blame. This additional element has to show why moral reasons have preference over others. Neither Williams nor Sobel go into the criteria for what gives some reasons preference over others at all, but I offer a suggestion for what might suffice here. The most obvious solution is that the weight of our reasons to correspond to the weight of our desires. I think this is the most plausible suggestion in virtue of Williams’ initial commitment to the Humean Theory of Reasons. Underlying everything is the assumption that what gives us reasons for action are our desires. Rational deliberation may affect these desires, but this deliberation is constrained

by an agent’s pre-existing motivational set. Facts about the world might change the strength of these desires but at root it seems that when we make judgments we are rational insofar as we do the thing that corresponds most accurately with what we want. Desires that have more weight entail reasons that have more weight. This is a fairly commonplace and internalist claim.

I would refrain from absolutely committing that this would be what Williams intended. He emphasizes about a vagueness about reasons, emphasizing that it ‘is often vague what one has a reason to do’.\(^{39}\) However, I take it that this vagueness does not preclude the plausibility of the idea that what one has most reason to do is what one has strongest desire to do if the choice is clearly between some strong desire and some weak desire. I labour the point because it is important to see how this then narrows the field of blameworthy agents and renders Williams’ original argument against the blame problem much less effective. Yes it may be true that many agents have a pro-tanto reason to behave morally on the internalist picture, but how much less likely is it that what the majority of agents most desire is to behave morally? I suggest that it is substantially less likely. Indeed, in many cases it seems as though morality requires a degree of self-sacrifice – you should do what is moral even though you really desire some other thing. So, if it is the case that what an agent has most reason to do is what he most desires to do then the problem from blame is more substantial than Williams initially tries to argue.

A different sort of solution to the problem is to bring morality into the process of rational deliberation. Blame is not merely contingent on what you have most normative reason to do, but the fact that this reason is a moral reason. If you could write the constraints of morality into a definition of rationality then following a period of rational deliberation an agent would always find that they had most reason to do the moral thing. Part of the problem with this is that Williams himself denies that the process of rational deliberation entails that ‘every rational deliberator is committed to the constraints of morality’\(^ {40}\). The reason he denies this is because it is not a claim you can just have for free. If being moral is a part of being rational

---

\(^{39}\) Williams, ‘Internal Reasons and the Obscurity of Blame’, 38.

\(^{40}\) Williams, 37.
then one has to explain why this is so, and this is not an enterprise that Williams wants to undertake. So, it is certainly the case that the Williams’ argument has a problem in that it allows for people who have nothing in their motivational set to make them behave morally, but it also seems that it has an additional, less obvious problem in that it cannot account for why people have anything more than pro-tanto reasons to behave morally. If to be blameworthy is to have gone against what you had most reason to do – not merely to have gone against what you had any pro tanto reason to do – then the internalist needs to explain how it is an agent has most reason to behave morally. If the claim is, as I suspect it would be, that the weight of reasons is contingent upon the weight of desires then this only leads to more problems, because it is clearly not the case that the morally right action always corresponds with the fulfilment of our greatest desires.

I imagine the sort of response the internalist would have against this criticism is that, if it is a problem for internalism, it is surely also a problem for the externalist too. Externalists might, for example, make a claim that moral reasons carry some sort of authority which means that an agent is always (excepting mitigating circumstances) obligated to do what is morally right, which means that they always have most reason to be moral. However, they then need to explain the source of this authority. A moral dimension might give us more reason to behave in a certain way, but this is not something one can just have for free either. And it may be that accounting for the weight of most reason on the externalist side requires as much legwork as explaining the hierarchy of reasons on the internalist side. This response is not without foundation, however an advantage for externalism is that it is not constrained by an agent’s psychology in the way this form, and many other forms, of internalism are. Any internalist theory which at root relies on something like the Humean Theory of Reasons has a problem in that they place the authority of an agent’s reasons within the agent’s psychology. In so doing they make it seem as though the weight of an agent’s reasons is also dictated by their psychology which narrows the field of blameworthy agents even more than one might initially have thought.

41 I take it that a Neo-Kantian view might do this. Certainly I take Korsgaard’s account to be one that would avoid the blame problem this way Korsgaard, The Sources of Normativity. This requires some metaphysical leaps which are much too intricate to be discussed here. In addition, though this view is internalist in the way it is formulated, it is not internalist in that an agent’s reasons are not restricted by their motivations.
Now, a further response from the internalist might be to just deny the supposition that an agent is morally blameworthy for going against what they had most reason to do. Perhaps it is true that an agent can be blameworthy for going against a pro-tanto moral reason while they had a greater internal reason to do something else. They are not irrational in the sense that they have acted as they had most reason to act, but they are still blameworthy for having acted against what they had a moral reason to do. My first response to this claim is that it seems to me that this goes against our intuitions about what we are doing when we blame. When the blamer claims that the blamee ought to have acted differently, this is a claim not merely that the blamee had a reason to act differently, but that they ought to have chosen that reason and acted in that way. If Jeanette tells Barbara that she is blaming her for her selfish actions then it seems clear that Jeanette is not claiming is that Barbara has a merely pro tanto reason not to behave selfishly. To the contrary, it seems that in blaming her, Jeanette is claiming that Barbara’s reasons not to behave selfishly have weight, greater weight than her reasons to behave selfishly. If Jeanette did not think that these reasons had much weight, then I find it dubious as to whether she would be bothered about Jeanette’s selfishness at all, let alone bothered enough to blame her.

My second response is that, if it is accepted that blame is contingent upon reasons at all then it has to be that blaming is only justified when the agent had a stronger reason to do the thing for which they are blamed than to do anything else. You can only face criticism on the basis of your rationality if you go against what you had most reason to do in any situation. Otherwise the criticism is unfair – assuming that the best that one can do in terms of acting rationally is to act in accordance with one’s reasons. If you claim that an agent can still be blameworthy and have acted rationally at the same time then the two have to be mutually exclusive. Saying this is equivalent to saying something like the agent acted in accordance with what they had most reason to do but their action was still bad and therefore they are morally blameworthy. This is a legitimate position but you cannot maintain this position and uphold the connection between blame and reasons.

In sum, while it may at first appear that the scope of the problem from blame can be mitigated somewhat by the argument that most agents have something in their motivational set which could give them reason to act morally, this claim is more troublesome that it first seems. For it is not enough to claim that an agent has a pro tanto moral reason to behave
morally, they have to have most reason to behave morally and the internalist argument\(^{42}\) cannot achieve this on its own. More problematic still, because of its contingence upon the HTR it seems probable that the internalist argument would equate the weight of reasons with the weight of the relevant desire, and this makes it seem even less plausible that what most agents have most reason to do will correspond to the moral thing to do. The internalist can try and argue that the externalist faces equally difficult metaphysical burdens, however because for the externalist reasons are not bound by an agent’s psychology, it is possible for them to give an account of what an agent has most reason to do which is less at odds with the blame problem. Alternatively, the internalist can try to deny the connection between blame and most reason but it seems to me that, not only does this go against our intuitions about blame, but seems to entail a denial of the connection between blame and reasons at all.

None of this is to deny that there will always be some agents on the internalist argument who have nothing within their motivational set to make them act morally. All Williams is trying to do thus far is narrow the number of possible candidates for whom this is true, but as I have tried to show, his claims in this regard are not as uncontroversial as they first appear. I now turn to his most famous argument in this vein ‘the proleptic mechanism’ argument, and attempt to show why it is wholly unsuccessful.

### 2.4 The Proleptic Mechanism

It can be unclear from Williams’ analysis what he means by the proleptic mechanism. This is because Williams’ analysis points to more than one way in which blaming acts proleptically. One claim is that blaming acts retroactively to enable an agent to acknowledge the reasons he actually has. According to Williams, what we blame an agent for is not so much a failure to recognise what he had most reason to do, but blaming represents a hope that the agent would come to a different conclusion ‘if he were to deliberate again and take into account all of the reasons that come before him’.\(^{43}\) This function is ‘proleptic’ in that subsequent to being blamed an agent will take into consideration the negative impact of being blamed as one of his reasons. So blaming can be seen as giving more weight to an agent’s existing reason to behave morally, or alternatively as providing an extra reason which has some weight in the

---

\(^{42}\) From hereon by ‘internalist argument’ I always mean Williams’ argument as I have construed it.

\(^{43}\) Williams, ‘Internal Reasons and the Obscurity of Blame’, 42.
agent’s overall rational judgment.\(^{44}\) Either way, one point that Williams appears to be trying to make here is that part of the function of blame is to try and get an agent to acknowledge that they really did already have a reason, an internal reason, not to do the action for which they are blamed. This possibility is not the proleptic mechanism, and I take it that this aspect of the argument was dealt with in the previous section.

The proleptic mechanism of blame applies to those agents who do not hold the reason for which it is claimed that they are blameworthy within their motivational set. For example, if you steal from me and I claim that there is a reason for you not to have stolen from me, I imply that you had a reason not to steal because it is bad to take things that do not belong to you; it is unfair given you have not paid for it; it breaches a basic right to property; it hurts me; and so on. Yet you had nothing directly within your motivational set that could have led you not to steal, and hence no reason not to steal. You desired very much the object you stole, and stealing was the only feasible way for you to obtain the object. In the strictest sense, the claim that I make when I blame you is a false one, because the reason that I claim you had is not a reason you could have discovered from a process of rational deliberation from your actual motivational set. Williams accepts that not all blamed parties will have within their motivational set the right motivation to do the thing for which they are blamed. To clarify, I presume that this example is supposed to function in cases where the agent is unaware that they will be blamed for their action. Otherwise it might seem that the agent already has a reason not to perform the action on the basis of the first claim I put forward in this section.

However, Williams then holds that in some of these cases the party is still blameworthy because there may be other dispositions within the agent’s motivational set to which the blame may, indirectly, relate. He points to the likelihood that people do not generally want to be blamed and furthermore he emphasizes that where this motivation to avoid blame is more than a simple aversion to negative consequences, it stems from an ‘ethically important disposition that consists in a desire to be respected by people whom, in turn, one respects’.\(^{45}\) Williams then writes that the claim that the agent had a reason to act other than they did becomes true. This is in virtue of the combination of ‘the disposition to do things that the

\(^{44}\) It is hard to know which because, as previously discussed, Williams is ambiguous on how our process of rational judgment should work.

\(^{45}\) Williams, ‘Internal Reasons and the Obscurity of Blame’, 41.
people he respects expect of him and...the recognition...of what those people expect’. I take it that what Williams is getting at is that it is true that as in the above example you did not, at the time of the action, have a reason not to steal. However, in virtue of the disposition to respect me and the new found recognition that I am disrespected by your actions, were you to deliberate rationally now then you would find that you have a reason not to steal. The reason being that I am disrespected by your action.

In sum, for Williams, blaming has two functions. In some cases it points to a specific internal reason that the insured has which ought to have prevented them committing the crime. In others, however, the blamer is not really pointing to a reason the insured actually had internally at the time they acted badly - but to one the insured ought to have held in virtue of the respect they have for the blamer. Via this ‘proleptic mechanism’ of blame it becomes a reason.

A query that might be still persist against the proleptic mechanism is the question of why the agent does not have a reason in the first place given that they have some disposition and given that it is a fact that the blamer will respond with blame. Williams does not explain this, however it is clear from the discussion that he thinks that the proleptic mechanism brings about a new reason and not merely that it highlights an old one. I would suggest that the conditions of full information only apply to facts about the world that are static. In the famous gin/petrol example referenced earlier it is truly a fact about the world that what is in the glass is petrol and not gin, as such the agent’s belief is false. But one cannot have a belief that is verifiably false about the future reaction of an agent until after the agent has reacted, as such I have to assume that when Williams caveats that an agent has no false beliefs that these do not include beliefs about the future. Hence why the wrongdoer does not have a reason to perform the moral action at the time of her wrongdoing.

Moving on, there are various criticisms which have been levelled against the proleptic mechanism argument. Here I give a breakdown of each of them to illustrate why the proleptic mechanism is really inadequate as an explanation of internal reasons and blame.

---

46 Williams, 42.
2.4.1 Problems around reasons

The first issue – which I term the most reason problem – is again from David Sobel and stems from the same argument put forward against internal reasons generally.\(^47\) He points out that it is easy to imagine the sort of person who does not always have most reason to avoid doing the sorts of things people will find blameworthy. Indeed, the kind of person who never did anything that someone might blame them for (irrespective of whether the blame is justified) seems quite pathetic. The problem, once again, is that the proleptic mechanism is not going to deliver (or at least is rarely going to deliver) what the agent has most reason to do. All it can deliver are pro tanto reasons for doing certain things and these remain insufficient to warrant blame.

However, I think the latter point, while not wrong, overlooks a more fundamental issue. The problem is not only that the proleptic mechanism cannot adequately account for what an agent has most reason to do, but that it produces the wrong reason. Blame is indicative of the accuser thinking you had a reason not to do what you did, and it is the failure to act on that reason that you ought to be blamed for. If you steal my car, then I will blame you on the basis that it is bad to steal cars (or some other reasons claim). This is the reason that I think you ought to have acted on. I am not blaming you for failing to recognise that you do not like being blamed and that stealing my car would lead to these negative consequences. The reason the blamer purports that you are blameworthy and the reasons you actually had prior to the action need to match.

There are several strands which make up the argument as to why the two reasons need to match. One argument is that there is a sense in which one should only adjust one’s behaviour if blame is warranted. When you consider the person who does not do anything for which they might be blamed, they appear so pathetic because they are not acting according to their own or even general principles, rather they are acting in response to the negative reaction of the blamer. But blame should only be warranted in the first place if the blamer has indeed breached some kind of moral principle. A friend at school might blame you for not giving her the answers to the English test – but that does not mean you should always give her the answers to tests for fear of being blamed. This is because her blame is unwarranted. It is

\(^{47}\) Sobel, ‘Subjectivism and Blame’, 166.
unwarranted because you have acted for moral reasons, or at least not for immoral reasons. So, you are neither bad nor irrational. Williams’ claim about the proleptic mechanism entails that in such situations your friend’s blaming you counts as a reason to act differently, but suggesting this only makes the internalist argument seem less convincing if what is supposed to count as a normative reason for you to act can be contingent upon the unkind behaviour of a friend.

The internalist might respond to this by claiming I have chosen a bad example. Where an agent is blamed for doing a good thing then they are clearly blamed unjustly, and as such it would be entirely appropriate if the agent upon deliberation rejected this as a reason to act. However, the problem is no less acute in cases where an agent has genuinely done something bad but no-one actually blames them for it. Consider, again, yourself as the car thief. It is of course possible that you steal my car and I never actually succeed in blaming you for it. Naturally I might shout and scream and blame ‘mystery-thief X’, but I never actually succeed in meeting you and giving you the kind of feedback that the proleptic mechanism requires - though you wholly deserve it. There is nothing directly within your motivational set which could have given you a reason to behave otherwise, although you do have the disposition of respect required by the proleptic mechanism argument. Is it really the case that the agent who has been actively blamed for stealing cars has a reason not to steal them while you, the agent who has never been caught, have no reason not to steal them? This would be a very strange conclusion. It should surely be that you had a reason irrespective of whether you are later actually blamed for it. The reason is what renders you blameworthy and what renders blame justified.

Here we come to the real crux of the problem with the mis-match of reasons which is that Williams has not really answered the problem from blame at all. As per my initial discussion of blame, an agent is blameworthy if they act in such a way that they had a moral reason not to act. However, on reasons internalism it is possible for an agent not to have had a reason to act but we still blame them for it. Agents’ must be blameworthy in order to be justifiably blamed, and as such they have to have had a reason: an external reason. However, if the proleptic mechanism argument succeeds at all, all it succeeds in showing is that, subsequent to the action actually happening, blame can give the blamed agent an internal reason not to do the same thing again. So it cannot render the original blame justified as the agent did not have this reason prior to the action. And in order to answer the blame problem it needs to show
how it is that blame is justified and how it is, therefore, that the insured is blameworthy. The only way that blame can be justified is if the reason the blamer cites for blaming is truly an internal reason of the blamed agent, or alternatively if blameworthiness is not contingent upon reasons at all.

It is unclear as to what criteria is supposed to make the blamed party blameworthy in the first place if it is not a failure to adhere to moral reasons. One option is to say that it is the badness of the action. But this explanation runs into problems given there are irrational agents (mentally ill people, children) who do bad things but are exempted from blame because they cannot reason sufficiently. This option also just begs the question against the externalist who will likely claim that the badness of the action entails a moral reason not to do it. A different response is that Williams does not intend that the proleptic mechanism argument will solve the problem from blame. Instead he just intends to show how blame can come to give the blamed reasons for the future. However, this would be a curious retort given that he starts with the recognition that ‘ought to have’ in the mode of blame corresponds to the agent having had a reason to do other than they did. He sets out the problem much as I did, and then proceeds to answer it by saying that the ‘proleptic mechanism’ of blame acts serves to give an agent reasons. In other words, he begins by agreeing that it appears justified blame is contingent upon reasons and then seeks to answer this problem by arguing that blame is forward-looking, that it gives agents reasons to not act in such a way again. So, it conflates two different issues: the question of whether reasons are a necessary condition of blameworthiness, and the question of whether part of the purpose of blame is to point out or to give an agent reasons for the future.

As a final note on this point, if one goes along with Williams and agrees that part of the function of blame is that it is reason-giving, what are the consequences for the internalist argument? I think one can go along with Williams and very easily agree that a part of the function of blame is to try and get people to adjust their behaviour. I stress, though, that this is only part of blame’s function – surely it has a backward-looking aspect too. However, I think many people would query whether the negative impact of blame should be so strong a motivation as to be reason-giving. This should become clear when we consider the cases of the unjustly blamed agent and the bad agent who is not blamed. One can come to see how an agent who is repeatedly blamed for telling the truth say, might come to think they have a reason not to tell the truth on the internalist picture. That is, it is plausible that repeatedly
being blamed for truth-telling and the negative consequences of truth-telling might lead the agent to no longer desire to tell the truth. Certainly, were the agent to then become secretive and deceptive, it would serve in an explanation of her action that she was always blamed for truth-telling. However, this only goes to show how the inappropriate application of blame could manipulate an agent into going against what we think she has a moral reason to do. We can see how she might come to believe that she has a reason, and we can see how this reason is internalized. But the question is does this entail that she actually has a normative reason? This is what Williams and the proleptic mechanism requires us to claim. Because he does not place the appropriateness of blame on reasons, and because an agent’s responsiveness to blame is grounded only in their disposition to be respected by others, it seems as such that the proleptic mechanism is as capable of giving an agent reasons to do non-moral actions as it is moral ones.

2.4.2 A Problem of Authority

A different but related problem is the problem of moral authority. It seems very bad to say that I should only do good things because if I do bad things the people who I respect will blame me for them. The sort of person who only behaves morally because they are afraid of punishment is unlikely to be considered a moral person (though inevitably they would do nothing wrong). To a Kantian idea, these people act in accordance with morality but not from it.\textsuperscript{48} To do what morality requires is to do something because it is moral, not because a person’s own interest happens to coincide with what is moral. For instance, consider the sort of person who chooses not to torture cats because it is difficult to catch them and they will bite him. It can be seen that it is good that he does not torture the cats, but the authority of this action is in the wrong place. He refrains from torturing cats not because it is immoral but because it is inconvenient. Now, to this the internalist might say that Williams sees the proleptic mechanism as relying on respect not on inconvenience. But does it make it any better to say that the agent has a reason not to torture cats only because the people he respects will blame him for it? The cat example is of course a very strong one. But any example can be used to make the point. If the authority of your moral reasons finds its root only in a disposition not to be disrespected by others then it can seem as though it finds its authority in the wrong place.

\textsuperscript{48} Kant, \textit{Critique of Practical Reason}. 
It is important to see that this is not only a claim about reasons but a claim about the character
of the moral agent. The person who does not torture cats only out of respect for certain judges
and a fear of their criticism seems at best very odd and at worst fundamentally bad. A person
who only saves a life out of respect for the people who will judge them appears selfish and
self-serving. And these traits may be in and of themselves blameworthy. The action is not
itself bad, but they could be blameworthy because they take their reason for acting from the
wrong sort of authority and are, as such, of bad character. Because the source of authority is
in the wrong place in the proleptic mechanism argument it seems as though it might in fact
make the agent more and not less liable to blame in the future.

Now the internalist may argue that these claims misconstrue the way Williams sees the
proleptic mechanism as working. The disposition of respect from others in conjunction with
the understanding that this is what they respect from you is supposed to cause your desires to
change. Such that you then actually have a desire not to torture cats, for instance, or at least
now lack the desire to behave in that way. The authority of the moral reason in the strictest
sense is not the external judge, it is just that their negative feedback in blaming influences
you and your desires, such that were you to deliberate again you would act differently. The
authority lies in your desires not the judge’s. However, I do not think that this distinction
makes a substantive difference to what I have argued. It does not seem to help to say that
your desires are the authority of your reasons but that the blaming agent can influence them.
The point is that what should influence your desire to behave morally is not only the negative
feedback from a judge. It does not assist to say that I have a reason to torture cats because I
like it. But you have blamed me for it and, now, following deliberation I have concluded that
I have a new desire because I understood that torturing cats was not what you expected from
me. It might be that authority is the wrong term in relation to this argument, but it still seems
as though a person whose desires and whose reasons are influenced only by blame from a
person they respect finds this influence in the wrong place.

2.4.3 A Logical Problem

As a final criticism, there is also something of a logical problem in the proleptic mechanism.
This argument is taken from John Skorupski who propounds that the attitude of respect which
is necessary for an agent to be receptive to proleptic blame is inconsistent. It is inconsistent because Williams argues that you will avoid blame out of respect towards the judge but one might query how far you can respect a judge who is blaming you for something which you in fact had no reason not to do.

As established, the proleptic mechanism is supposed to work retrospectively. It is not true that at the time of whatever blameable thing the agent has done that they actually had a reason not to do it. They had no desire in their motivational set not to do it, and they could not have reached one following rational deliberation. Rather Williams professes that in claiming the agent had a reason, the blamer hopes that the agent will see that they have a reason not to perform that way. But if one goes along with Williams and agrees that internalism is the correct theory of reasons, then we are left with three premises which lead to a somewhat unfortunate conclusion:

Premise 1: Williams’ Reasons Internalism is the correct explanation of our reasons for action.
Premise 2: In order to be blameworthy, an agent must have had most reason not to do the thing that he did at the time that he did it.
Premise 3: For blame to be justified, an agent must be blameworthy

The necessary conclusion of these premises is that the agent who is blamed but who did not have most reason to act morally is blamed unjustly. This is problematic for Williams because it is a condition of the effectiveness of the proleptic mechanism that the blamed agent has a disposition of respect for the blamer. But in what way could the blamed agent respect someone who blamed them unjustly? The answer is important because if, as Skorupski proposes, the agent cannot respect them then the proleptic mechanism is impossible. The condition of respect required for the proleptic mechanism to come about is not a condition that could exist in any case where blame would function only proleptically.

One potential route out of this problem is that it may be that what Williams requires is not that we respect the blamer as a good judge of when blame is appropriate, but just because we like them, or they are our boss, or a family member etc. However, to link this back to my

49 Skorupski, ‘Internal Reasons and the Scope of Blame’. This is a kindle edition so unfortunately I cannot give page numbers.
50 At before, I take it that the disposition of respect for others only becomes relevant after the fact.
previous objection, this then makes the authority of the reason even more arbitrary. Not only do I find that I had a reason not to do what I did because the person I respect is blaming me and I do not like blame, but my respect for that person may be borne out of my thinking that they are particularly funny, for instance. If a judge is of the sort we ought to respect in regard to blame then they should be a good judge of when blame is appropriate. If blame is only appropriate when the agent had most reason not to commit the action then no judge in the context of the proleptic mechanism argument is a good judge. In order for your blaming me to matter, I must respect you as the sort of person who is a good judge of when blame is appropriate, a good moral judge.

Alternatively, it could be that what Williams is claiming is that respect is a foundation of morality, and as such blame from a disrespected person is an expression of something being morally bad. You are a person who has been disrespected and you blaming me just says something about my action being disrespectful. In some cases this is true, if I always talk over you then this is a clear case of disrespect. Your telling me how this disrespects you might lead me to change my ways, I might now see that I have a reason not to do it again. I had the disposition to respect you but I did not understand what respect involved, and your blame informed me of it.

However, this is a troublesome explanation. One way in which it is troublesome is that it clearly does not apply in all cases. Talking over you is bad because it disrespects you, but torturing cats is clearly not bad because it disrespects you. I do not even think that you can claim it is bad because it disrespects some societal norm – torturing cats for pleasure is more fundamentally bad than that. So it seems that this explanation can only function in cases where the moral wrong is clearly one of disrespect towards the blaming party. But, more problematically, the example given does not seem to be an example of proleptic blame but of a case where blame is in any case already appropriate. Because it seems to be the case that were I rationally full-informed as to what the concept of respect actually entailed as well as being in possession this disposition of respect towards you, then following rational deliberation I would not have desired to talk over you in the first place. The disposition of respect is too strong in this case for blame to function proleptically. As such, I see it that what Williams takes the disposition of respect to entail is that the blamed agent respects the authority of the blamer. But how one can be expected to respect someone who blames unjustly remains mysterious.
2.4.4 Summarising

In sum, I find the proleptic mechanism to be wholly problematic as a response to the problem from blame. I think the most damning objection is that Williams does not even really answer the problem. He offers proleptic blame as a response which involves the agent having a reason, but it is neither the same reason that they are blamed nor is it a reason they had at the time of the offence. The agent is, as such, not blameworthy and the blame is not warranted – so the conflict between blame and reasons as first set out remains. But even if one ignores the problem of blame as first set out and just assumes that blame is forward-looking, Williams cannot offer a coherent account of why the blamed party should respect the moral judge or why their judgements should have authority. And, furthermore, even granting that they did have authority it is hard to see how this authority would not be arbitrary or how it could be the right sort of authority for moral reasons. So, all in all the problem from blame stands and is unmitigated by the proleptic mechanism. I move later to offer a more plausible way to avoid the problem, by divorcing blameworthiness from reasons. However, prior to that, I want to point to a way in which the problem is compounded.

2.5 A broader problem with the internalist argument

When one looks at the argument for the proleptic mechanism, and the move to defend internalism from the problem from blame more generally, it highlights an oddity in the internalist account. If it is true that blame is contingent upon reasons, and if reasons internalism is also true then when blaming someone we are making a judgment about their psychology. This point is usually just summed up by saying that blame entails the existence of external reasons. But here I want to consider what it means if it just is the case that both reasons internalism and our concept of blameworthiness are correct. Blaming, then, becomes quite odd, as Sobel writes:

..it is not enough for the internalist to be able to point to point to something blameworthy by their lights in cases where blame seems appropriate. They must persuade us
that the good thought behind the urge to blame in such cases derives from the presence of this internal reason.\footnote{Sobel, ‘Subjectivism and Blame’, 165 my emphasis.}

When blaming someone the internalist makes a claim that the blamed party had more reason to behave in another way, and this reason must be an internal reason. For it to be an internal reason, there has to be a corresponding desire. So blaming someone on the internalist framework equates to claiming that the agent would have had a desire to behave morally – or at least not immorally – had they rationally deliberated.

This picture strikes me as strange in several ways. Firstly, it just is not as plausible as a practical explanation of what happens when we blame as the externalist explanation. When blaming someone, the externalist can point to a reason in the world – for example that there was value in performing the right action and disvalue in performing the wrong one. Granted they have then got to explain how it is that value engenders reasons and how those reasons can motivate but, the point is, these reasons exist irrespective of the agent. The only judgment the externalist need make about the blamed party is that she is a person capable of recognising external reasons as reasons, i.e. that they are not a child or mentally incapacitated in some way. The internalist, on the other hand, is making a much stronger claim about the blamed person. They have to claim that the agent already has something in their motivational set which gives them a reason to behave in a certain way. Not only that but they have to claim that the blamed person already had something within their motivational set which gave them most reason not to do what they did. So they are required to make a claim \textit{about the wrongdoer}, as to the content of the wrongdoer’s motivational set and as to his desires following rational deliberation. I am sceptical as to whether we make the latter sort of judgment when we blame.

Secondly, there is an epistemological problem. We cannot possibly know the content of people’s motivational sets. So when we say ‘I blame you for stealing’, at best what we are saying is ‘I suspect that you would not have had a desire to steal had you rationally deliberated from your motivational set’. This may be a fair claim if you know the person you are blaming well, but it becomes much more of a stretch if you are blaming a stranger. Consider what would happen if you went around making claims about the motivational sets

\begin{footnotesize}
\begin{itemize}
\item Sobel, ‘Subjectivism and Blame’, 165 my emphasis.
\end{itemize}
\end{footnotesize}
of people you did not know very well. Would it not be odd if I just told you, a total stranger, that you liked blueberries. This is not the same as saying that you would like blueberries if you tried them, and therefore that you have a reason to try them. Blaming someone is to tell them that they had a reason, and therefore that they had a desire, not that they would have a desire iff. (excluding of course the deliberative constraint).

Now it might be that we can make some plausible assumptions as to what would be within someone’s motivational set, but to this I would again reiterate my claim that this sort of deep psychological appraisal is just not what we think we are doing when we blame someone. Williams, I think, recognises this, which is why he suggests that where blame is used proleptically one does not profess to make a claim about the agent’s psychology, but is rather hopeful that the agent will change their ways. However, as aforementioned, this claim does nothing to mitigate the original problem from blame, and neither does it dispute that where blame is not proleptic it is making this sort of claim. But when I blame you, I do not pretend to know your psychology. I do not profess to know your desires. Rather, I do not even care about you desires. I claim that you had a reason not to do that bad thing completely irrespective of what you wanted.

2.6 Moving on…

To summarise I do not think that there is a way in which reasons internalism can be reconciled with a conception of blameworthiness as being contingent upon an agent having had most reason to do the right thing. It may be that Williams is right and that many blamed parties are truly blameworthy. However, more work needs to be done to show how it is an agent comes up with what they have most reason to do for this assumption to be fairly assessed. I have not provided an explanation for how this could be done, however I suspect that it would require some addition to Williams’ account of the rational deliberation procedure. Even so, there will still be many characters who do not have most reason to behave morally in a given situation. As the proleptic mechanism argument fails, I find that none of these people will be blameworthy according to an account of blameworthiness which is contingent upon reasons. But how can it be that so many people are blamed unjustly? Is it

52 Williams, ‘Internal Reasons and the Obscurity of Blame’, 42.
sufficient that we hope that these wrongly-blamed people will change their ways for us to blame them unjustly?

This point, along with my more general concern regarding the fittingness of reasons internalism with the practise of blame leads me to conclude that, in order for reasons internalism to succeed, there must be something other than a failure to respond to reason that can render an agent blameworthy. There must be some account of blameworthiness which better fits with the internalist project. In the next chapter, I seek to find this more fitting account of blameworthiness. If it cannot be found, then it seems that there is no way back for the internalist.
3 Blameworthiness and Reasons

Thus far I have assumed a very basic conception of blameworthiness. Indeed, it is not really a conception of blameworthiness at all, but rather a claim that a necessary condition of blameworthiness is that an agent acts against their best reason. I have claimed that in order to be blamed for something, an agent must have had a reason not to do the thing that they did, and that this reason must have been what they had most reason to do. That is, they had more reason to save the drowning child than they did to go and get ice cream, and therefore can be blamed if they choose to go and get ice cream. If they had more reason to go and get ice cream (say if some mad super-villain will wipe out all the children unless he has his ice cream)\textsuperscript{53}, then the agent is excused from blame.

The conflict from blame comes from this idea that when an agent is blamed she had most reason to do something else. If internalism is true then we blame lots of people who seem to have no reason to behave morally, or more reason to do something else still. Let us now assume that it is the practise of blame that is mistaken, rather than reasons internalism, then I take it that the assumption is that it we ought not to blame these people. It would be unfair to blame them because they did not have the relevant choice such that they could have chosen to behave morally, or it is inappropriate because they are unmotivated by moral reasons so criticizing them on the basis of their reasons would be pointless. This appears to be what the internalist would endorse. Williams makes the former argument in ‘Internal Reasons and the Obscurity of Blame,’ putting it that the externalist risks browbeating the wrongdoer by claiming they could have performed an action that they could not have in fact performed.\textsuperscript{54} Meanwhile, the latter argument has been defended more recently by Stephen Darwall in ‘Morality, Blame and Internal reasons.’\textsuperscript{55} However, the purpose of this chapter is not, in the first place, to uphold these arguments. Biting the bullet might be what Williams and Darwall advocate for reasons internalism but given then that some of the worst characters are those who would not be blameworthy – the Hitlers and Caligulas of the world – I take it that this is not a particularly satisfactory option. Instead, I want to see if an agent can be blameworthy without opening internalism up to the blame problem at all. Or, if the latter cannot be achieved, to at least find an account that diminishes the problem.

\textsuperscript{53} Or more reason still to do something else
\textsuperscript{54} Williams, ‘Internal Reasons and the Obscurity of Blame’, 45.
\textsuperscript{55} Darwall, Morality, Blame, and Internal Reasons.
As I see it, the condition that an agent must have been going against what they had most reason to do in order to be blameworthy rests on two further assumptions about blame and when it is warranted. The first assumption, as I mention above, is that blame requires that an agent had a choice as to how to act and chose to act badly. It seems the agent must have had fair opportunity not to do what they did to avoid the negative consequences associated with blame. I take it that this claim is intuitively very plausible. If it was not possible for an agent to perform an action then it seems distinctly unfair to blame her if she fails to perform it. To see how this claim comes about consider what options are available to the bad internalist actor. If he has nothing in his motivational set that could make him behave morally then he will likely still have a choice. He will have a choice between various prudential reasons for instance: between doing the Hoovering because the house is dirty or between watching the telly because there is a big game on. He might even end up doing the right thing – for instance he might keep a promise to a friend but do so only because he thinks his friend will take him on holiday if he does so. What is unavailable to the agent is the opportunity to perform for the right reason. If blame represents a criticism that an agent failed to perform for the right reason then it is unfair to blame an agent who could never have performed for this reason. Hence it can be claimed that in order to be blameworthy and agent must have had a choice such that they could have performed for the right reason.

The second assumption is that blame represents some sort of failure in an agent’s rational judgment such that they did really have most reason to behave morally but judged poorly that they had more reason to do something else. This second assumption is more important than it might seem because on internalism it can be true that an agent has no reason to behave morally or that they might have more reason to behave immorally. If the agent then judges that they have no reason to behave morally or that they have more reason to behave immorally then there is nothing wrong with their rational judgment. Their judgment about the reasons they have and the weight of the reasons they have is correct. So they cannot be criticised for having the wrong reasons on the grounds that it is some rational fault. I highlight this point because it is no good if an account denies that an agent need have had a choice but then relies on the idea that the agent is prima facie rationally faulty for having chosen the wrong reasons. This would just presuppose that an agent’s normative reasons are

56 see Wallace, *Dispassionate Opprobrium*.
57 Here we find ourselves caught up in Kant’s famous ought implies can principle
58 They could however still be criticized on the grounds that their reasons or judgments are morally bad, as we shall see.
external reasons. So for internalism to succeed blame must not be a form of rational criticism such that an agent is rationally faulty for having the wrong reasons.\textsuperscript{59}

From this, I take it that in order to avoid the conflict with internalism an account of blameworthiness must meet two conditions:

The Non-Choice Condition – is that the account must not rely on a condition of choice such that it must have been possible for the agent to act in the morally right way in order for them to be deserving of blame.

The Rationality Condition – is that if an account fulfils The Non-Choice Condition, then it must further allow that the agent can be entirely without rational fault and still be blameworthy.

It should be noted that both conditions are highly problematic. The Non-Choice Condition would deny the assumption that an agent ought to have been able to choose to do the right thing which just appears on intuition to be a necessary condition of blameworthiness. If we assume that blame is something that ought to be deserved, then it is difficult to justify blaming someone who was unable to act in the way we expected them to. This condition is also in a sense prior to the other condition because for an agent to be blameworthy on the basis of most reason, it must first be the case that they have the reason at all.

The Rationality Condition, meanwhile, is difficult because, as I have been arguing all along, blaming seems to assume that the agent did have most reason to behave morally but failed to act on this reason. They are rationally faulty then either for failing to recognise that they had a reason at all, or for acknowledging that they had a moral reason but then failing in their judgment that this was what they had most reason to do. Blaming as such seems like a form of rational criticism. What makes this especially problematic is that where The Non-Choice Condition is fulfilled one tactic is to try and avoid the choice problem is to fall back on the condition that an agent can be rationally faulty irrespective of their choices. But this account would then fail The Rationality Condition. Much more will be brought out about these claims in the course of this chapter.

\textsuperscript{59} I talk of rational faultiness rather than irrationality to differentiate this from cases where an agent has judged correctly that they have a reason to perform an action but then fails to perform it.
The purpose of this chapter, then, is to try and find an account that can meet both The Choice Condition and The Rationality Condition. The plan is as follows: in the next section, 3.1, I clarify the relationship between reasons and blameworthiness. The purpose of this is to establish what we are seeking an account of - moral responsibility – and to begin to establish grounds for an account which can fulfil the two conditions I have put forward here. I will try to show that one can still agree to the plausible claim that responsiveness to reasons is a necessary condition of moral agency without falling foul of the choice problem or committing to a rational fault view. In section 3.2, I briefly outline three different proposals for what constitutes moral responsibility: attributability, answerability and accountability to see which of these is capable of helping the internalist and in what ways. I find that there are two plausible alternatives for the internalist: 1) to adopt a unified account of moral responsibility as answerability – this tries to avoid the problem from blame altogether; 2) is to adopt a bipartite theory of responsibility which splits moral responsibility into accountability and attributability. On the second option the problem from blame persists under the full conditions of accountability but not attributability, so if it can be agreed that only ‘attributability blame’ is appropriate in the hard cases for internalism then the problem from blame can be avoided. The subsequent sections, 3.3 and 3.4, are dedicated to exploring these two options.

### 3.1 Moral Responsibility, Choice and Rationality

I have claimed that I am going to look at accounts of blameworthiness in order to try and avoid the conflict with reasons internalism, however it would be more accurate to say that the focus of the discussion in this chapter and the next is on accounts of moral responsibility. I take it that if an agent is morally responsible for an action and they do something bad without appropriate justification or excuse then they are blameworthy. I also take it that if an agent is morally responsible for an action and they do something good then they are praiseworthy. So an agent must meet the conditions of moral responsibility in order to be blameworthy.

---

60 I later put it that that the appropriate form of moral appraisal in an attributability sense is aretaic appraisal rather than blame as per Watson, ‘Responsibility and the Limits of Evil’.. However for the purposes of laying out the argument I trust that ‘attributability blame’ suffices.

61 I take this to be a fairly uncontroversial claim, however some philosophers only take the negative responses (as in blame) to be constitutive of holding morally responsible. Someone can then only be morally responsible when they act badly see Wallace, Responsibility and the Moral Sentiments.
Here I want to explain briefly defend the two claims about choice that I made in the introduction from a couple of criticisms.

### 3.1.1 Backward-looking vs. Forward-looking Accounts

One argument for thinking that choice is necessary for moral responsibility comes from the idea that one must have some control over what one does in order to be responsible for the outcome. Of course, one can be responsible in some way for something just in case they are the causal originator of that thing – for instance in the way that an earthquake is responsible for the destruction it causes. However, moral responsibility must require more than just being the causal originator of an action because it leaves the agent open to blame – and we could not truly blame an earthquake. As I have noted, a common intuition about blame is that it ought to be deserved and this appears to entail that the agent ought to have had a fair opportunity to avoid performing the bad action. As such it requires that the agent ought to have had a choice such that they could have acted morally.

The claim I have made about choice is *backward-looking*, in that it suggests that blame ought to be deserved, and it makes a claim about the conditions the agent was subject to at the time of the wrongdoing. One way of denying this claim is to put forward an account that is only forward-looking. These accounts, like that of J. J. C. Smart (1961), are consequentialist accounts that regard blame and its legal paradigm of criminal punishment only in terms of controlling the wrongdoer and preventing future recurrences of their wrongful action. And it should be noted that these accounts therefore necessarily avoid the problem from blame. This is because all that would be required of the agent to be appropriately blameworthy in such cases is that they have done something morally bad and that they are capable of being controlled – that is there is at least a hope that they would respond to blame by not committing such a crime again, or that they would apologise, or some other positive consequence.

---

62 see Levy, ‘The Good, the Bad, and the Blameworthy’. for a defence of this point
63 Claims about choice are inevitably tied up in the debate regarding the truth or falsity of determinism. I just assume here that agents are capable of the relevant sort of control. See Fischer and Ravizza, *Responsibility and Control*. for more insightful discussion
This thesis is not overly concerned with what blaming is, however theories of moral responsibility that rely on conceptions of blame as only forward-looking have some fairly obvious weaknesses. For one thing, they ignore the backward-looking aspect where the blamer claims “You should not have done that!”. Naturally, in part this implies “You should not do that again..” but that is surely not the only thing that it implies. There is a recognition that what was done is bad, and that its badness is deserving of this recognition, if not further punitive responses like resentment and punishment. Equally, it can be argued that there must be more to blaming and punishment than only the control of behaviour if, for example, we accept that we try to control the behaviour of animals, or children, or the mentally ill but we do not blame them. You can tell a dog off for jumping up in order to try and make it desist and teach it not to do so in future, but that is not equivalent to blaming the dog for jumping up. This entails that there must be more to blame than the aim to control the blamed, and more to blameworthiness than the capacity to be controlled.

More plausible accounts of blameworthiness are able to encompass both the forward and backward-looking aspects of blame. These accounts acknowledge that part of the function of blame is as a mechanism to try and control behaviour and to prevent a recurrence of the blameworthy action. However, they also assume that part of the function of blame is to bring an agent to account for her wrongful action – to recognise, as it were, herself as the author of the action and to acknowledge its badness. As such, the backward-looking aspect requires that an agent is morally responsible for the action in a way that is not presumed merely by the forward-looking aspect. All that is required for the forward-looking aspect to function is that someone has done something that means her behaviour requires amendment. But to be truly blameworthy in a backward-looking sense requires more. Given I take it to be backward-looking, and therefore a question of desert, this can be taken to give more weight to the claim that an agent must have had a choice.

3.1.2 Acting for reasons

The second claim I made was that blame is a form of rational criticism. One strong argument in favour of this claim is that it seems to include and exclude the right sort of people within the sphere of moral agency. Agents who could meet the conditions of being morally responsible are moral agents. To clarify, they are not prima facie morally responsible for an
action, but if the action is immoral and they are responsible for it then these are people who are liable to be blamed. They may, of course, avoid blame if they have some justification or excuse. And if they are truly blameworthy then it does not necessarily entail that blame has to be expressed. But they are separated from non-moral agents who may fall out of the sphere of moral agency temporarily – because they go temporarily mad for example – or because they are outside the sphere of moral agency permanently – because they are non-human animals, mentally-disabled or children.

If blaming is a form of rational criticism then it follows that moral agents would be the sort of people who act for reasons. This capacity is not usually limited to merely acting on reasons, but the ability to act of the basis of judgments about reasons. One explanation for this is that limiting the scope of moral agency only to creatures that act for reasons at all makes it over-inclusive. My cat chases string seemingly to fulfil some sort of desire. In a weak sense this could be a reason for my cat, but we would not usually think it is sufficient for moral responsibility. The sort of rational requirement needed for moral responsibility seems to be not only that an agent has reasons but that they are capable of stepping back and making judgments as to which reasons to act upon. They must be capable of deliberating about their reasons and of making judgments as to the weight of their reasons. If an agent lacks these capacities then they cannot be a moral agent.

And the ability to act on the basis of judgments about reasons does indeed include and exclude the right sorts of people within the sphere of moral responsibility. As Angela Smith argues, that creatures that do not have the relevant rational faculty, like non-human animals, are not blameworthy. If my cat knocks over a very expensive vase I might be angry, but I would not blame the cat. He is not capable of acting on rational judgments and as such is outside the sphere of moral responsibility. In a different sort of group are adults excused from blame because they were in a period of madness. These adults usually have a rational capacity, that is they are usually capable of recognising and acting on their reasons, but their illness has rendered them incapable either of seeing their reasons at all, or of making rational judgments. Their illness might lead them to believe they have reasons that they do not have,

---

64 Although in the case of children it seems likely that their inclusion will increase over time as they gradually meet more of the conditions
65 The capacity for deliberation between reasons is particularly pressed in Korsgaard (1996), but rational deliberation is a feature of many accounts including, of course, Williams (1989)
66 Smith, ‘Control, Responsibility, and Moral Assessment’, 388.
or to place undue importance on the wrong reasons – they count ceiling tiles rather than feed
the baby. Children, meanwhile, have different degrees of reasons-responsiveness at different
stages in their development. This could account for why we sometimes think it is appropriate
to blame children in some sense but at other times do not. That is, it is perhaps appropriate to
blame a teenager for their perpetual rudeness but it seems inappropriate to blame a toddler for
destroying your lipstick.

This argument might be seen to support the claim that blame is a form of rational criticism
and that, therefore, in order to be blameworthy an agent must have acted against what they
had most reason to do. However, one can agree that reasons are a necessary component of
moral agency without agreeing either that blame requires that the agent had a choice, or that
the agent has gone against what they had most reason to do. One can instead claim that an
agent who acts for reasons at all is an agent who acts intentionally. Their actions can be seen
as an expression of their character. For instance, an agent who acts for morally good reasons
appears to be an agent of morally good character. Their acting upon good reasons expresses
their good will. An appropriate response to this is praise. On this line of thinking, the capacity
to act on the basis of judgments about reasons is necessary only because reasons are required
for the agent to truly be in control of their actions. I call this condition of moral agency ‘weak
reasons-responsiveness’.

This point is perhaps better explained by considering an agent who is not blameworthy
because their action says nothing about them. If a man robs a bank because he was being
hypnotized then he is not acting for reasons. His action has no intent, and it says nothing
about him that he robs the bank. One cannot evaluate his character on the basis of this action
because it was not done for reasons, and as such he is not responsible for what he does. In a
more everyday example, we do not usually criticise those who act compulsively. If Jane has
obsessive-compulsive disorder and must repetitively lock the door fourteen times before she
is able to leave the house then it is inappropriate to criticize her for it because her behaviour
cannot be helped. Jane probably understands that she has good reasons not to repeatedly keep
locking and unlocking the door, but this judgment has no effect on her action. So one way of
thinking about reasons in relation to moral agency is that in acting for reasons at all, agents
show good or ill will and this leaves them open to moral appraisal.
It is a further question whether reasons-responsiveness is necessary for moral agency because an agent must necessarily be able to act for moral reasons, such that they have a choice, or whether they need to be able to recognise what they have most reason to do. From now on, I refer to the kind of reasons-responsiveness that is problematic for reasons internalism as ‘strong reasons-responsiveness’. This is the sort of reasons-responsiveness that requires that an agent actually had most reason to behave morally but failed.

I take it that one can see how the weak reasons-responsiveness outlined is unproblematic for reasons internalism. It is unproblematic because this sort of reasons-responsiveness is only concerned with the fact that the agent is motivated by reasons at all. It does not matter the agent does in fact have a normative reason to behave morally, what matters is that the agent’s motivating reasons display good or ill will. It is clearly the case that the bad actors under reasons internalism are acting for reasons; it is just that they are morally bad reasons. So, if one thinks that reasons-responsiveness is important for moral agency only insofar as creatures that act for reasons act in a way that is indicative of their character, then an account which put it that blame is based on criticism of character could avoid the blame problem for reasons internalism. It fulfils The Non-Choice Condition because an agents actions can be indicative of their character irrespective of whether they had a choice as to whether to perform those actions. If Mr. Mogg is of a grumpy disposition such that he is only ever motivated to be grumpy and to grumble and complain, then one might still think him responsible for his grumpiness irrespective of whether he had a choice about it. And there is nothing which requires that it would fail to fulfil The Rationality Condition – whether it does so would depend on the individual account.

I highlight the distinction between weak and strong reasons-responsiveness with various motivations. One motivation is that I think it is hard to argue that reasons-responsiveness plays no part in moral responsibility at all. Even if holding morally responsible is contingent upon character, I take it that in order for your actions to mean anything in relation to your character then you have to be the sort of creature who acts for reasons.\textsuperscript{67} A second motivation is that I think it is important to note that internalists can agree that reasons-responsiveness is a valid condition for moral agency without committing to the stronger claim that agent’s must be able to respond to particularly moral reasons. That is, an agent can be reasons-responsive

\textsuperscript{67} It is a separate question as to whether you must know the reasons you are acting on, see Arpaly, *Unprincipled Virtue*.  


insofar as they have reasons and make judgments on the basis of their reasons without it being true that they ever respond to moral reasons. As such we can set up the argument that is relevant for internalism as between those theorists who endorse only weak reasons-responsiveness on one side, and those who endorse strong reasons-responsiveness on the other.

3.2 Attributability, answerability, accountability

I have framed the problem as between weak reasons-responsiveness (for internalism) on one side and strong reasons-responsiveness (against internalism) on the other. The focus now shall be on distinguishing between accounts of moral responsibility that require the weaker kind of reasons-responsiveness and those that require the stronger. Handily for our purposes, the distinction between weak and strong reasons-responsiveness maps on to a distinction between two sorts of moral responsibility: responsibility as accountability and responsibility as attributability. As a kind of disclaimer, it should be noted that my intention is not in the first instance to find the best, most plausible and coherent theory of moral responsibility full stop – that would naturally be far too great a task to undertake here. Rather, the aim is to look at some prominent accounts of moral responsibility to find an account which avoids or solves the blame problem for reasons internalism. 68

It will help to begin to draw up a basic conception of moral responsibility first, and to note any significant departures from this conception in discussions later. Many philosophers69 follow Strawson in saying that to hold an agent morally responsible is to judge her to be an apt candidate for the Strawsonian reactive attitudes.70 Roughly put, this means that an agent is morally responsible for her actions insofar as blame or praise are appropriate. I take it that blame could include sentiments like resentment or bad feeling (as Strawson intended), but I also take it that it might be enough to believe such sentiments are appropriate if they are not felt. 71 A further Strawsonian characteristic is that blame expresses a demand for good will.

68 notably from Watson, ‘Responsibility and the Limits of Evil’; Scanlon, What We Owe to Each Other; Smith, ‘Attributability, Answerability, and Accountability’.
69 It is hard to find a subsequent philosopher who has not been influenced by Strawson, but for example Wallace, Responsibility and the Moral Sentiments; Watson, ‘Responsibility and the Limits of Evil’; Fischer and Ravizza, Responsibility and Control; Levy, ‘The Good, the Bad, and the Blameworthy’; Talbert, ‘Blame and Responsiveness to Moral Reasons’.
70 Strawson, ‘Freedom and Resentment’.
71 Wallace, Responsibility and the Moral Sentiments.
For Strawson, this is a general demand that is held in virtue of our participation in interpersonal relationships, and it is when it is breached that we blame. I take it that the idea of blame as upholding this demand when it has been breached is also a widely held view.\(^{72}\) Because it appears that the demand for good will could vary within relationships or from person to person, for simplicity here I take it that the demand represents a demand to behave morally: to meet a moral standard.\(^{73}\) The result is an account in which an agent is morally responsible insofar as she is an apt candidate for blame, where blame is a Strawsonian reactive attitude which upholds a demand that the wrongdoer behave morally.\(^{74}\)

It is important to note that judging blame to be appropriate does not entail that the blamer must actually blame the wrongdoer in order for them to be held morally responsible. There might well be other considerations that affect whether it is appropriate to blame someone in any given situation, including prudential considerations. It should also be noted that blaming as I see it does not entail further sanctions like punishment. Whether such sanctions are appropriate is a separate question. Further, the construal above puts it that an agent can only be blameworthy for her actions. However, it might also include attitudes and omissions\(^{75}\), or restrict it to attitudes alone\(^{76}\). I do not think the object of blame much affects the ensuing arguments or the argument against internalism, and where I write that an agent is blameworthy for her action this should not preclude that she is in fact blameworthy for her attitude or vice versa.

However, I take it that the view of blame construed does generally match what we think of as blame in the everyday sense. Often blame does involve bad feeling towards the wrongdoer. I take it that expressing blame verbally might involve asking for justification from the wrongdoer for their actions and, if the wrongdoer can offer no justification or excuse, telling the wrongdoing of their wrongdoing and explaining the demands that they are expected to meet. This does not have to be as direct as I make out. For instance, it can be very clear from indirect action or from what the blamer has said to others or even from body language that a


\(^{73}\) It appears more usual to define moral responsibility merely by the claim that an agent is morally responsible if she is an apt recipient of the reactive attitudes. However, I think in all of the accounts I cite here, blame also entails – roughly put – a demand to meet some moral standard that has been breached. Hence why I include it here.

\(^{74}\) I just use the negative from now on for clarity

\(^{75}\) Levy, ‘The Good, the Bad, and the Blameworthy’.

\(^{76}\) Scanlon, *What We Owe to Each Other*.  

14
person might blame you. However, irrespective of the many nuanced ways in which the blamer might express their blame, it still seems that the blamer’s attitude is one of blame and that this expresses a demand to behave morally that has not been met by the wrongdoer. I refer to this sort of response at other times in the thesis as ‘blame proper’. This is necessary to distinguish it from another weaker sort of moral appraisal.

I have put it that an agent is morally responsible insofar as she is an apt candidate for blame, where blame is a Strawsonian reactive attitude which upholds a demand that the wrongdoer behave morally. As blame on this account is about holding up a moral standard and demanding an agent meets it, one sort of theory – accountability – puts it that an agent can only be blameworthy if it is fair and appropriate to hold them to this standard. Fair in the sense that the agent had sufficient opportunity to avoid the wrongful action and appropriate in the sense that the agent is the sort of being of whom you can make such demands. An advantage of this theory is that it links a plausible claim about responsibility – that you are responsible for something if you choose to do it – with a response that we also plausibly think of as blame – that being blamed involves being asked to justify what you did and to account for your actions. However, clearly it is these sorts of accounts that are incompatible with the blame problem. They necessitate that the agent had a choice as to which reasons to act on but chose to act immorally. I assume that given that accountability requires that an agent had a choice as to which reason to follow that they would also agree to the second point that the agent is rationally faulty because they acted against what they had most reason to do. So let us take it that accountability as moral responsibility would fail both The Non-Choice Condition and The Rationality Condition.

A different theory of moral responsibility need rely only on the weaker reasons-responsiveness condition, however, if it relies on reasons at all. On this account, an agent is responsible insofar as her actions are an expression of her character. She is responsible for her actions in the sense that they are attributable to her. Attributability as moral responsibility avoids the problem from blame because it relies only on the weak condition that agents act for reasons at all such that their actions are really attributable to them. However, attributability as moral responsibility faces some substantial obstacles. The first is whether it

77 Levy, ‘The Good, the Bad, and the Blameworthy’.
78 Watson, The Trouble with Psychopaths 1.
79 See Darwall, The Second-Person Standpoint, 88. for a defence of this point, in addition to my defence in Chapter 2.
is a form of responsibility at all, given agents seemingly have no control over their characters. The response put forward by Gary Watson is that it is a form of responsibility because it warrants moral appraisal. A blamer can appraise an agent’s moral character: “What a bad person!” and if their action warrants this appraisal of their character then, it can be argued, it is something for which the agent is morally responsible. I take it, therefore, that the attitude that is appropriate under attributability is different from that in the roughly Strawsonian model first construed.

A further difficulty, however, is that the sort of moral appraisal that seems to be entailed by attributability is not blame proper – saying ‘he is bad’ is not equivalent to blaming him. While some in the debate are happy with this sort of aretaic appraisal in regard to attributability, others put forward an account in which the conditions of attributability are sufficient to warrant blame proper. The answerability theorists, notably Tim Scanlon (1998) and Angela Smith (2008), argue forcefully that an agent can be blameworthy in the fullest sense of the word for judgment sensitive attitudes which issue from their character. They are blameworthy because they are answerable for these attitudes. Since being called to answer for yourself is ordinarily considered to be a feature of blame proper, then if you can be called to answer for attitudes that are rooted in your character then it seems you can be truly blameworthy on attributability conditions.

Much more will be said about accountability, answerability and attributability, however from the picture we have thus far I hope it can be seen that some alternatives are open to internalism. One potential route is to argue with the answerability theorists that the conditions for attributability warrant blame proper. This option could mitigate the problem from blame for internalism because it supposedly relies only on the weaker reasons-responsiveness condition. Insofar as an agent acts for reasons at all then they could be blameworthy. Since the difficult characters for reasons internalism all act for reasons then it seems that they would no longer be difficult cases. It would be entirely apt to blame them irrespective of whether they actually had a reason to behave morally.

---

80 Levy, ‘The Good, the Bad, and the Blameworthy’.  
81 ‘Responsibility and the Limits of Evil’.  
82 Watson, ‘Two Faces of Responsibility’.
A different but still plausible route is to argue with Watson for a bipartite theory of moral responsibility. The weaker sort of attributability that he endorses is insufficient on its own to entail all that moral responsibility constitutes since it only engenders aretaic appraisal rather than blame proper. However, if one could accept that some agents are attributability blameworthy without being accountability blameworthy (while other agents are both) then this again offers a potential solution for the internalist. It could then be that the difficult characters for reasons internalism, though not accountability blameworthy, are at least blameworthy in an attributability sense. This might explain why it seems that we blame characters who under internalism have no reason to behave morally.

In the next section I put forward a case for answerability. I deal with answerability first because I see it as the weaker account for several reasons that I will bring out in the discussion. One reason is that I find the move from attributability conditions to accountability blame unconvincing without some sort of appeal to choice (then it would fail The Non-Choice Condition). This would then render internalism susceptible to the blame problem once more. On a different interpretation I am concerned that the account fails The Rationality Condition and seems in some ways to entail externalism about reasons. So a worry is that it is more fundamentally incompatible with internalism than I have presumed. In the following section I then put forward the case for a bipartite theory of responsibility and the consequences that would have for internalism.

3.3 Answerability as moral responsibility

Answerability as moral responsibility is the name often given to a group of similar views which hold that an agent is morally responsible for their judgment-sensitive attitudes. I take this to be roughly the views of Tim Scanlon and Angela Smith. A good outline of these views is also put forward by David Shoemaker as part of his account of a tripartite theory of moral responsibility. Smith, for instance, writes that ‘to say that an agent is morally responsible for some thing, on this view, is to say that that thing makes it appropriate in principle, to ask her to defend or justify it’. Any attitude which an agent holds which is formed on the basis of judgments about reasons is open to this sort of criticism. Smith cites

---


84 Smith, ‘Control, Responsibility, and Moral Assessment’, 370.
‘desires, emotions, beliefs and other attitudes’ as the sorts of things for which agents are morally responsible and for which they can be asked to answer when being held morally responsible. These attitudes are judgment dependent because ‘they generally reflect and are sensitive to our (sometimes hasty, mistaken or incomplete) judgments about what reasons we have, and they are generally responsive to changes in these judgements.

I take it that on an answerability view what makes the agent responsible for the attitude is supposed to be that these attitudes reflect her judgments, and that these judgments are themselves constitutive of what an agent values, of her character. These attitudes are attributable to her in virtue of this. For now, I assume that the answerability theorist would agree that this requires only the weak sort of reasons-responsiveness, such that an agent acts (or more accurately possesses attitudes) on the basis of her rational judgments at all and as such would avoid the problem from blame for internalism. One indication this is what the internalist means is that answerability theorists argue for the inclusion of psychopaths whose particular deficiency is that they cannot act for moral reasons. However, as we shall see in some ways this stance is unclear. As I have noted, a common charge-levelled at such a view is that an agent cannot be responsible in this way because they have no choice as to their character. To be responsible for something seems to entail that you have some control over the outcome of that thing, but attributability does not entail that.

Answerability theorists, though, take it that attributability is the proper basis for moral responsibility. They argue this because, roughly put, blame addresses a demand that the agent’s attitudes meet a moral standard. Smith writes that moral assessments involve ‘implicit demands for justification, demands that are made on the basis of moral standards we expect reasonable persons to accept’. Scanlon meanwhile puts it that moral criticism questions an agent’s judgment sensitive attitudes and calls for the agent to revise them. One can make this type of demand of an agent on the basis that their attitudes are in principle ‘under the control of reason’. If what blaming is is a demand for justification from the

---

85 Smith, 370.
86 Smith, 370.
87 Scanlon, ‘Reasons, Responsibility, and Reliance’.
88 Much more is said about this criticism of attributability in the next section.
89 This aspect is the same as in the roughly Strawsonian view I drew up earlier. I take it, however, that the answerability theorist tends to deny that blaming involves any particular bad sentiment or punitive response.
90 ‘Attributability, Answerability, and Accountability’, 369.
91 Scanlon, What We Owe to Each Other, 287.
92 Scanlon, 272.
wrongdoer, then they are blameworthy just in case they can provide answers to this demand. They can provide answers to this demand if they possess judgment sensitive attitudes; if they are agents who act on the basis of judgments about reasons at all.

On the brief picture I have drawn answerability can, at the outset, solve the blame problem. This is because, what we are left with is an account of moral responsibility that includes within the realm of moral agency anyone who can be answerable for their judgment sensitive attitudes. This is anyone who acts or whose attitudes are based on judgments about their reasons. This includes the difficult characters for reasons internalism because the very problem is that they act on the basis of reasons (its just that these reasons seem to be the wrong ones). But the move answerability theorists try to make is to suggest that this capacity is sufficient for such an agent to be blameworthy to the full extent normally associated with accountability blame. Being called to answer for your actions is a feature of blame proper, not a weaker sort of descriptive appraisal that claims you are bad. As such, if answerability succeeds and if it is compatible with internalism then there is no problem with regard to blame.

However, I think various criticisms can reasonably be levelled at answerability as moral responsibility. One move is just to deny that all attitudes are judgment sensitive in the way the account requires. For instance, I take it that Jay Wallace denies that emotions are judgment sensitive in the way that would be required for you to be morally responsible for them.93 I think that this objection is a worthy one. Nevertheless, there are more pertinent concerns for our purposes here. I follow Levy in his argument that answerability seems in some ways to rely tacitly on the condition that the agent has a choice and chooses to act badly, though this is a condition answerability theorists claim to deny.94 This is problematic because if it transpires that the agent must have had a choice as to their judgment sensitive attitudes, and further that this choice must have included the option to behave for moral reasons, then reasons internalism is susceptible to the blame problem on an answerability account.

---

93 Wallace, Responsibility and the Moral Sentiments, 131–32.
94 Levy, ‘The Good, the Bad, and the Blameworthy’.
Levy puts it that answerability theorists argue that judgment-sensitive attitudes are, in principle, within the control of the agent.\textsuperscript{95} Indeed, what makes an agent responsible for these attitudes, rather than merely a possessor of these attitudes is that they are subject to change. They are judgement sensitive such that if an agent’s judgments about reasons change then the attitude should change to reflect this. To borrow Levy’s example, these attitudes are not like our hair or eye colour, we are responsible for them because they are sensitive to judgments about what reasons we have. This is what makes asking an agent to answer for them appropriate in a way that it would not be appropriate to ask her to answer for her hair colour. However, if this is the case then it seems clear that, in making judgments about reasons, agents are making choices about what are or are not reasons for them. Otherwise it is difficult to see what making a judgment would actually entail.

Scanlon’s response to this is that attitudes are subject to rational judgements ‘in principle.’\textsuperscript{96} They are within a realm of things that are in and of themselves controllable but need not necessarily have been controlled at the time, nor over which that particular agent need actually have active control. This is because in principle we can be called upon to defend our attitudes irrespective of whether we actually deliberated, or irrespective of whether we could have deliberated – attitudes are just of the class of thing that can in principle be defended and justified by reasons. This justification is important because it is the demand for justification which distinguishes moral criticism from mere description.\textsuperscript{97} So because blame entails that an agent answer for their attitudes, it seems anyone who has judgment-sensitive attitudes is potentially liable to be held responsible for those attitudes. As such, any agent who makes judgments about reasons, or who at least has the capacity to make judgments about reasons, is a moral agent.

But this is something of an odd claim. As Levy notes, to be under the control of something ‘in principle’ just is not sufficient for responsibility. He writes ‘I don’t have a kind of ertsatz control over my car if the steering wheel falls off; the fact that cars are in principle controllable does not alter my lack of control in that particular circumstance.’\textsuperscript{98} And this seems to be fundamentally correct. What matters is surely actual control over one’s attitudes, not that attitudes are the sort of thing that are in principle controllable. If this were true then

\textsuperscript{95} Levy, 10.
\textsuperscript{96} Scanlon, \textit{What We Owe to Each Other}, 272.
\textsuperscript{97} Smith, ‘Attributional, Answerability, and Accountability’, 381.
\textsuperscript{98} Levy, ‘The Good, the Bad, and the Blameworthy’, 10.
any possessor of an attitude could be blameable irrespective of whether they were capable of answering for it. The emotions of an extremely depressive person would I think constitute the sort of attitudes that are in principle the kind of thing that an agent could answer for, but I do not think it would be fair to argue that the agent is responsible for them when they make her behave erratically just because they are the kind of thing that could be justified by reasons. We assume in such cases that she is expressly not responsible because she is in the throws of something beyond her control. No-one is going to ask her to account for her bad mood or erratic on the basis that she is in the throws of something beyond her control.

Moreover, curiously, Scanlon sometimes seems to agree with this assessment. He writes that an agent acts badly if they are in full awareness of the relevance of some considerations to a moral principle which counts as a ‘conclusive reason’ against a certain course of action, but they choose to perform the action anyway; or if they perform it due to ‘faulty self-governance’. 99 On this assessment then it seems that choice does matter, because an agent must actively choose not to perform the action or fail to perform it through faulty self-governance. Faulty self-governance can be taken to mean several things. It could mean that an agent judged something that was not in fact a reason to be a reason. 100 It could mean that an agent judged something to be a reason but then was lazy or forgetful and failed to perform the action. But, either way, I would argue that all of these options entail choices, because for an agent’s actions to be faulty there has to have been an opportunity for the action to be faultless. I do not think that these choices need always be consciously made, voluntarily-taken choices, but it seems that it was at least the case that the agent had the opportunity to avoid doing what they did. This interpretation would bring about the blame problem because the account would then fail The Non-Choice Condition.

To this, I suspect the answerability theorist would reply that an agent’s judgment can be faulty irrespective of whether they know it is faulty. The can behave for the wrong reasons irrespective of whether they know what the right reasons are. Scanlon puts it that one can be criticised rationally for their bad chess moves irrespective of whether they are aware that they are rationally faulty. 101 This claim is dubious on a number of grounds, not least of which is that when an agent is criticised on the basis of their rationality it does not usually constitute

99 Scanlon, What We Owe to Each Other, 269.
100 In which case it would presuppose externalism, a point I come to later..
However, for our purposes, what is much more significant is that this move renders answerability incompatible with internalism as I have presented it, because bad agents for internalism are not rationally faulty. So the account would fail The Rationality Condition as I put it earlier.

In order to explain why the account would fail The Rationality Condition it helps to compare how answerability theories operate with how the internalist theory operates, with regard to the formation of reasons, desires and attitudes. On an answerability account an agent is responsible for her judgment sensitive attitudes. If these attitudes are morally bad, and if they cannot be justified by her reasons when she is called to answer for them, then she is potentially blameworthy. I take it that on both Scanlon’s and Smith’s account the process is seen as being something like the agent taking various things that she values to count for or against her doing certain things, that is to be reasons for her. She deliberates between these reasons and this process constitutes her rational judgement. The judgments she makes then go on to inform her attitudes. So the explanatory direction moves from reasons, to rational judgment, to (potentially blameworthy) attitude.

This is perhaps explained more clearly by example. My father is an avid supporter of Ipswich Town Football Club. Every season the same reasons persuade him buy a season ticket: he has supported them all his life; he is holds loyalty very highly; he generally enjoys watching football. These are weighed against other reasons: tickets are very expensive; Ipswich Town are not a very good football team and therefore the football may be terrible. His rational judgment leads him to the belief that it is rationally sensible to buy a season ticket, he has that attitude and he goes and buys one. On an answerability account, my father’s belief in certain considerations counting towards him buying the ticket, that is his belief in what his reasons are, influences his attitude. This, I would suggest, differs from the internalist account in that the direction of explanation is the other way around. On Williams’ account the attitudes that you have, and the other component parts of your motivational set are what give you reasons. Following rational deliberation it might be that you most want – and as such what you have most reason to do – is different than you thought, but it cannot come from something entirely outside of your set at all.

Let us return to Scanlon’s claim that an agent acts badly if they are in full awareness of the relevance of some considerations to a moral principle counts as a ‘conclusive reason’ against
a certain course of action, but they choose to perform the action anyway; or if they perform it
due to ‘faulty self-governance.’¹⁰² Faulty self-governance, Scanlon explains, can include an
agent’s ‘failure to notice this reason simply out of indifference or because he is excited at the
prospect of some imminent success.’¹⁰³ One can see how an agent’s actions could be faulty
on the direction of explanation that I put forward for the answerability account above: If it
turns out that my father has promised my mother that he ought in fact to give up the football
because he has a compulsion to gamble on the result and has lost a substantial amount of
money then a failure to factor this reason into his deliberation does seem like faulty self-
governance. This is something we would take to be a reason for him, in an external sense,
irrespective of what he wanted or whether he considered it a reason, so he is faulty because
he has not acted on the basis of what he had most (external) reason to do. However, if, on an
internalist account, my father truly had nothing within his motivational set to make him want
to keep his promise to my mother following rational deliberation, then this consideration is
not a reason. So in what way is he then faulty? He has deliberated and come out with the
belief that what he would most desire is to buy a season ticket; as such he has most reason to
buy the ticket. His behaviour, even in full awareness of the promise to my mother, is
completely in line with rational self-governance.

Now one could argue that the way I have construed the move from reasons to judgment to
attitude in the first place is not quite right. It could be that the reasons an agent judges to be
reasons are influenced by her pre-existing attitudes, and that judgments about these reasons
then shore up the pre-existing attitudes or cause the agent to form attitudes. Scanlon is an
externalist about reasons but I suspect that he would agree to this slightly messy account as
being what the formation of our judgment-sensitive attitudes is like in practice. Indeed,
although he writes against Williams in the final chapter of ‘What we owe to each other,’¹⁰⁴ I
take it that his general assessment in the chapter is that it does not make much difference if
one is an internalist or an externalist about reasons.

The natural solution is that if my father is not rationally faulty, he is morally faulty such that
he is bad for buying the ticket (indeed this might be the point Scanlon is trying to make given
his subsequent discussion). But indeed, even assuming that this is the case, if answerability is

¹⁰² Scanlon, *What We Owe to Each Other*, 269.
¹⁰³ Scanlon, 269.
¹⁰⁴ Scanlon, *What We Owe to Each Other*, 363-73.
a form of responsibility because the agent is supposed to answer for what they have done, I fail to see how my father could meet that requirement in the right way. When answering for why he bought the ticket, my father – being, of course, well-versed in the workings of reasons internalism – could reply ‘well, I just was not at all motivated to keep my promise even following a period of rational deliberation, as such it just was not a reason for me’. This would be a fair and legitimate response. My point is not that my father is not bad for breaking the promise – he is clearly bad; my point is that his rational judgment is not faulty.

The problem as I see it is this: one way of interpreting answerability is that the agent’s reasons reflect her character because they are contingent upon choices she makes as to which reasons to adopt. These reflect her fundamental values. However, it seems that in this case the agent’s responsibility for her judgment-sensitive attitudes is in virtue of the fact that these are reasons she chooses insofar as she makes rational judgments. Or rather, even if she does not actively choose them, she has the capacity to choose them. This capacity is necessary so that she can be called upon to defend her attitude with these reasons if necessary. However, on this account answerability then fundamentally relies on choice. As we have seen this is problematic to internalism because the account then seems to require stronger reasons-responsiveness and then fails to fulfil The Non-Choice Condition.

On the other interpretation, it does not matter whether the agent has a choice about what reasons constitute her rational judgment. She is faulty just because her judgment is rationally faulty and on this basis she can be blamed. However, in order for her judgment to be rationally faulty there must be right reasons – moral reasons – that she ought to have acted for. Her judgment is faulty because the reasons she would give to justify why she acted do not match what her actual normative reasons are. This account can explain why the agent is answerable because blame then functions as a form of rational criticism after which one must put forward their reasons, however on this account answerability fails to fulfil The Rationality Condition.

The brief argument I have provided cannot be expected to form a comprehensive case against answerability, however I hope it is sufficient to cast doubt on whether answerability can really help internalism, either because it relies tacitly on choice and therefore cannot meet The Non-Choice Condition, or because it relies too heavily on the idea that blame is a form of rational criticism and therefore fails to meet The Rationality Condition. More will be said
against answerability in Chapter 4. I now move on to the other theory that I think could help the internalist: attributability and accountability as a bipartite theory.

### 3.3 Attributability and Accountability as a Bipartite Theory

The purpose of this section is to outline more clearly an alternative to answerability: the bipartite attributability and accountability theory of moral responsibility. This view was put forward and discussed by Gary Watson in the well known essay ‘Two Faces of Responsibility.’ I take it that this view holds that there are two types of moral responsibility: attributability and accountability, and these two types of responsibility can operate side by side. Attributability puts it that an agent is morally responsible insofar as an agent’s morally relevant actions or attitudes are attributable to her; and accountability which is often taken to be the view that an agent is morally responsible for her actions or attitudes insofar as it is appropriate and fair to respond to her with the reactive attitudes. The previous section dealt with answerability which puts it that attributability conditions are sufficient for blame proper. The bipartite theory, meanwhile, puts it that to be faulty under attributability warrants only aretaic appraisal, while the wrongdoer must meet the conditions for accountability to truly warrant blame. The purpose of this discussion to open up a route out of the blame problem for internalism because, although bad characters cannot meet the conditions for accountability, it seems they do meet the conditions for attributability and can be subject to the weaker form of appraisal on the basis of the latter.

Attributability is, in a sense, prior to accountability because an action must be attributable to you before you can be accountable for it. Attributability theories are concerned with what an agent’s actions or attitudes say about an agent’s character. They are often termed ‘deep-self’ theories in that your actions and attributions, insofar as they are done with intent,

---

105 Watson ‘Responsibility and the Limits of Evil’.
106 This view is denied by David Shoemaker Shoemaker, Responsibility from the Margins. in ‘Responsibility from the margins’ where he argues attributability, answerability and accountability are each independent forms of responsibility with independent conditions. I take it as fairly self-evident that any agent who is blamed under accountability can also be appraised aretaically. If someone does something blameworthy then it is hard to argue that they are not also bad, mean, selfish, callous or some other morally-charged adjective. So if attributability responsibility is engendered by aretaic appraisal then it seems an agent must be attributability responsible prior to being accountability responsible.
107 I take it that this term was first coined by Wolf, Freedom within Reason.
reflect some aspect of your character that is subject to appraisal. These sorts of views face two significant and related challenges: one, is how such a view could entail that an agent is morally responsible, particularly given that there is a sense in which we cannot help our characters; two, is whether (in light of point one) attributability legitimately entails blame, appraisal or mere description. I cannot hope to substantively defend attributability from the first challenge here – but given that I do not intend to argue for attributability as a comprehensive theory of moral responsibility perhaps that is not such an issue. All I hope to show is that we can justifiably attribute claims of ‘badness’ or ‘goodness’ to an agent when they perform bad actions, and that therefore bad characters can be appraised aretaically even if they cannot strictly be blamed.

Regarding the first challenge, a claim against attributability is that it is superficial and vastly over-inclusive. This is the sort of claim levelled by Susan Wolf.\(^{108}\) She writes that ‘when we praise or blame an individual in the superficial sense, we acknowledge that the individual has good or bad qualities, or has performed good or bad acts. But when we hold an individual morally responsible for some event, we are doing more than identifying her particularly crucial role in the causal series that brings about the event in question’.\(^{109}\) Wolf’s complaint is that moral responsibility is not equivalent merely to causal responsibility. She sees attributability as being guilty of the two. And this then widens the field of possibly blameworthy agents to dogs and cats, hurricanes, children, people with mental illnesses and so on. Therefore attributability risks making the bounds of moral agency ridiculously over inclusive.

But such a challenge is unfair. Naturally, one aspect that is required to be morally responsible at all is that you performed the action in question (or hold the relevant attitude if you can indeed be blameworthy for attitudes). It is therefore true that you are the causal originator of the bad event – but this condition would be true in order for you to be blameworthy in any sense, not merely an attributability one. Hence, what is usually claimed is that the agent is both the causal originator of the bad action and that they meet some sort of attributability conditions. These might include the agent having the relevant sort of character or that the

\(^{108}\) Wolf, 40.
\(^{109}\) Wolf, 40–41.
agent acts for their own ends. A highly plausible claim is that the agent must be rational such that their actions are capable of actually issuing from their character. This is what allows attributability accounts to be compatible with the claim that reasons-responsiveness is a necessary condition of moral agency (as I put it earlier).

However, a substantial difficulty for attributability as moral responsibility is how an agent can be responsible in such a way as to be blameworthy given that they have no choice as to their characters (as per the discussion of answerability above). If blame is a matter of desert such that an agent must deserve the negative responses associated with blame, then it seems unfair that they should suffer these responses given that attributability does not require that they had a choice as to whether to be bad or good. The response given by Gary Watson to this problem is that attributability is sufficient to warrant moral appraisal and that alone makes it a form of moral responsibility. It entails a claim about the moral worth of the agent, rather than merely the moral worth of the action, so it is in this sense that the agent is responsible.

But then what does holding someone to be attributability blameworthy actually entail? That is, what sort of responses are appropriate to an agent being negatively morally responsible in an attributability sense? To elucidate further, Watson who cites a case from Peter van Inwagen in which a colleague of van Inwagen’s denies the reality of moral responsibility. Yet, when that same colleague’s books are stolen he exclaims: “That was a shoddy thing to do!” To van Inwagen, the exclamation “That was a shoddy thing to do!” means that the colleague believes in moral responsibility. He has theoretically denied the reality of moral responsibility but in practice, at least at the outset, it appears that he is holding the thief morally responsible. However, as Watson attests the situation is somewhat more complicated than it first appears. To call behaviour ‘shoddy’ is to see it ‘as a poor exercise of human evaluative capacities, as characteristic of someone who cares little about standards of excellence in human affairs’. To evaluate someone’s behaviour as shoddy is to judge them to be of bad character, ‘shoddiness’ in this scenario is a vice. So in one way, van Inwagen is right and his colleague is mistaken if he denies moral responsibility at all and yet calls the agent’s behaviour shoddy. The wrongdoer is being held to be morally responsible in an

---

110 A view of the first sort can be found in Arpaly, Unprincipled Virtue; Watson, ‘Two Faces of Responsibility’. I think it sensible to endorse both that an agent has the relevant character and that they act for reasons.

111 Watson, ‘Two Faces of Responsibility’.

112 Watson, 230.
attributability sense because the attribution of shoddiness is being applied to him. So attributability is the attribution of some quality to an agent which has a moral relevance: selfish, lazy, bad etc.

Attributability differs from accountability in that it can apply in non-moral cases. This is because the aretaic perspective is in some cases moral, and in other cases clearly not. The colleague’s exclamation of ‘that’s shoddy’ clearly has a moral dimension. The action, theft, is morally bad, and to be a thief is also morally bad. However, to be a ‘bad businessman’ or a ‘lousy artist’ is not, in and of itself, morally bad. Further, as Watson argues, even if we take ‘all such appraisals to be moral in a broad sense, they are independent of the particular moral norms that are invoked in accountability’\textsuperscript{113}. What I take him to mean here is that the broader sort of moral concerns around being a bad person, do not amount to the moral considerations (or indeed moral reasons) that are the subject of accountability blame. I take it that accountability blame differs because it entails a demand – it requires something of the agent (as per my stipulative account at the start of 3.2). I take it that for Watson these interpersonal demands form the basis of morality. Demands entail that an agent has failed to respond to the relevant moral considerations – for Watson I take it that the demands are constitutive of moral norms. However, either way the demands bring about the problem with choice. ‘You should not have done that’ entails, at least, that there are strong moral considerations that you ought to have taken into account that you did not. However, attributability does not require this sort of claim. I can say ‘you are bad’ irrespective of any claims about your responsiveness to moral considerations.

To summarise, in attributability the agent is responsible insofar as their attitude or action can be attributed to their character. The conditions that must be met on the part of the agent are that they really do have the attitude or commit action, and that they have the relevant sort of character that they can be good or bad and – for the reasons I put forward earlier – that they act for reasons at all. As such, attributability requires only weak reasons-responsiveness and, in principle, can fulfil The Non-Choice Condition and The Rationality Condition. Accountability, meanwhile, is contingent upon choice as I described it in the introduction and requires that an agent had most reason not to do what she did. When someone is appraised in an attributability sense we can say that they are bad or shoddy or obstinate etc. When an

\textsuperscript{113} Watson, 231.
agent is truly blameworthy in an accountability sense however, we believe that the reactive sentiments are appropriate in response to her attitude or action, such that we truly blame her and hold her to account for what she has done.

This distinction has the potential to be good for Williams, and – helpfully – this is precisely what he proposes in ‘Internal Reasons and the Obscurity of Blame’.\(^{114}\) There, he puts forward the possibility of a man who has no internal reason not to beat his wife; that is there is truly nothing within his motivational set that could give him reason not to beat her. Williams acknowledges that, in a strict sense, we cannot really blame this man for beating his wife, because blameworthiness requires that the agent had reason not to do what he did. However, he then claims that we can still say that it is ‘bad’ that the man beats his wife. The attribution that the husband is bad is distinct from whether or not he is blameworthy on Williams’ account. And the distinction that Williams draws appears to be cognisant with the distinction between accountability and attributability blame put forward by Watson. The two differ in that Williams does not appear to regard aretaic appraisal as ‘blame’ while Watson accepts that some part of what constitutes blaming is aretaic appraisal. However, this difference seems to be nothing more than semantic. Both distinguish between the practice of aretaic appraisal, or blame in an attributability sense, and the more standard sort of accountability blame. I would also suggest that the fact that Williams puts forward aretaic appraisal in lieu of stronger accountability blame, suggests that it is in some sense blame or is supposed to take the place of blame in situations where the agent truly has no moral reasons. He writes that we can still say the man is ‘ungrateful, inconsiderate, hard, sexist, nasty, selfish, brutal’.\(^{115}\) His point is essentially that we cannot blame the bad husband, but at least we can still evaluate him in some morally relevant way. And, according to what Watson has propounded, this point is not fundamentally inconsistent.

Let us assume that there is a distinction between true accountability blame and the sort of aretaic appraisal that constitutes attributability blame. Attributability can then meet the two conditions I have proposed but this point does not, on its own, solve the problem from blame first put forward. In order to do this, the internalist would need to show several things. The first is that there are truly cases in real life where we blame people but only in an attributability sense – or rather perhaps where the case can reasonably be put forward that we

\(^{114}\) Williams, ‘Internal Reasons and the Obscurity of Blame’, 39.

\(^{115}\) Williams, 39.
ought to blame these people only in an attributability sense. As I will go on to show, I think this case can reasonably be made for psychopaths and perhaps even other individuals of bad character. The second claim that needs to be substantiated is that those who are excluded from blame in an accountability sense but are still liable to be blamed in an attributability sense are the same people who are not accountability blameworthy on Williams’ account.

For instance, it could be fair to say that those adults who have a personality disorder which impairs their emotional responses are likely to be excluded from accountability blame, and Williams’ internalism can accommodate this. However, it is far less clear in the case of Williams’ wife-beater, for instance, whether he truly is not worthy of accountability blame. A further facet of this point is that we just do actually seem to blame bad people in an accountability sense, though on Williams’ account they can only ever be attributability blameworthy. The leads to difficult questions: for instance, if it is the case that Hitler was not motivated – or insufficiently motivated – following a period of rational deliberation not to order the extermination of the Jewish people then internalism allows that he had no reason not to. The question is: Is it enough to say only that his actions were bad, rather than that he ought to be called to account and punished for them?
4 Bad Characters

The purpose of this chapter is to ascertain whether the characters who are problematic for internalism ought to be excluded from the sphere of moral agency. As I have put it, when we blame it assumes that the wrongdoer had most reason to behave morally. This entails that the blamed party must have been able to act on the basis of the moral reason and that this reason carried the most weight. The following cases represent the kinds of characters that I think are problematic for reasons internalism under the conditions of accountability. Their names are, with the exception of the psychopath, somewhat arbitrary. I take it, for instance, that an extremely egoistic person in real life could well also be a psychopath. What differentiates the three in my examples is the depth of their responsiveness to moral reasons – the psychopath being completely unresponsive, the mob gangster being somewhat unresponsive and the egoist being responsive but not sufficiently responsive. The aim will then be to see if we think it is acceptable to exclude these characters from the sphere of moral accountability, and if we think it is sufficient to blame them (or rather aretaically appraise them) under the conditions of attributability.

Case 1) The psychopath. I take the psychopath\textsuperscript{116} to be an agent who has absolutely nothing within their motivational set which could give them reason to behave morally. They are completely morally blind. Not only do they have nothing within their motivational set which could make them behave morally, but they never could have. I take this to be the case because they are not capable of understanding moral requirements and are therefore unmotivated by them. To put this within the framework of Williams’ internal reasons, they would not only fail to have moral reasons subsequent to a period of rational deliberation, but they lack any disposition at all which could lead them to ever have any moral reasons in the future. As such, they do not and could never have the opportunity to act for the right reasons. Hence, they would be outside the sphere of accountability according to internalism on the grounds that they do not have the relevant choice.

\textsuperscript{116}This is not intended to be a psychological definition of a psychopath. It merely notes that such morally blind agents are often considered psychopaths in the philosophical literature. Irrespective of the label, I take it that the reader can acknowledge the possibility of an agent who is completely unreachable by moral address and unresponsive to moral considerations.
Yet, at the same time psychopaths are capable of acting for reasons in the same way as other morally normal actors, so there is nothing that precludes them from having internal reasons in Williams’ sense. They can be vicious and manipulative and seem to fall into the collection of characters who we would ordinarily consider to be morally blameworthy.\textsuperscript{117} They are in this way unlike a bit of furniture that gets in the way, or a hurricane that tears through a city. They act with intent for reasons, and they are the sort of character who could answer with reasons if they were held to account.

Case 2) The mob gangster. He is someone who finds that he has some desires to do the right thing following rational deliberation in some cases but not in others. As such he has some moral reasons some of the time. A classic example of this sort is the mob gangster who loves his family very much, is very loyal and even shows bravery in some circumstances. This man drives his ailing mother to the hospital for every appointment, and yet sees no reason at all to refrain from shooting a policeman in the face during a bank robbery. On internalism, if the mob gangster truly has no desire at all not to injure the policeman following rational deliberation then he has no reason not to do it. This is problematic for internalism because it seems that in certain circumstances the mob gangster truly has no reason to do the right thing and is, as such, not blameworthy. Again, it seems as though the mob gangster falls foul of the choice problem because it was not possible that at the time of the bad incident for him to have acted upon the right reason.

Case 3) The egoist. The egoist is someone who most often has some desire to do what is morally right, but always has a stronger desire to do whatever is best for themselves. As I have previously argued, this seems to mean according to internalism that the egoist has more reason to do what they most want than what is morally right. They are not lacking in moral reasons at all, but they are of a sort whose moral judgment we would usually consider to be faulty. If it is agreed that blameworthiness has to involve some failure of moral judgment, as I have argued, then the egoist is problematic because they are not failing in their rational judgment when they choose to do something that is morally bad. Although the egoist is an extreme example, I take it that most of us are on some occasions like the egoist, and as such I see this case as particularly problematic for the internalist.

\textsuperscript{117} For a detailed psychological account of psychopathy see Hare, Neumann, and Widiger, ‘Psychopathy’.
The egoist is an interesting case however, because they are not necessarily exclude on the grounds of choice. Given the egoist has some motivation to behave in the morally right way then it could be argued that it is possible for the egoist to behave in the morally right way. However, according to internalism the egoist would be rationally faulty for behaving in the morally correct way. And as I have argued, if one thinks that blame entails that the agent had most reason to act morally and failed then the egoist too must be excluded from accountability blame. Otherwise it would entail that they were being rationally criticised for acting on the basis of what they had most reason to do.

4.1 Are Psychopaths Blameworthy?

Handily for our purposes, the question of whether psychopaths are morally blameworthy for their bad actions is generally bisected in the philosophical literature by the same two views I have dealt with here. On one side are those answerability theorists who argue forcefully that psychopaths are morally responsible in virtue of the fact that they hold judgment sensitive attitudes. On the other side, usually, are accountability theorists, or those who take a bi- or tripartite view of responsibility, who argue that psychopaths are not blameworthy precisely because they cannot recognize or respond to moral reasons. As I have put it previously they fail to be responsible because of the choice problem.

The purpose of this discussion is to see whether psychopaths ought to be blameworthy generally – irrespective of the truth or falsity of reasons internalism. This is because externalism is also compatible with the psychopath being excluded from the sphere of moral accountability. The psychopath, as I have construed him, is incapable of understanding moral considerations. One can easily hold that moral considerations are in fact reasons, external reasons, and agree to the condition that the agent must be capable of understanding these moral reasons in order to be held accountable. So, if it turns out that psychopaths ought to be excluded then this represents an argument in favour of accountability as moral responsibility. This argument will be compatible with internalism because psychopaths will necessarily be excluded under accountability on an internalist view. However, it would then remain to be seen whether the mob gangster and the egoist could also be excluded on the same grounds. If, on the other hand, it seems the answerability theorist is correct then this would be
problematic because it would require an interpretation of answerability that seems to rely on the truth of externalism (as I showed earlier).

I will argue that psychopaths ought to be excluded from the sphere of moral responsibility in line with the conditions of moral accountability. It further seems that they could be held responsible under weak attributability conditions. This is, as such, beneficial for internalism. The argument is principally in the negative – against answerability – because, as I have stressed, I take the condition of choice to be inherently plausible so the burden of proof is on those who wish to refute it. The argument refers to the answerability theorists I have discussed previously – particularly Scanlon^{118} -but also draws heavily on the work of Matthew Talbert^{119} who argues specifically for the inclusion of psychopaths within the sphere of moral agency on answerability grounds.

To re-cap, briefly on the answerability view, an agent is morally responsible for an attitude or action if they are answerable for it. This answerability amounts to them giving reasons to justify their actions. And they can give reasons for their actions if they are the sort of agent who acts on the basis of judgement sensitive-attitudes. There is no requirement, however, that the agent necessarily has the capacity to respond to moral reasons in the formation of these attitudes. This means that the scope of moral agency is pretty much equivalent to the scope of rational agency, as Talbert writes:

> the divide between blameable and non-blameable creatures is marked by the distinction between beings who are assessors of reasons in a very general way (but not necessarily responders to specifically moral reasons) and beings that lack rationality to such a degree that we would not describe them as making decisions on the basis of judgments about reasons^{120}

So are psychopaths rational in the way required? Psychopaths (as characterized in the philosophical literature at least) are often considered capable of acting for prudential reasons, for instance. Indeed, Talbert describes a murderous psychopath as ‘one who acts with cool

\[^{118}\text{Scanlon, What We Owe to Each Other; Scanlon, ‘Reasons, Responsibility, and Reliance’.}\]
\[^{119}\text{Talbert, ‘Blame and Responsiveness to Moral Reasons’.}\]
\[^{120}\text{Talbert, 519.}\]
deliberation and for reasons of her own. This claim, however, is more controversial than one might think. Stephen Finlay points out that psychopaths have deficits in self-regarding emotions and are often guilty of impulsive and imprudent behaviour. If this were so then it might mean that the psychopath is irrational to such a degree as to be outside the sphere of moral agency altogether. However, let us assume that this is not the case. It is certainly seems that even if the psychopath is incapacitated in this way that there may be other people who are also reckless and imprudent who are not psychopaths. There is nothing particularly about these people that excludes them from moral agency, so it seems probable that the psychopath is sufficiently rational as to be responsible in the sense required by answerability. What is distinct about the psychopath then, is that they are usually capable of acting on the basis of judgments about reasons (as answerability demands) such that their attitudes are judgment sensitive. However, moral reasons can never actually be reasons for them.

Because psychopaths act on the basis of judgments about reasons, it seems to be the case that they act maliciously, with intent. Their actions can therefore be said to display the sort of ill will that is indicative of their bad character. This leaves them open to moral appraisal. On an answerability account this is sufficient for blame proper. However, as I have shown, on an attributability/accountability view it is insufficient to warrant blame proper. Having bad character leaves one open to aretaic appraisal, but accountability is ordinarily seen to require that the agent is capable of recognising moral requirements as reasons. For instance, Watson has it that holding someone responsible in an accountability sense is to hold them to moral norms (which he takes to be norms that recognise the standing of other individuals). The psychopath cannot understand the authority of these norms and as such cannot be a moral agent in the necessary way, so it is inappropriate to blame them.

What would make it fair and appropriate to blame the psychopath only on the basis of their rational judgments? The psychopath’s incapacities mean that he cannot judge that a moral consideration is a reason for action. Scanlon’s response is that, though a psychopath cannot make the judgment that a moral requirement is a reason for action, he can make the judgment that a moral requirement is not a reason against so acting. Scanlon writes:

---

121 Talbert, 517.
123 Watson, The Trouble with Psychopaths, 1, 8.
..a rational creature who fails to see the force of moral reasons – who fails for example, to see any reason for being concerned with moral requirements at all or with the justifiability of its actions to others – can nonetheless understand that a given action will injure others and can judge that this constitutes no reason against so acting.  

Scanlon’s contention is that the psychopath, when considering which course of action to take, has the capability to understand the consequences of their action. When they are deciding upon a course of action to take, a judgment that these consequences are not a reason for the psychopath is still a judgment, and one that leads to an attitude or action for which the psychopath could be blamed.

There are two issues that I think arise from this claim: the first is whether the psychopath is, in fact, capable of the sort of judgment that Scanlon supposes; the second is whether such a judgment does indeed warrant blame proper, or whether more is required as usually entailed by accountability. So the first is a challenge to the claim that psychopaths, specifically, are blameworthy; the second questions more generally whether answerability succeeds as a comprehensive account of moral responsibility. I begin by examining the first claim – which is expanded upon substantially by Talbert – and then seek to explore the second.

With regard to the initial claim, Talbert argues that one is justified in blaming the psychopath because it is clear that the psychopath is capable generally of making judgments as to what are and are not reasons. We know this, he claims, because we can tell a counterfactual story in which the psychopath does not perform the action in question. Though the psychopath cannot judge that he has moral reasons, he can judge that he has prudential reasons to avoid so acting. This, importantly, also gives the psychopath the possibility of avoiding the sort of unwelcome response that constitutes blame. From this, I take it that Talbert thinks that the psychopath’s understanding of what reasons are, coupled with the idea that they have a choice about which considerations are or are not reasons means that the psychopath’s failure to respond to moral considerations is indeed the result of a judgment on their part. That is, they are truly making a judgment that a moral consideration is not a reason and it is on the basis of this judgment that they are faulty.

---

124 Scanlon, What We Owe to Each Other, 288.
But is the psychopath really making a blameworthy judgment? In some cases, it might be that the psychopath does in fact consider the consequences of his actions and rejects them as reasons for him. Now it cannot be that the psychopath makes any judgment with regard to an agent’s moral standing or with regard to moral obligations (because he cannot understand the nature of moral standing or moral obligations). However, it could be that the psychopath considers that an action will hurt someone and discounts the idea that this would constitute a reason not to perform the action. Further, Talbert claims that the psychopath knows that these considerations count as moral reasons for other people and yet still chooses to reject them as reasons for himself. However, even if the psychopath can understand instrumentally that the outcome of an action is that someone will be hurt, for instance, they surely cannot understand why this consequence would have any normative force. Therefore when thinking about a moral consideration, the psychopath does not have a choice about whether it is a reason – he is necessarily going to reject it as a reason because (on this stipulative account) that is what psychopathy is. This reiterates my worry from the previous chapter that answerability is covertly contingent upon the agent having a choice as to which reasons to adopt. This is a condition of moral accountability and if this is so then the psychopath ought to be excluded.

It is important to see, however, that this objection is not only an objection on the grounds that it is unfair to blame someone when they did not have a choice. My claim is that it seems the psychopath is restricted in his choices in terms of the judgments he can make about moral reasons. He is restricted because when faced with judging ‘is x a reason not to do this?’ where x is some moral consideration, he is unable to answer in anything but the negative. He cannot answer anything but the negative because he cannot understand why moral considerations can have normative force. So say the question is ‘is the fact that this action will cause harm a reason not to do this?’ then the psychopath’s response will only ever be ‘no’ or ‘yes, but only in light of some prudential considerations (like that causing harm might cause trouble for the psychopath’. Now, you might respond, this is a choice – as in it is theoretically possible for the psychopath to avoid performing the bad action even if only for prudential reasons. This is true, but it is important to note that the answer to the question is based on other judgments about reasons. The correct response is something like ‘yes, because causing pain is bad’, but what I think the answerability theorist fails to note is that the

---

126 Talbert, 8.
judgment that causing pain is bad is still a judgment about what is or is not a reason. And on this question: ‘Is the fact that causing pain is bad a reason not to cause it?’ the psychopath can only ever answer in the negative. What I am querying is not only whether it is fair to judge someone on this basis, but whether it even constitutes a judgment. So, it seems that a psychopath is not blameworthy because it is doubtful whether they are responsible given they do not have the relevant choice.

Talbert’s response to the choice objection is that there may be instances of ordinarily blameworthy agents in which the fact that the agent could have behaved other than she did does not explain why her action renders her an appropriate recipient of blame. He considers an agent who acts on a self-serving consideration just as it occurs to her without contemplation. He suggests that even if it is possible that the agent would have acted differently upon reflection, it is unclear that this consideration counts as a condition of her blameworthiness. This is because it is unclear that the agent’s action ‘involves a choice against a moral value’. But the agent in Talbert’s case does make a choice against a moral value, it is just not the same moral value that Talbert assumes. The agent’s fault in this case, it seems, is that they act unthinkingly when moral values are at stake. She chooses not to deliberate about the various outcomes when she ought to have deliberated. And this in and of itself is blameworthy. Indeed, you can imagine such an agent replying when blamed ‘I’m sorry – I just did not think!’ This lack of foresight is potentially an excuse that renders the agent guilty only of the lesser crime of acting unthinkingly, rather than the greater crime of acting maliciously. For the psychopath meanwhile, this kind of excuse is simply unavailable because no amount of thinking could have brought him to the right conclusion. So there is a substantive way in which choice affects the appropriateness of blame in Talbert’s example. This would all uphold that a psychopath is not blameworthy because they do not have a choice.

A different tactic employed by both Scanlon and Talbert is that we criticize other non-psychopathic agents on rational grounds even if they are incapable of understanding the requirements up to which they are held. This tactic points to interpretation I discussed in the previous chapter where an agent is just prima facie blameworthy if they hold bad judgment sensitive attitudes because they are rationally faulty. As I noted, Scanlon puts it that he ‘is

127 Talbert, 16.
open to criticism for his stupid chess moves even if [he] is unable to see that they are stupid’. \(^{128}\) Gary Watson criticizes this point on the grounds that a rational failure is unlike a moral failure in that a rational failure does not warrant blame, and I think this reply is basically correct.\(^{129}\) I have assumed that blaming does constitute some sort of rational criticism. However one can both agree to this and still agree with Watson that rational criticism on its own does not automatically entail blame. It might be that behaving in a rationally faulty way is a necessary condition for moral blameworthiness but that does not entail that it is a sufficient condition.

However, I think that a more damning response is Scanlon’s example given that it is just not particularly plausible. If I criticize someone for their poor chess moves then I assume that it is at least possible that they could understand the standard to which they are being held (even if they do not at the time). For instance, say I choose to play chess against a friend who has a brain injury which has substantially lowered their IQ such that they now struggle to grasp complex strategic concepts. It might be infelicitous to play chess against them at all, but it would certainly be infelicitous to criticize them for their faulty chess moves if I truly believed that they were incapable of ever understanding these standards. I take it that when one is criticized on rational grounds, the assumption is that the criticised party could understand the grounds on which they are being criticised even if the possibility of them ever actually understanding it is very small. That is, even if for instance it would require them learning much more about chess than they could ever really be bothered to do. But the psychopath is not just lazy in regard to understanding moral requirements; his understanding is impaired. One can agree to the idea that blame is a form of rational criticism but deny that it is ever acceptable to rationally criticise someone where they are completely incapable of understanding that the standard against which they are being judged. As such I take it that it is infelicitous to criticize psychopath in the way Scanlon supposes. So, though I agree with Scanlon that blame is a form of rational criticism, the psychopath should not criticised because there was no way of him acting in the rationally correct way.

As I have put it, it appears infelicitous to criticise someone on grounds that they are incapable of understanding. This sort of concern brings out a more serious one for answerability which is that Scanlon and his followers think that the fairly weak capacity of reasons-responsiveness

\(^{128}\) Scanlon, ‘Reasons, Responsibility, and Reliance’, 151.

\(^{129}\) Watson, *The Trouble with Psychopaths 1*, 12.
required by answerability is enough to warrant blame proper. The moral criticism answerability affords, according to Scanlon, ‘supports demands for acknowledgment [of wrongdoing] and for apology, or for justification or explanation’. But even if the psychopath is capable of making rational judgments in the way answerability requires, is this sufficient for this sort of response? Or are the conditions required by accountability necessary for the psychopath to be blameworthy?

On Scanlon’s view, blame is a demand that the agent answer for what they have done (with apology, justification etc) and two things underwrite this demand. One, as I have been discussing, is the capacity of the blameworthy agent to have judgment sensitive attitudes, which entails that they have a basic capacity for rational judgment (though not the specific capacity of responsiveness to moral reasons). This entails that they can answer for what they have done. The second are interpersonal conditions which are in effect that the agent does something morally wrong which impairs the relation between blamer and blamed. I take it that Scanlon takes something of a Strawsonian view in this regard where morality is contingent upon demands we make of each other in virtue of our participation in interpersonal relationships. This entails that the agent has done something blameworthy and that the blamer is in a position to blame. But are these the only conditions that ought to be required in order for the blamed to answer a demand from the blamer to account for their actions?

I would suggest not. For instance, in the case of the psychopath, though he has a basic capacity for rational judgment it seems that this capacity is insufficient to underwrite the demands being made of him. One difficulty is that to make this demand of him would be pointless: what kind of a response could the psychopath make to these demands? As Gary Watson notes ‘nothing they could do could intelligibly be construed as an apology or acknowledgement’ and this is what is demanded by blame (even on an answerability account). The reason that psychopaths cannot respond is that there is nothing that could make them see that moral requirements are reasons for them, so their reply is never going to be adequate. They are never going to be able to give an excuse, or a justification, or an

---

130 Scanlon, *What We Owe to Each Other*, 272.
131 Scanlon, 272.
apology – so if this is the purpose of blame then one could be forgiven for querying whether such a response is appropriate.

The difficulty, I think, comes at root from an imbalance in the answerability account. On one side the blamed party need only be capable of acting in accordance with rational judgments at all, while in order to blame the blamer must be capable of making judgments that the wrongdoer has behaved immorally and that it has damaged their relationship. However, the latter set of conditions are much more stringent, more demanding, than those conditions of the blamed. Yet, if we think that moral responsibility is supposed to be about shared moral standards that agents are expected to meet then the capacities of the blamed have to at least match those of the blamer in order for blame to be justified. So the psychopath needs also to be capable of judging what behaviour is immoral and whether their relationship is damaged in order to fall within the realms of moral agency but he does not and cannot do that. I take it that what is missing from answerability is the idea that all moral agents are subject to reciprocal moral norms and that everyone who is a moral agent should be capable of ‘recognizing and responding to the validity of these norms’ 

133 I use ‘norms’ here to quote Watson, but I take it that, broadly speaking, this matches the outline of moral responsibility first construed in which there is some demand to meet and uphold a moral standard. One does not have to agree about the sort of things that make up a moral standard to agree that it is only fair and appropriate to blame someone if they can understand that standard.

Hence, I conclude that psychopaths are not accountability blameworthy. I take it that they are not accountability blameworthy because they cannot understand the demand to which they are held blameworthy. As such they are not faulty for failing to meet the demand because there is no possible way in which the demand could have been met. Holding them to proper accountability blame is therefore both inappropriate: because they could never respond in the right way to the demand that is made of them; and unfair: because they have no choice and therefore no fair opportunity to avoid blame. They cannot be responsible for their rational faults because they cannot help it. This reiterates the conditions I put forward back at the start of Chapter 3.

133 Watson, 10.
4.2 Should psychopaths be attributability blameworthy?

I think it can reasonably be held that the psychopath remains blameworthy in the attributability sense. As I noted at the beginning of this discussion, psychopaths do act for reasons and these reasons are often bad, malicious or cruel. Indeed, part of what makes psychopaths appear so worrying is that they intend to hurt others. Even if they cannot understand the normative force of moral requirements, this intention frequently puts them against us such that one might attribute ascriptions of badness to them. As Gary Watson notes, psychopaths intend to subordinate the agency of others to their own.134 These intentions are indicative of the relevant sort of control required for an action to be attributable to an agent. Indeed the actions of a psychopath are not compulsive, they often require a high degree of manipulation and planning135. Moreover they are morally significant in their maliciousness, so one can justifiably attribute a claim of badness to the psychopath. It is bad that they do these things and it is right that this sort of appraisal is given.

As I noted in Chapter 3 the agent who is appraised aretaically is subject to some morally relevant ascription. Often it seems this appraisal is third-personal – ‘She is very naughty’, ‘That was a terrible thing to do’, indeed even ‘Gary is a psychopath’ might be seen as a sort of aretaic appraisal where it is used in the more common every day sense rather than in a clinical assessment. Such appraisal might be addressed to the recipient: I might say ‘Sarah you are so selfish’. But the difference between this and holding someone morally accountable is that nothing is required of the wrongdoer. I can tell the psychopath that they are bad, malicious, a terrible human being, but I cannot expect any answer from them. I cannot and ought not expect them to change their ways as a result of such appraisal. Furthermore even though one might feel horrified at a psychopath’s action, ‘resentment and indignation proper’ are unwarranted.136 I take it that one might feel badly towards the psychopath, but upon the understanding that they are a psychopath one might expect the resentment to be diminished, or at least there will an addition be an understanding that the sentiment is not warranted.

134 Watson, 14.
The psychopath is not accountability blameworthy and hence no longer a problem for reasons internalism. However, what of the other characters who are problematic for reasons internalism? The mob gangster who has some moral reasons but not others, or the egoist who has some reason to behave morally but not more than they find that they have reason to behave in their own best interest? These are characters who it seems are not \textit{totally} incapable of understanding and being motivated by moral considerations – that is they have some moral reasons, but either they do not have the right ones or they have them in insufficient amounts. In order for these characters to be unproblematic to the internalist then they too need to be excluded from the sphere of moral accountability. In the next section I consider whether it is right that this is the case, and whether it is sufficient to only blame them in an attributability sense.

4.3 The Mob Gangster and The Egoist

So far I have tried to establish that it is sufficient to hold the psychopath morally responsible only in an attributability sense rather than an accountability sense. This means that a bad psychopath deserves only aretaic appraisal rather than blame proper. The psychopath is excluded from the moral sphere because he cannot and could never meet the demands of morality such that blaming him would be appropriate. It is unfair to blame him because he did not have a choice and he cannot be rationally faulty because he is not responsible for this fault. The psychopath does not meet the relevant conditions, the question is whether we agree that the mob gangster and the egoist also do not.

There is a critical way in which the mob gangster and the egoist differ from the psychopath. The psychopath is excluded from the moral sphere because he is completely incapable of understanding moral considerations and is as such unmotivated by them. This is not the case with either the mob gangster or the egoist in my example. It is not the case that the mob gangster and the egoist are fundamentally incapable of understanding moral considerations because they are in some cases motivated by them. In the case of the mob gangster they possess some moral reasons but not others. That is they can understand why they have moral reasons to visit their sick mother in hospital for instance, but that same understanding does not motivate them to avoid shooting police officers in the face. In the case of the egoist they clearly understand moral considerations because they are to some degree motivated by them.
Where they fail is in being sufficiently motivated such that they have most reason to behave morally. These characters do not lack the depth of understanding that renders a psychopath blameless. We take it that they are capable of this understanding but they just happen to be unmoved, or insufficiently moved.

This assessment is borne out by the view that we consider psychopaths do be deeply psychologically abnormal, and indeed psychopathy is of course an incurable mental illness. But a failure to be motivated by moral considerations even when you can understand them does not seem like a psychological abnormality – indeed it seems almost boringly commonplace. I, for instance, would like to think of myself as someone who cares about the environment and believes that we have some sort of moral duty to protect it. Yet, I frequently buy water bottles and disposable cups of coffee from the university cafeteria. I understand that it would be better for the environment if I did not and that I might even have a moral reason not to do so. But I have a much stronger to desire to drink coffee and I have no desire to lug around water bottles or reusable coffee cups, and I still think this even after much rational consideration. I am in this way like the egoist because, on internalism, I have more reason to continue to drink from disposable cups even though they are bad for the environment and less reason to behave morally and buy a keep-cup.

Now, one criticism might be to argue that my discussion around psychopaths was inaccurate. You could hold that the psychopath does not fail to understand moral considerations, rather he understands them and is just not at all motivated by them – as such he is not distinguished from the mob gangster and the egoist by his lack of understanding. But even if this were true, the psychopath remains fundamentally unmotivated by moral considerations in a way that is not true of the mob gangster or the egoist who are both sometimes and to some degree motivated. There is no way that a demand to behave morally could ever be met by the psychopath, but there is a way in which the demand to behave morally could sometimes be met by the mob gangster or egoist – so, even if only in this way, they are constitutionally different.

I would suggest then that the mob gangster and the egoist are not unreachable by the language of moral address, they are incorrigible. This would entail that in the case of the mob gangster, for instance, though he might truly have had no reason to shoot the police officer in the head at the time he shot him, could come to have a reason not shoot police officers in the
head in the future. This sense of possibility can range from entirely probable to very, very small such that it demands a long and unlikely series of events to come to pass. It could be that the gangster’s son is killed as a result of gang related violence and suddenly he sees that he most wants to live a good and moral life. Or it could be that the gangster is arrested and, out of a disposition to avoid further punishment, goes on to have years of rehabilitation. Upon his release, while still not really wanting to behave for moral reasons, a chance encounter with a Buddhist monk leads him to go on a path of self-discovery subsequent to which he finds that he does most want to behave morally. It may be scenario required for the gangster to see that he has reasons never comes to pass, but the point is that there is always some possibility, however small, that the mob gangster could act for moral reasons – whereas with the psychopath there is none.

So these cases are different because it might be appropriate to blame them in a forward-looking sense in that it is theoretically possible that they could change their behaviour. However, as I argued forcefully during the discussion around the proleptic mechanism, it is not enough to argue that an agent could come to have this reason in the future. In order to be blameworthy the agent has to have the possibility of behaving for the right reason at the time that they commit the wrongdoing. Though we want to blame the mob gangster and the egoist, we cannot because they either do not have this possibility, or have it but have more reason to do something else so they cannot be criticised for their failure. The problem is that I think we can agree that there are people like the psychopath who are constitutionally incapable of behaving morally. What we doubt is that people who sometimes behave for moral reasons could actually not have moral reasons in some circumstances. Crucially, I think this doubt comes about not because they fail to be motivated by moral considerations – as I have argued this is depressingly commonplace – but because we just assume that such agents ought to be capable of seeing moral reasons for what they are: aka normative reasons.

A strong argument against my point here, however is that we do think that some non-psychopathic agents ought to avoid being blamed precisely because we can see how they are incapable of being motivated by moral considerations – even if they can in principle understand moral considerations. This sort of argument can be made for the kind of bad character akin to Watson’s famous Robert Harris example.\footnote{Watson, ‘Responsibility and the Limits of Evil’.} On one hand, Harris was a
double-murderer who appeared of extremely bad character. Along with his brother, he commandeered the car of two teenage boys in order to rob a bank. Before the robbery he shot the two boys at point blank range and proceeded to eat their hamburgers. Watson writes that Harris was hated by the other inmates in his prison for his total immorality and was frequently involved in violent altercations with them. They said that they would celebrate when he was executed because he was such a nasty piece of work. On the other hand, Harris was an unloved and abused child. Unwanted by his mother, he was born prematurely when his extremely violent father kicked her in the stomach and so induced labour. His sister reported that their mother seemed to blame him for this, and subsequently showed Harris no love at all for the remainder of his life. He would tug at her for attention and be kicked away. His father continued to be violent towards his mother, Harris and his siblings. He was bullied in school and had no friends. When he offended as a teenager he was sent to prison where he was raped. So, although Harris is of bad character, we can see how that bad character has come about. We can also see how it is that he could fail to be motivated by moral considerations as reasons having never experienced any praise or love throughout his formative years.

According to my line of discussion, I take it that Harris falls somewhere in the realm of the egoist or the mob gangster. Harris is not constitutionally incapable of seeing the force of moral considerations like the psychopath. Indeed, upon his execution in the gas chamber, he seemed to apologise to the father of one of his victims. This suggests a capacity for the kind of remorse of which the psychopath is constitutionally incapable. Indeed, it seems very plausible that if Harris had been the recipient of a more ordinary sort of upbringing that he would not have committed such terrible crimes but may have been capable of living a relatively ordinary sort of moral life. This, again, is not true of the psychopath. Harris appears to have the capacity to act for moral reasons. Yet, it also appears that Harris either is not motivated by moral considerations, or that he is far more motivated by his own prudential considerations. In an internalist sense, Harris did not have most reason to behave morally. And, if we agree that accountability blame requires that an agent acts upon what they had most reason to do, then Harris is not accountability blameworthy.

While this was not the intention of Watson’s example, I take it that Harris and characters like him are useful for examining the complexities of the reasons internalism/externalism debate. This is because, on one hand, it seems as though Harris is clearly blameworthy. He has all the
capacities required for rational agency. There is no doubt that he committed the crimes of which he is accused. We expect him to be able to give account of himself for the crimes he committed. Yet, on the other hand, it is wholly unsurprising that Harris commits the crime he does. We might think at face value that he has reasons not to do what he does, but at the same time we can understand why these reasons are not reasons for him. That is, we can see how it is that moral considerations completely fail to motivate Harris and other agents of similar backgrounds. Harris can be seen as illustrative of how non-psychopathic characters, characters like the mob-gangster, or the egoist, could be outside the sphere of accountability. It is certainly not a given that the capacity to respond to moral requirements entails that one is automatically accountability blameworthy.

As it goes, I think that Harris lies within the realm of the morally accountable, however there is no need to defend that view here. This is because, even assuming that Robert Harris and his ilk are outside the sphere of accountability blameworthiness, for every character of this type there is another sort of mob-gangster or egoist who elicits no such sympathy and who can be granted no excuse for failing to respond to moral requirements. It is clearly true that not every person who acts badly is either a psychopath, or could possibly exempted for reasons of prior maltreatment. We can all think of cases where someone is a ‘bad apple’ irrespective of their upbringing. They may have loving loyal parents who did all they could and yet still they grow up to be hardened criminals, manipulators or persons with no moral backbone. Moreover, what is most problematic for internalism is that these people need neither be very many nor very bad. We might presume, for instance, that many bad adults who have been bad all their lives in spite of their good upbringing are indeed psychopathic (given the prevalence of psychopathy in the general population). Or indeed, we might presume that many of the people who commit the worst crimes are either psychologically abnormal or were grossly maltreated such that moral reasons are not, truly, reasons for them. But, in order to avoid the problem from blame, reasons internalism requires that there is no one who fails to have most reason to behave morally who is not exempted for psychopathy or some other plausible exempting factor. I see this as patently false.

And in these cases, what we need to establish is whether blame on an attributability only basis is all that is warranted – because on the internalist view that must necessarily be the case. However, I think it is self-evident that we do not believe that attributability in these cases is all that is warranted. Talbert puts forward the following example:
Imagine the way prisoners in a concentration camp surely blamed and condemned their murderers. Certainly, the prisoners’ condemnation expressed the conviction that their treatment was impermissible and the demand for recognition of their moral standing as human beings. However I do not think that these demands and claims lost their point when they failed to move hardened concentration camp executioners. 138

Now Talbert puts forward this argument as an argument against the inclusion of psychopaths, but his fault in this case is not that we should not blame people who are particularly incorrigible, but that these incorrigible Nazi murderers are not the same as psychopaths. Now one can grant that some of these people are neither like the mob gangster or the egoist in that they really did have most reason not to perform the action and only did it through coercion or because they were following orders. But it is hard to believe that in the course of history that none of these agents was like the mob gangster or the egoist – if they were not then it is hard to believe that such regimes could ever have succeeded. And assuming this is the case, then it is clear that we do not think that the fact that the executioners might have had no internal reason to behave otherwise means that they are not blameworthy. To the contrary, I propose that we assume that such people clearly have reasons not to do what they are doing, and we assume that the prisoners are legitimate in their demands because the executioners have reasons.

It certainly does not seem sufficient to reply that the prisoners’ pleas are unjustified, that their blame is unjustified, but that the guards are at least responsible in an attributability sense so at least the prisoners can say that the agents are bad. This does not seem to suffice at all. This would entail that the guards are – as Angela Smith puts it – ‘passive victims’ of their moral psychology. 139 But we just plainly do not think this is true of such people. To regard these people as bad in an attributability way is like saying that they are bad but they cannot help it because they cannot help their psychology. But intuitively we do not think this is the case. Rather, we assume that, irrespective of their motivational set, that these agents have control over the way that they act and that they ought to act for moral reasons even if they do not want to.

139 Smith, ‘Control, Responsibility, and Moral Assessment’, 390.
Here, it seems, we come to the crux of the blame problem for internalism. Though the exclusion of psychopaths on the basis of moral accountability seems legitimate because they are fundamentally incapable of acting on the basis of moral reasons and therefore answering the demands made by moral responsibility, the same claim cannot be made about mob gangsters or egoists. In these cases it seems blame proper is warranted and aretaic appraisal would be wholly inadequate. But to accept this, I think, requires that one deny the truth of reasons internalism and accept the truth of reasons externalism. This point is made stronger when one considers that externalism is capable of justifying why the psychopath would be excluded from the sphere of moral accountability but not the mob gangster or the egoist. The psychopath can be excluded because they are fundamentally incapacitated such that they cannot recognise that moral reasons are in fact reasons. Whereas for the mob gangster or the egoist, they are morally responsible because they could recognise moral considerations as reasons but just choose not to do so. This point needs significant justification which I shall not give here, however what should be noted is that externalism is compatible with moral accountability in regard to who should or should not be included, whereas I do not think internalism ever could be.

4.4 Concluding Remarks

I argued in the first half of this thesis that it seemed there was no way of avoiding the unfortunate conclusion that some blameworthy characters are not blameworthy if we adopt both reasons internalism and a conception of blameworthiness which is contingent upon the agent having acted against what they have most reason to do. In the second half I have argued that this would not be so problematic if those who were excluded were truly morally blind – psychopaths as I have termed them here. However, an issue that becomes evident in my argument against Willi’s internalism is that the pool of problem candidates appears to be much greater than might first be thought. Some characters – mob gangsters in my terminology – appear to have some internal moral reasons but not others. This is problematic if we think that the agent must have had a choice over their actions such that they could have performed for the right reasons. Other agents – egoists – have some moral reason but insufficient reason on my account to behave morally. If blame is a form of rational criticism such that an agent can only be blameworthy if they fail to act on the basis of their best reason, as I have argued here, then egoists too are a problem.
I find that the weakest part of my argument is this contention that an agent has most reason to behave morally and the assumption that on an internalist argument what an agent has most reason to do is contingent upon what they the greatest desire to do. I take it that neither of these are particularly outlandish claims, however if one denies my contention about most reason or if more could be written in to the capacity of rational deliberation such that what an agent had most reason to do was in some way linked to morality then some of the arguments I have put forward here would be mitigated. However, I want to stress that the blame problem would clearly still be a problem given the existence of people like the mob gangster, and I hope I have drawn attention to the depth and complexity of the problem here. The exclusion of the mob gangster is contingent upon the issue of choice and, as I have argued, blame proper does not seem to be warranted where choice is missing.

It may be further that there are attributability accounts that I have missed that do a better job of arguing that blame proper is appropriate on attributability conditions. However, I take it that answerability is the most prominent example of this sort and it seems to either rely tacitly on the sort of choice that is problematic for internalism, or just presupposes that reasons are external. Overwhelmingly, what I hope has come out is that blaming is just that much simpler when one is an externalist about reasons.
Bibliography


http://dx.doi.org/10.1017/CBO9781316136478.


http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198715672.001.000
1/acprof-9780198715672.

Skorupski, John. ‘Internal Reasons and the Scope of Blame’. In *Bernard Williams*, edited by
Alan Thomas. Cambridge: Cambridge University Press, 2007.[Kindle]


Smith, Angela M. ‘Attributability, Answerability, and Accountability: In Defense of a


149–70. https://doi.org/10.1353/cjp.0.0070.

Strawson, P. F. ‘Freedom and Resentment’. In *Freedom and Resentment and Other Essays*,

Talbert, Matthew. ‘Blame and Responsiveness to Moral Reasons: Are Psychopaths

Tognazzini, Neal, and D. Justin Coates. ‘Blame’. In *The Stanford Encyclopedia of
Philosophy*, edited by Edward N. Zalta, Spring 2016. Metaphysics Research Lab,


Press, 1996.

Watson, Gary. ‘Responsibility and the Limits of Evil’. In *Agency and Answerability: Selected

http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199753673.001.000


Williams, Bernard. ‘Internal and External Reasons’. In *Moral Luck*, 101–113. Cambridge:

———. ‘Internal Reasons and the Obscurity of Blame.’ In *Making Sense of Humanity: And