The Normativity of Obligations

Trust, Blame, and Interpersonal Agency

A dissertation submitted for the degree of Doctor of Philosophy

Faculty of Philosophy

University of Oxford

Alexander Odgaard Heape

Mansfield College

Trinity 2018

Abstract

The Normativity of Obligations: Trust, Blame, and Interpersonal Agency
A dissertation submitted for the degree of Doctor of Philosophy
Faculty of Philosophy
74401 words
Alexander Odgaard Heape
Mansfield College
Trinity 2018

This dissertation defends Deontic Pragmatism – the view that the normativity of obligations is grounded in reasons to engage in interpersonal agency. I defend an account of obligations as correctness standards grounded in specific kinds of reasons. On this view, all obligations are directed from one person to another because these reasons are facts about a specific bipolar relation between people. To be a moral person is, in part, to comply with one's obligations for such reasons, by caring about other people and how they rely on oneself. Such caring attitudes are constitutive of the kinds of relationships people have reasons to have with one another. Indeed, the trusting attitudes such relationships also constitutively involve presuppose that others hold such caring attitudes towards us. In turn, the blaming responses to which trust disposes us presuppose that people fail to respond to such reasons, by failing to hold such caring attitudes.

I argue that this nexus between obligations, trust, blame, and relationships is played out across the entire moral domain. I then defend an account of the reasons people have to hold such attitudes. On this view, reasons that ground correctness standards, including obligations, are shared by people in virtue of their engagement in activities. Specifically, it is correct for people to have moral relationships, because they have reasons to intend to engage in interpersonal activities that involve mutual dependence relations. And they have those reasons, I argue, whatever other activities they are engaged in. Since such dependence is presupposed by the attitude of reliance involved in trust, the reasons that ultimately ground our obligations are grounded in a relation that is itself presupposed by the attitude of trust.

Contents

Introduction

When asked why they are doing this or that, most people find it fairly easy to come up with an answer. They put on the kettle because they are making a cup of coffee, they stay inside because it is raining, and so on. We all know what we are talking about when we give these explanations; we know what 'rain' and 'cups of coffee' are, and why facts about rain and coffee are relevant to answering such questions. Questions about morality are often importantly different. If asked why we do not lie, steal, or injure people, most of us would not hesitate to say that it is because such acts are wrong. But most of us also find it unusually difficult to articulate what we are trying to say when we say that some action is 'wrong'.

This answer is difficult to articulate in part because it can be unclear what consequences wrongful actions have for those who commit them. At this level too, morality contrasts with ordinary practical reasoning. There is a sanction for going outside in the rain: you will get wet if you do. There is also one for making coffee with lukewarm water: it has a bad taste. Whether one should stay indoors or put on the kettle are often clear questions in part because the consequences to oneself of failing to do them are clear. Of course, wrongful actions often do have consequences too. Many wrongful acts are covered by law, and there is a system in place for ensuring that violations of the law are met with tangible sanctions. But it is not always illegal to act wrongly, and the law is not always enforced. So why obey the moral law if one can avoid the sanctions of the juridical one?

1

In her essay 'Modern Moral Philosophy', G.E.M. Anscombe argued that the difficulty of answering moral questions when posed in such terms was precisely the reason why moral philosophy ought to rid itself of the concept of obligation. This term and its cognates such as 'ought' and 'wrong' were, Anscombe argued, relics from a bygone era of Divine Command theories that conceived of morality as law and immoral behaviour as punishable offence, a conception she held to be unsustainable in the absence of a divine lawgiver. Instead, Anscombe proposed that philosophers study the psychology of agency and virtue, an area of research which was, as she put it, 'conspicuously lacking'.[1]

These days, the lack is less conspicuous, and we now have substantive competing theories about the nature of attitudes like desire, intention, and blame that make up our moral psychology. The development of these theories is, I believe, important progress. And I argue in this thesis that a proper understanding of our moral psychology can itself support an explanation of what obligations are, why they are normative, and why they have the particular binding force that set them apart from other kinds of normative considerations. The very type of moral philosophy Anscombe encouraged thereby illustrates why her central contention was wrongheaded. Rather than obscuring the questions central to moral theorising, the concept of obligation can help us answer them.

As I suggested above, a particularly difficult question is why it matters at all that people do what they are obligated to do. When we ask why people should not lie, for instance, there are many possible answers. We might say that lying is likely to be worse for the liar. If one were disposed to lying, it is is likely that one would be found out at some point, with dire consequences; better not to countenance lying,

---

[1] Anscombe (1958, p. 1).

then. But there is a familiar concern with this type of explanation. Most of us do not think in these terms when we conclude that lying would be wrong, nor do we believe this of those whose moral halo shines brighter than our own. Alternatively, we might explain the obligation not to lie, as some moral theories do, by saying that it would be irrational to lie, or that refraining from lying will help bring about the greatest amount of pleasure over pain.

The concern, however, can be pressed against these explanations too. Most of us do not believe that morally decent people are concerned, exclusively or primarily, with rationality or with pleasure and pain. When pressed repeatedly, the concern arises that any such explanation is bound to be unsatisfactory. On the one hand, it is no good if a moral theory simply tells us that we should do what is right because it is right. That would not be much of an explanation. On the other hand, it is no good if it can only identify a reason to do what is right that would not motivate morally decent people, or that would seem irrelevant or inappropriate to them. That would make it difficult see what the answer given has to do with the question that was asked.

A theory of obligations that takes moral psychology as its starting point offers a hope of avoiding this kind of dilemma.[2] If a reason to do what one is obligated to can somehow be traced back to the psychology that characterises moral people, it is less likely that such a reason should seem alienating to them. This emphasis on moral attitudes characterises a variety of recent approaches to moral theory. One point of focus concerns attitudes like guilt or remorse, through which a person holds him- or herself accountable for having violated an obligation. A second

---

[2] This is sometimes referred to as 'Prichard's Dilemma', but that label is somewhat misleading, since Prichard (1949) denied that moral theories were supposed to provide substantive explanations in the first place.

concerns attitudes like resentment and indignation, through which we hold other people accountable for their actions. And a third concerns the deliberative point of view, from which one asks how the fact that one is obligated to do something should guide one's reasoning. In different ways, these attitudes are said to involve thoughts that help indicate the normative grounds of obligations.

I believe there is an equally significant type of attitude that involves a distinctive way of thinking about obligations, one that has largely been ignored so far. Just as we can think about how our own obligations constrain our interactions with other people, so we can think about the role that other people's obligations play in their deliberations about how to interact with us. And the attitude that is characteristically responsive to the role that other people's obligations play in their deliberative mode of reasoning is, I shall suggest, the attitude of trust. My aim in this thesis is to argue that a sustained focus on the normative content of this attitude opens up an argumentative route to the normative source of obligations that is both novel and appealing, one that takes their source to be human beings' capacity for a certain type of agency. I call this theory Deontic Pragmatism.

The type of agency that Deontic Pragmatism takes to be fundamental is thoroughly interpersonal. Ultimately, it explains the normativity of obligations by appealing to practices that human beings engage in together, that aspect of our practical lives that concerns what the Greeks called 'ethics'. To adopt 'Ethics' as the name of particular ethical theory would, however, be misleading, and 'Deontic Ethics' is hardly better. Thus, I settle on the name Deontic Pragmatism for lack of a better one, though I do not intend to suggest any particular affinity with the views propounded by so-called pragmatists in the American tradition of Peirce, James, and Dewey.

Deontic Pragmatism's central idea has been explored before. That the phenomenon of interpersonal trust reveals a normative source of obligation is perhaps the central topic of the Danish theologian K.E. Løgstrup's 1956 book 'The Ethical Demand'.[3] However, Løgstrup's own conception of this demand, the notion of trust he operates with, and the sense in which the normative source of that demand can be derived from that phenomenon, are all subject to intense exegetical debate. Moreover, attempts to situate his ideas within contemporary analytic moral philosophy are few and far between, and a project deserving of an entire thesis in its own right.[4] I have thus found it best to eschew such exegetical questions and leave it to others to determine the extent to which the view defended here is akin to Løgstrup's, and to explore instead what seems to me (rightly or wrongly) to be the central insight.

Deontic Pragmatism is fundamentally different, I believe, from any existing account of obligations and my aim here is to defend it as a particularly appealing one. I find it appealing partly because it seems to me to incorporate in one coherent package elements of Consequentialist, Contractualist, and Kantian views of morality that are appealing in their own right. So while this view is in an obvious sense a competitor to these alternatives, it would be a mistake to consider it a direct rival to any of them. Indeed, I believe that it illustrates the possibility of reconciling some of the central claims that motivate these different theories, and that this is a weighty consideration in its favour.

I will also argue that it allows us to formulate an independently plausible response to the dilemma discussed above. Deontic Pragmatism does this by illustrating how ethical theorising that proceeds from observations about our moral

---

[3] Løgstrup (1997).

[4] See Fink and Stern (2017); Andersen and Niekerk (2007) for some recent exceptions.

psychology need not take for granted any justification of the attitudes that make it up. On the contrary, it allows us to argue that these attitudes and our obligations have a single, independent normative source. At a different level, then, part of the motivation for exploring Deontic Pragmatism is to determine the extent to which we can attain such unity.

Such unity is attainable, I will argue, because the explanation Deontic Pragmatism offers of these normative phenomena appeals only to a single, fundamental normative property, the property of being a reason. By drawing systematic distinctions between different sets of reasons, it derives normative standards for both attitudes and actions from the same basic source: interpersonal activities. In this way, the case for Deontic Pragmatism is simultaneously a case for the idea that reasons are normatively fundamental, and part of the project here is to explore the extent to which such a 'reasons-first' approach to obligations can succeed.

I begin by exploring in chapter 1 the idea that facts about obligations are facts about reasons. I argue that the idea that obligations are a particular type of correctness standard, fixed by how certain kinds of reasons weigh against each other, captures much of what appears to be distinctive about obligations. Part of what makes them distinctive are their characteristic 'bipolar' features, the way in which they obligate one person *to* another. These bipolar normative features suggest that the kind of reason that settles this particular kind of correctness standard obtains in virtue of some relation between people. By discussing the relation between blame and trust and the way that both presuppose the existence of obligations, I suggest that this is the relation involved in trust itself: the relation of reliance.

In chapter 2 I defend in further detail this Reliance View of obligations, arguing that it gives us a plausible account of both promissory and special obligations.

This view is plausible, I shall argue, because it provides an independently plausible account of the normative nexus between three different types of attitude and the role they play in relationships. Relationships are constituted by attitudes of care and concern that modify the weight of people's reasons to do what the person cared about relies on them for. That difference in weight can in turn explain why reliance based reasons can ground obligations directed to some people rather than others. That people comply with these obligations out of care and concern is precisely what is presupposed by the attitudes of trust constitutive of these relationships. And this in turn explains why the failure to respond to the obligations should dispose people to blame; it is because such behaviour is indicative of the lack of care and concern that such relationships depend on.

In chapter 3 I argue that this account can be extended to give us a plausible account of all obligations. I argue for this by exploring a version of the Contractualist notion of a maximally extensive moral relationship, but I suggest that the most plausible version of this idea is too think of that relationship as constituted by the attitudes of respect and recognition traditionally emphasised by Kantian theories of morality. By building on a specific second-person interpretation of such attitudes, I argue that the model suggested in chapter 2 applies to this relationship as well because it involves default attitudes of trust and reliance towards everyone. On this view, obligations owed to everyone obtain in virtue of reliance relations that connect any two individual people within this maximally extensive relationship. So-called imperfect duties, on the other hand, obtain in virtue of reliance relations that connect subsets of individual people within it. An important source of motivation for this view, I shall argue, is its ability to explain variations between individual people's standing to blame third-parties.

In chapter 4 I ask how one might defend the idea that there are reasons for people to have the attitudes that constitute these various kinds of relationships and ground the obligations they involve. I shall argue that Deontic Pragmatism's commitment to the idea that reasons are normatively fundamental presents it with the so-called Wrong Kind of Reason problem. I then discuss how competing accounts of obligations might avoid this problem and argue that each of them leads to some fairly uncomfortable conclusions. If we cannot solve the Wrong Kind of Reason problem, the best option for Deontic Pragmatism is to abandon its commitment to the fundamentality of reasons and accept at least some of these conclusions. On other hand, if we can solve the Wrong Kind of Reason problem, we would not have to accept these conclusions, and this would give theories that accept the fundamentality of reasons an advantage over those that do not.

In chapter 5 I argue that there is a solution to this problem, one that provides independent support for Deontic Pragmatism. According to that solution, correctness standards are generally constituted by reasons that obtain in virtue of engagement in activities. Thus, reasons make it correct to hold some attitude just in case they are reasons shared by everyone engaged in some activity. This shows us how we might defend the claim that it is correct for everyone to hold the attitudes that constitute the moral relationship. By providing a non-normative analysis of the phenomenon of second-personal interaction, I argue that this is the generic attitude of intending to engage in joint activities with other people. On this view, it would be correct for everyone to have that attitude if they are engaged in some activity in virtue of which they most reason to have such an intention.

Finally, in chapter 6, I defend the claim that everyone *is* engaged in some such activity. I begin by defending the normative framework set out in chapter 5 in fur-

ther detail. I first address the objection that reasons cannot provide normative support for attitudes at all. I then address the objection that agency itself cannot be a source of normativity. I argue that neither of these objections succeeds. I therefore move on to show how a defence of the claim above is possible on Deontic Pragmatism's premisses. I shall suggest that such a defence is possible, if we adopt an account of the weight of reasons that sits naturally with the idea that reasons are explained in virtue of activities.

On this view, the weight of a reason is proportional to the complexity of the activity it promotes. Because joint activities are complex in the relevant sense, there is more reason to engage in them than not, for anyone who is engaged in any activity. Put simply, the reasonable life is one of interdependent agency. The type of relation that grounds reasons for everyone to be responsive to the reliance of others is thus precisely the relation that the attitude of reliance itself presupposes, the dependence of one person's agency on another's. And since reliance is constitutive of trust, the attitude of trust thereby brings out the normative grounds of obligations.

As I shall suggest, Deontic Pragmatism thereby offers an explanation of the normativity of obligations that is highly attractive. It is simple, general, and economical, and it treats obligations merely as a special case of the normative domain, special because it is the part of this domain that is explained by the particular interconnectedness of human agency. If this view is correct, Anscombe was right that moral psychology was central to moral theory. She was right about this, however, because our moral psychology is precisely where we can find the resources to explain what is both special and important about the normativity of obligations.

Chapter 1

Obligations, Blame, and Trust

1.   Deontic Pragmatism

This dissertation is a defence of the view I call Deontic Pragmatism. It is a deontic view because it holds that there are obligations. It is a normative view because it holds that obligations have normative properties. It is also an explanatory thesis, as it says that there is an explanation for why obligations have normative properties. Further, it is a fundamentality thesis. It holds what is sometimes called the 'primacy of reasons thesis' of 'reasons-first thesis'; that all normative properties obtain, when they do, in virtue of reasons. It is also a parity thesis. It says that the normative properties of all obligations are ultimately explained in the same way. Finally, it is a pragmatist view because, it holds that this explanation is fundamentally about practices or agency. Hence its name: Deontic Pragmatism.

   This chapter sets out the basic tenets of Deontic Pragmatism (or Pragmatism, for short) and clarifies how they fit together. They fit together, I shall suggest, as a thesis about a normative nexus between obligations, blame, and trust; one that explains the particular type of 'directedness' that many obligations seem to have, and why they should ultimately be explained by practices. I begin by setting out

Pragmatism's basic structural features. Most important here is the idea that obligations are normative because reasons are, so I will start by clarifying what 'because', 'obligation', and 'reason' mean here.

Many accept that normativity is fundamentally about reasons. However, since there are different ways of understanding that idea, let me clarify my own view. I believe that all claims about what is 'normative' or 'normativity' are ultimately about what there are reasons for, but I do not take this to be a matter of definition or conceptual truth. If we were to discover, for instance, that there was no non-trivial property picked out by the term 'reason', it would not follow from this that there were no normative properties. We might discover that our beliefs about reasons were false, and that sentences or propositions about normativity and normative phenomena were really about, say, rationality rather than reasons.

Similarly, while I believe we should accept that obligations are normatively 'binding' in the sense that we always have more reason to comply with them than not, this does not seem to me to be a conceptual truth about obligations. Rather, these claims about reasons and obligations are substantive claims in the sense that they could turn out to be false; that is why we must defend them, if we think they are true. Part of this endeavour, I believe, is to offer a plausible explanation of *why* they are true. And one of the central claims I shall defend is that the *most* plausible explanation of why obligations are binding, in the normative sense, is an explanation in terms of reasons.

Reasons, I shall assume, are facts. They are facts that count in favour of responses. A 'response' can either be an action or an attitude, and I shall often use 'response' as a generic term to cover both. When all the facts are in, we can ask what you have *most reason* for. This question will not always have a straightforward

answer. For instance, you might have one reason to eat a banana cake, another reason to eat a chocolate cake, and no other reasons. If these reasons have the same *weight*, then it will not be true of either cake that you have most reason to eat it. In this case, we should merely say that you have most reason to eat *some* cake. If, on the other hand, these reasons are not equally weighty, then you will have most reason to eat one cake rather than the other.

I shall also assume that a reason is always a reason for some particular person. The fact that a particular cake contains walnuts may be a reason for you not to eat it, without being a reason for me not to eat it. This would be the case, for instance, if you are allergic to walnuts, while I am not. These are claims about *objective reasons*. If you are allergic to walnuts, the fact that a cake contains walnuts will be an objective reason for you not to eat it, whatever you think about it. These are the reasons that Pragmatism holds to be normative. However, if you do not believe that this particular cake contains walnuts, then the fact that it does may not be a *subjective reason* for you not to eat it. Something is a subjective reason for you to eat a cake only if you believe it, and only if it would be a reason for you to eat the cake if true. Because your beliefs may be mistaken, facts or propositions do not provide you with reasons merely because you think they do. But your beliefs about reasons may nevertheless be important.

Suppose you believe falsely that some cake is delicious. When this belief explains why you eat it, your *motivating reason* is your subjective reason.[1] Had the world been as you believed it to be, this would have been a reason for you to eat the cake. As it was, it was not. But now suppose you also believe falsely that the cake

---

[1] There is a minor complication here. Your belief might 'explain' why you eat the cake in purely causal terms. For instance, if you suffer from a funny virus that causes you to eat whatever is front of you when you think about cakes, your belief that there is a cake would explain why you eat it, but not because its content is a motivating reason for you.

was free from walnuts. In this case, you would have most objective reason not to eat the cake. For you would have no objective reason to eat it, and a good objective reason not to eat it. Eating it would be a big mistake. If you had known or believed that the cake contained walnuts, you would have been *criticisable* for making this mistake. Given your subjective reasons, however, this might be false. If your subjective reasons favoured eating the cake, and if you were not in a position to know about your objective reason not to eat it, you might have an *excuse* for eating the cake.

When your objective reasons match your subjective reasons, then your objective reasons are also your *motivating* reasons, or the reasons you respond *for*.[2] For simplicity, I shall often assume such a match. On some views, however, a fact can be a reason for someone only if there can be such a match; that is, only if he could act *for* that reason.[3] For now, I shall reject this view on the basis of a familiar objection.[4]

Imagine that you love successful surprise parties, but hate unsuccessful ones. There is, in fact, a surprise party behind the door in front of you. So that is a reason for you to enter the door. But it could be a reason for which you entered the door only if it were a motivating reason. It could be a motivating reason only if you thought it was true. But, of course, if you thought there was a surprise party behind the door, it would no longer be a surprise. And then it would no longer be a reason for you to enter the door. So this view says that it cannot be a reason at all, even

---

[2] This assumes, of course, that you are not *akratic* i.e. that you fail to be motivated by reasons that you genuinely believe are there. Though important in their own right, I shall generally set aside questions about akrasia.

[3] This is also known as the 'response condition' on reasons. See Shah (2006); Kelly (2002); Parfit (2011a, p. 51).

[4] This case is first discussed in Schroeder (2007, p. 33). I emphasise 'for now' here because we will be discussing a slightly different version of the response condition in chapter 4.

before you think it is true. Since that is false, this 'response condition' on reasons is also false.[5]

A final point to note is that some facts are reasons only because other facts are. That there is a party may be a reason for you to go, only because doing so is a way of doing something that you have another reason to do, such as having fun. But not all reasons can be derivative in this way. If there are objective normative reasons, some facts must non-derivatively count in favour of responses. Though we shall be discussing derivative reasons along the way, it is these non-derivative reasons that Pragmatism is ultimately concerned with. Thus understood, it holds that there is an explanatory relation between our objective, non-derivative reasons and our obligations. But there are different views about the nature of this explanatory relation.

Normative reasons are often referred to us 'justifying' reasons, because they are the reasons which justify our thoughts and attitudes. And Pragmatism is the view that the normative properties of obligations are ultimately explained in terms of practices. This suggests a very natural place to start looking for normative reasons on this view: our justificatory practices. When we ask what reason people have not to commit murder, for example, a common explanation is that murder is *wrong*. Similarly, when asked why they are not willing to break a promise, many would be prepared to say that they consider themselves *obligated* to keep it. So it is natural to think that reasons not to do such things obtain in virtue of facts about wrongness and obligations. This is also a substantive philosophical view.[6] But Pragmatism asserts that the direction of explanation goes the other way; that obligations obtain

---

[5] As we shall see in chapter 4, this point is important because Pragmatism is actually harder to defend if we reject the response condition than if we accept it.

[6] See e.g. Heuer (2010); Darwall (2010); Rawls (1971, p. 113); Scanlon (2007).

in virtue of reasons. Those two claims are circular, however, unless we hold that obligations simply *are* reasons. And this view is open to a serious objection.[7]

The law, for instance, is a body of requirements that purports to constrain our conduct. Like obligations, statutory provisions specify things one must do, and other things one may not. And we often explain why there are reasons not to do things by saying that they are 'wrong' *de jure*, or 'illegal'. But suppose the law forbade saving drowning children. One could coherently question whether *that* would be a reason to let drowning children die, without betraying any misunderstanding of what a legal requirement is. So facts about these kinds of putative constraints are not *themselves* facts that count in favour of things. Like the law, the objection goes, obligations are not *really* normative in and of themselves.

One might respond to this by denying that obligations *are* reasons. Instead, one could think, facts about them are higher-order facts *about* reasons. If we think of reasons as being like weights on a scale, one might hold that, rather than being weights on the scale, obligations are facts about what the weights on the scale favour. This would explain why, when we say that something is obligatory, e.g. keeping a promise, we are usually not suggesting that this consideration is one reason among many; as if it were still an open question whether there was most reason to keep it. Rather, we are asserting something like a presumptively conclusive directive: Promises are to be kept, period.[8] But this view has other problems.

First, one can have most reason to do something without being obligated to do it. You might have two options that are identical in all respects, except that one of them will get you an extra piece of cake. Even if you have most reason to choose

---

[7] Crisp (2006, p. 13); Anscombe (1958).

[8] An idea familiar from Hare's treatment of moral language as essentially prescriptive. See Hare (1963, pp. 168–9).

the cake option, we would not normally say that you are *obligated* to choose it. Second, someone who accepts *that* there is most reason to save drowning children can reasonably ask *why* there is. The claim that saving drowning children is obligatory would be a substantive answer to that question. And substantive answers to 'why' questions are explanations. But if 'being obligatory' were identical to 'being what there is most reason for' it is unclear how to make sense of this. After all, nothing explains itself.

If we insist that reasons are normatively fundamental, it seems we cannot easily appeal to views like these in explaining what we mean by saying that obligations are normative 'because reasons are'. And one can motivate further scepticism towards this idea by pointing to the fact that obligations appear to have normatively salient properties that set them apart from reasons altogether, in particular at the level of deliberation.[9]

For instance, those who appreciate the normative force of obligations will tend think of them as a source of *constraints*. If they come across a drowning child, they will normally consider the options that are incompatible with saving this child as out of the question. Reasons often do not work in this way. The fact that drinking coffee would make me less tired may be a reason to drink coffee, but it does not similarly constrain me. Good and strong-willed people can appreciate the normative force of this reason, without thinking that coffee-less options are out of the question.

This difference between options that enter one's deliberative frame and those which do not seems to be rooted at a deeper level. Many of us would consider it an indication of a person's character how he deliberates. Giving serious consideration

---

[9] Scanlon (1998, pp. 155–8); Wallace (2013); Greenspan (2010) give different statements of this general idea. See also Dancy (2004).

to the inconvenience of saving a drowning child would often be taken to indicate that one had failed to appreciate why one should; inconvenience is not supposed to be relevant at all. Again, the difference with ordinary reasons is tangible. Giving serious consideration to the inconvenience of getting coffee is not normally taken to be an indication that one fails to appreciate the force of a reason to.

Lastly, obligations differ from reasons in the way they typically interact. At the deliberative level, the competition between most reasons is less charged. A reason to drink coffee because it stimulates one might compete with a reason not to drink it because one is already dehydrated. But failing to feel conflicted about this would not indicate a failure to respond to the normative facts. And even if one *did* feel conflicted here, it would be a gross mistake to think that the *type* of conflict is the same. Whether or not one decides to drink the coffee, one can move on, content that the matter has been resolved. In contrast, an upright person who had to break a promise to keep another *would* normally feel a conflict.

Obligations thus seem to constrain us differently from how reasons do, both in terms of the *options* that enter our deliberative frame, the sorts of *reasons* we consider relevant for the purposes of deliberating about these options, and the types of responses we consider cases of *conflict* to call for. The concern is that somehow reducing obligations to reasons will force one to flout these distinctions. Of course, each of these claims may be denied. Nevertheless, I take it to be a virtue of any account of obligations that it accommodates them. So in the following section, I will sketch the account of the structural relation between reasons and obligations that seems to me most promising for Pragmatism to pursue. And this account, I shall suggest, does not merely avoid the kinds of objections just discussed. It does

so in a way that allows it to explain, in terms of reasons, why obligations have these normative properties.

2.   Kinds of Reasons

On a simple and natural way of thinking about deliberation, what you think about when you deliberate are your reasons. And deliberation is a matter of comparing the weight of the various reasons that count for and against the options that are available to you. So think again of reasons like weights an old-fashioned scale. Now suppose we say that an obligation obtains just in case one side of the scale outweighs the other, *and* those weights are of a certain kind. On this view, if you are thinking about your obligations, then you are thinking about your reasons. Moreover, if you assert that there is an obligation, you are not merely asserting that the scale is balanced in a certain way. For your claim would allow that the scale *could* be balanced in this way, even though there was no obligation. Those two properties would not be identical. Indeed, your claim would be compatible with the claim that obligations obtain, when they do, only because reasons do.

This view may seem to be open to a quick objection. It may be best, even morally best, if one gives most of what one owns to charity. Doing so may be what one's moral reasons favour. But most of us believe that doing so would also be above and beyond the call of moral duty, or 'supererogatory' as philosophers say. If it is beyond the call of moral duty, however, it is not obligatory, so what one is obligated to do cannot just be what is favoured by some particular kind of 'moral' reason.[10] But this objection would be too quick. Importantly, it concedes that there is a generic sense of 'moral reason' in which at least *some* such reasons are irrelevant with respect

---

[10] See e.g. Heuer (2010); Darwall (2017a, pp. 263–4) for this objection

to determining what one is obligated to do. We might as well conclude from this that the class of reasons that settle our obligations is more restricted than this class of 'moral reasons'.[11] It is no objection to the purely structural idea that obligations obtain in virtue of certain kinds of reason.

This idea also seems to me to provide enough of an answer to the concerns raised above. To show how, it will be useful to add a bit more detail. Imagine first that all of our weights on the scale are metal weights. We can draw distinctions between different kinds of metals, like lead, and brass, and so on. We can also draw distinctions between different kinds of shapes that metal weights might have, like round, and square, and so on. So suppose we say that, for each option one has, there is a corresponding scale. And, for each option, deliberating about it involves determining which side of the scale outweighs the other. So, let us say that the square weights are those that count in favour of the option, the non-square weights are those that count against it.

Now we can put all our weights of any particular kind of metal on the scale. Let us start with those made of electrum. The square electrum weights go on one side, the non-square electrum weights go on the other. We could do the same thing with the weights that are made of billon. And we could do the same thing with any particular set of weights of any particular type of metal. So suppose we say that the kinds of weights that are relevant to determining whether you are obligated to prevent death when you easily can are all of the electrum weights. This would explain why, if you understand the normative properties of this obligation, some

---

[11] It might be further objected that this reply leaves the notion of supererogation unintelligible, but this would again be too quick. The objection casts supererogation as a type of case in which some class of reasons favour a non-obligatory option. The idea that obligations obtain in virtue of a certain kind of reason (which does not include all of those favouring the supererogatory option) can make sense of that all on its own.

weights just do not go on the scale when you deliberate about whether to save a child drowning in front of you. For you, the fact it would be inconvenient to do so would be like a lead weight. It might feel heavy, but it just does not count.[12]

It would also explain why, for you, certain kinds of options do not seem to be available. Because if you already take the electrum weights to fix the fact that you are obligated to save the child, other sorts of weights will be relevant for you only with respect to deliberating 'downstream' from that point, about different *ways* of saving the child. And the kinds of weights that go on *this* type of scale will all be of a kind that do not weigh in favour of options incompatible with saving the child. So, for you, there is simply no reason for those sorts of options to enter your deliberative frame in the first place.

Now it does not follow from this that just because you are faced with a drowning child, you would not consider *any* other reasons relevant. Because you might understand that you have other obligations than that of preventing easily preventable death, such as the obligation to keep your promises. The weights that determine this obligation might be of a different kind of metal, say, billon. And it is possible that we get different results on the scale when we put different kinds of weights on it.

In these cases, the natural question for someone who appreciates the normative force of *all* their obligations is what these weights favour *together*; that is, what they are obligated to do all-things-considered, given all such *pro tanto* obligations. So the natural thing to do is to put both the electrum weights and the billon weights on

---

[12] I am not suggesting that such considerations are irrelevant in the sense that their normative force is 'silenced' or 'excluded' by other other reasons, in the senses that John McDowell and Joseph Raz have in mind, cf. McDowell (1998, pp. 17–18, 55–6, 90–3); Raz (1999, 40 ff.) Their normative force, I would say, is not diminished, but rather relevant to a different normative standard. See Wallace (2013) for an illuminating discussion of this distinction.

the scale. Of course, if there are other sorts of obligations, then the weights that fix those facts should go on the scale too. In the end, there might be all sorts of metal weights on this scale: electrum, billon, goloid, shibuichi, and so on. And those might turn out not to share any fundamental properties, in the way that various types of jade do not. There might be no answer as to what obligations really *are*. But if the reasons that ground them *are* related in the way that electrum, billon, goloid, shibuichi are, then there will be such an answer. For, as it happens, all of those metals are silver alloys.

Indeed, that would be an important discovery in its own right. But it might also explain other interesting properties of obligations, such as how they interact, why some are weightier than others, and why we should respond differently to the reasons that ground them than to other reasons; as when the fact that there is sterling silver in your pocket is a reason for you to be happy, when the fact that there is nickel is not. Since Pragmatism is a view with such explanatory ambitions, I submit that this is the most promising account of the basic structural relation between reasons and obligations for it to explore: Obligations are *pro tanto* normative correctness standards, strict deontic facts that are settled by sets of reasons of particular kinds.[13] Our task, then, is to provide some answers concerning the fundamental nature of these reasons.

To this end, I will now move on to distinguish two different notions of obligation and clarify which of them Pragmatism takes to be fundamental. In Section 4, I discuss the role that this kind of obligation plays in the negative, blaming aspect of our moral psychology, and in Section 5, I discuss its positive role in the attitude of trust. Finally, in Section 6, I suggest that this normative nexus between obligations,

---

[13] I shall assume that the terms 'obligation', 'duty', and 'requirement' pick out the same normative property.

blame, and trust supports an attractive, unified account of the reasons that ground our obligations.

3.   Obligations and Bipolarity

The idea that an action is wrong is often expressed by saying that one *must not* do it. Similarly, something is said to be obligatory when one *must* do it. In contrast, when there is most reason to do something we often merely say that one *should* do it, or that one *ought* to do it.[14] Thus understood, it is natural to think that the type of property that Pragmatism should give an account of is this 'must-be-done' property.[15] But some obligations appear to have salient features that cannot be explained purely in terms of such a property.

Consider the case of Prometheus who has promised Priam to water the garden. As most of us would agree, Prometheus is thereby obligated to water the garden. But we might also say that Prometheus is obligated *to* Priam to do it. That, after all, seems to be why only Priam is someone to whom Prometheus would owe an apology if he failed to, and why only Priam would be in a position to forgive Prometheus for such a failure. And the same seems to be true for other sorts of obligations Prometheus might have concerning Priam; obligations not to torture him, or deceive him, or steal from him.

Now suppose instead that Prometheus is also obligated not to destroy the planet. If there were such an obligation, it would seem to be importantly different. For whatever its source, it does not seem like Prometheus would be obligated *to* the planet not to destroy it. After all, it is not as if the planet can forgive him or accept

---

[14] See McNamara (1996) for a discussion of this semantic difference.

[15] Parfit, for instance, uses the corresponding notion of a 'mustn't-be-done' property to refer to what he calls the 'indefinable version of the concept *wrong*', cf. Parfit (2011a, p. 165), emphasis in org.

his apology. Of course, one may disagree that there is anything like an obligation not to destroy the planet, but this does not matter. The point here is merely that there appears to be an intelligible distinction between obligations that one has *to* another individual, and obligations one does not.

As a piece of stipulative terminology, let us call the former *bipolar* obligations and the latter *monadic* obligations. This distinction is important. For even if *no* actual obligations are 'monadic' in this sense, the distinction itself highlights the fact that it is a non-trivial feature of obligations not to break promises, torture, deceive, kill, etc. that they have a certain kind of normative *directedness*. As Michael Thompson eloquently puts it, bipolar obligations seem to involve an 'arc of normative current' passing from one agent-pole to another,[16] one that goes together with the directedness of phenomena like forgiving and apologising.

I believe any satisfactory account of obligations must be able to explain, at least, why such bipolar obligations not to steal, kill, and lie are normative. And I think that if we can give a satisfactory account of this, then we can give a compelling account of what it might be for a monadic obligation to be normative (if it turns out that there are any). So what I will do in the present section is to discuss some accounts of what this apparent 'directedness' of bipolar obligations consists in. I will then explain why I think the reasons-based approach should leave them behind and suggest a way of moving forward.

On one suggestion, an obligation is bipolar just in case the violation of it *wrongs* another. For me to obligated *to* you not to kill you, for instance, is for it to be true that I would wrong you by killing you. This view is both simple and intuitive. But identifying bipolar obligations merely by appeal to the concept of 'wrongness' seems

---

[16] Thompson (2004).

to me unhelpful, unless we have a firmer grasp on that notion.[17] One way to address
this problem is by adding more substance to it. According to one prominent ac-
count, the property of being 'wrong' is itself to be understood in terms of a warrant
for attitudes like blame.[18] The distinction between monadic and bipolar obligations
can then be drawn along a distinction between two kinds of blame.

On the one hand, we have attitudes like resentment. Resentment is what I am
warranted in feeling when you wilfully step on *my* gouty toe, as opposed to someone
else's. On the other hand, we have attitudes like indignation. Indignation is what
I am warranted in feeling when you wilfully step on *someone's* gouty toe, whether
it is mine or someone else's. The idea that obligations are constitutively related to
blame then leads to a natural pair of ideas: What you have a monadic obligation to
do is what it would be wrong for you not to do, by warranting *anyone's* indignation.
What you have bipolar obligation to do is what you would wrong someone by not
doing, by warranting *their* resentment.[19]

This distinction plausibly characterises many obligations, such as those not to
step on feet, break promises, and lie. It is appropriate for us to resent people who
inexcusably do such things to us, for them to apologise to us, and for us to forswear
such resentment, if we accept the apology. But consider instead my obligation not
to murder you. On the present distinction, this can be a bipolar obligation *only
if* you would have a warrant to resent me, if I murdered you. The problem is that

---

[17] One may, of course, insist that the notion of 'wronging' is somehow primitive, as e.g. Thompson
(ibid.) and Wallace (2013) suggest. Wallace (forthcoming) sets out a detailed bipolar conception of
moral obligation along these lines.

[18] This idea goes back to Mill's claim that there are some things which we wish that people would
do 'yet admit that they are not bound to do; it is not a case of moral obligation; we do not blame
them' Mill (1998, p. 93). See Scanlon (1998, Chap. 4); Skorupski (2010, Chap. 12); Darwall (2006,
Chap. 4-5); Gibbard (1990, p. 40f) for more recent expositions.

[19] Darwall (2012) defends this view. The distinction between personal and impartial reactive
attitudes first appears in Strawson (2003).

it seems there is no way you could have such a warrant. If I murdered you, you would be dead and cease to exist. And those who are dead do not have warrants for attitudes because they cannot have them. Taken literally, then, this distinction seems to imply that obligations not to murder people could not be bipolar. And this is odd since that type of obligation seems to have as good a claim to being bipolar as any.[20]

Similar problems afflict alternative versions of this view. For instance, we might instead suggest that for me to have a bipolar obligation to you is for you to be warranted in being *disposed* to hold attitudes like resentment against me for violating it. This view may avoid the problem above. But it may also face other problems. Suppose I give you a promise knowing that you are a notorious promise breaker yourself. Because you break promises to other people all the time, it would be hypocritical of you to resent me for breaking my promise to you. It thus seems plausible that *you* would not have a warrant to be disposed to resent anyone for breaking promises to you. But it would seem to me a mistake to infer from this that *I* had no obligation to you to keep my promise to you. I see no reason why one could not have a promissory obligation to a hypocritical promise breaker.

These points do not refute this view, but they do illustrate a general problem one faces in understanding bipolar obligations *directly* in terms of warrants for attitudes. They also help to contrast it with another of thinking about bipolar obligations. For on a slightly different view, my bipolar obligation to you obtains in virtue of a relation between us, one that exists *independently* of any warrant for you to respond

---

[20] This is especially plausible on the view that bipolar obligations correspond to individual claim rights. Cp. Darwall (2012, p. 347): 'consider the fact that stepping on another's foot would violate her right. Or *equivalently*, that it would violate a (bipolar) obligation to her', my emphasis. If we have any rights against people, the right not to be murdered by them is surely one.

to my violation of this obligation. Let me briefly discuss two different versions of this idea.

On one version, the relation that grounds a bipolar obligation is that of *normative control*.[21] For instance, the fact that Prometheus' promise to Priam invokes a specifically bipolar obligation is explained by the kind of control that Priam can exercise over Prometheus' normative situation. Most salient is the power he has to release Prometheus from the obligation to keep the promise, but there may be others. For instance, if Prometheus breaks the promise, Priam may also release him from the subsequent obligation to apologise. A second version of this type of relational approach holds that the relevant relation is that of an *interest*.[22] Essentially, Prometheus' promissory obligation is directed to Priam because Priam has a specific and particularly weighty interest in the promise being kept, one that Prometheus is in a unique position to advance.

Both these views are subject to well-known problems.[23] The normative control account seems to get the order of explanation the wrong way round. For Priam to have any control over Prometheus's obligation to him, that obligation has to be there in the first place, independently of any control that Priam might exercise. The interest account, on the other hand, seems unable to distinguish interests which ground directed obligations from interests which do not. It may be, for instance, that someone else than Priam has a *greater* interest in Prometheus keeping his promise, but it should not follow from this that Prometheus is thereby under an obligation to *this* person to keep his promise to Priam.

Once again, the point here is not to conclusively refute these views. For our purposes, it is enough to point out that, as ways of developing a reasons-based ap-

---

[21] Owens (2012) defends this type of view.
[22] This idea is defended in Raz (1984); Raz (1986, pp. 168–70, 175–6).
[23] May (2015) provides an overview of these and further problems.

proach to directed obligations, they merely push back the problem. Insofar as 'normative control' is supposed to be a *normative* property, a proponent of the reasons-based approach must offer a further account of where the normative reasons which ground this property come in. And insofar as the plausibility of the interest account relies on a distinction between interests which ground obligations and interests which do not, it merely restates the major challenge for reasons-based approach to obligations that we identified in the previous section: to offer a distinction between reasons which ground obligations and reasons that do which.

For these reasons, I think Pragmatism is better off taking a different approach, one which combines central elements of the views discussed here. There is a compelling case to be made that bipolar obligations are directed precisely because they obtain in virtue of a certain type of relation between people. And that idea is particularly plausible, I will suggest, because there is a constitutive relation between blame and obligations. For, as we shall see, at least *one* type of blame is constitutively related to an attitude that both ascribes obligations, and presupposes a specific type of normatively significant relation between people.

4.   Blame

The aim in the present section is to identify some features of blame that will be important going forward. The aim is not, however, to defend an account of what blame *is*, nor to provide any analysis of its necessary and sufficient conditions. The phenomenon of blame is arguably too complex and varied for any such analysis to be forthcoming. Yet, arguments that appeal to facts about blame need not be unsound for that reason. Few now believe that an analysis of knowledge is forthcoming, but that has not stopped epistemologists from studying knowledge, nor should it. And

while blame may ultimately prove impervious to analysis, it does have identifiable features which set it apart from other kinds of attitudes. Those are the features we will be interested in identifying here.

Along with a number of other philosophers, I hold that blame involves at least two presuppositions: that the blamed party had sufficient reason not to do what one is blaming him for, and that he had the capacity to respond to such reasons.[24] It is irrational, for instance, to blame someone for failing to keep a promise when one realises that they were in a coma, or that their failure to keep the promise was a necessary side-effect of doing something one accepts they had even better reason to do, like preventing grave injury to another person. But blame also seems to involve further substantive judgments, both concerning the nature of the normative support involved, and the attitudes of the person blamed.

Blame is a kind of criticism, but it differs from other kinds. It is a type of criticism, for instance, to assert that someone is 'irrational' or 'unreasonable'. Someone might be subject to these forms of criticism for having inconsistent attitudes, or for failing to take any means to their intended ends. But we can also criticise people for failing to do what they have reasons to do without blaming them. One can have most reason to drink coffee merely because one likes the taste, but, in that case, actually *blaming* a person for failing to would be an excessively moralised response.

The charge of 'unreasonableness' particular to blame posits the mistake involved as a type of transgression of a 'constraint' of the type we identified above in discussing the deliberative characteristics of *obligations*. This claim is further supported by reflection on the kinds of considerations that can undercut the appropriateness of blame. While some circumstances can justify the violation of an obli-

---

[24] Skorupski (2010, Chap. 12); Darwall (2006, p. 98); Gibbard (1990, pp. 299–300); Shafer-Landau (2003, pp. 181–3).

gation, these are generally cases in which *other* kinds of obligation come into play, as when one is justified in breaking a promise to save a life.

As the phenomenon of excuses brings out, blame may also fail to be appropriate when such justification is lacking. A person might justifiably believe that he had most reason to push a button, for instance, without being in a position to know that doing so would trigger a deadly bomb. Such ignorance may be an excuse, but it does not imply that killing people was justified under the circumstances. Of course, ignorance may itself be subject to blame, but it is not always. Similarly, a temporary lack of control from being pushed, or the duress of being threatened on one's life can constitute excuses of different kinds.[25]

There is thus a gap between the judgment that something is wrong, impermissible, or unreasonable, and the judgment that it calls for blame. And this gap, it seems to me, is not filled merely by the judgment that the agent is 'responsible' or 'accountable'. Suppose, for instance, that I am threatened with serious injury, unless I mildly injure another person. If I do not injure this person, he could reasonably feel grateful, holding me responsible for preventing harm to him. If I do injure him I am no less responsible, but this does not imply that it would be appropriate to blame me.

This suggests that the gap here concerns a judgment about people's *attitudes*. This can be further supported by considering cases in which people's intentions do not succeed. Suppose you are ill. I want you dead sooner rather than later, so I intend to murder you with a lethal drug. At the last moment, someone manages to secretly switch the drug with one that will cure you instead. In this case, I did not *act* wrongly. But I would nevertheless have disregarded my obligation not to act

---

[25] See Rosen (2014) for an illuminating discussion of the relation between blame, duress, and excuses.

in the way I *intended*. Now, since you would know that I did not actually murder you, you cannot rationally blame me for *that*. Even so, it is not clear that there is *nothing* you could blame me for. It seems to me that you could blame me for trying, or simply for having the intention. And this would involve a judgment concerning your attitudes.

Blaming someone, then, involves at least three judgments: that the blamed party had sufficient reason not to act in the way he did, that he has thereby violated an obligation, and that this violation reflects the attitudes he thereby responded on the basis of. But this complex of considerations also calls for further types of 'reactions' of other kinds than judgments.[26] It is widely held that blame constitutively involves emotional responses in the key of anger, though it may take a variety of shapes.[27] This seems to me correct, but it is not the only relevant type of response. In many cases, pure angry affect would be a curious response to serious mistreatment.

Suppose, for instance, that someone sends you a poisoned invitation to their dinner party, and that you gleefully accept before learning about the poison. There would be something seriously amiss about getting angry over this without revising your intention to go the party, or at least without taking yourself to have overwhelmingly strong reason to. Even if accompanied by the sorts of judgments above, such anger would not merely lack the practical dimension that explains much of the importance we assign to blame. It would also be difficult not to take it as an indication either that you did not really appreciate the significance of what had happened, or of demeaning servility on your part.

These considerations support a more recent view, according to which blame involves revisions in one's intentions to share relations and good will with those

---

[26] I defend a more substantive account of what this complex content amounts to in chapter 3 and of the normative relations between its components in chapter 6.

[27] See e.g. Owens (2012, Chap. 1); Wallace (2011); Wolf (2011).

blamed.[28] And while these two views are sometimes taken to be rivals, in particular by their proponents, it seems to me a mistake to think that they must be. In the case above, it would be equally curious if you simply registered this attempted poisoning as a reason to cancel your acceptance of the invitation, revised intentions to invite this person for your own dinner parties, or ceased hoping that goods things would happen to him. That one should go through these responses with serene calm or placid melancholy jars with the sense that blatant disregard for one's life is something to get worked up about.

Consequently, I hold that these sorts of responses are equally central to blame. But because I am not defending an account of what blame *is* or what its necessary and sufficient conditions are, I shall merely assume that putative instances of blame that do not involve such responses are, at best, defective forms of blame. And I shall generally take the fact that there would be reasons for these types of responses to be strong evidence that blame would be appropriate and *vice versa*. Reasons for these types of affective responses are *inter alia* reasons for the corresponding types of cognitive responses, and our judgments concerning the conditions under which they are appropriate generally go together.

Nevertheless, I take these to be substantive claims rather than conceptual truths. We can imagine beings who routinely judge that others violate their obligations while having weighty reasons not to, but who have no capacity for emotional states like anger. Similarly, we can imagine a world full of immense value inhabited by beings who routinely judge that things are valuable, but who have no capacity to value things. It is not clear, I think, that these beings would be irrational, or that they would not share our concepts of 'obligation' and 'value'.

---

[28] Scanlon (2008, Chap. 4).

This analogy with value illustrates an important point, however.[29] It is not a trivial intuition that the lives of these non-valuing beings in a world of immense value would be lacking in an important respect. They would be missing out. And, unlike any of these beings, we *do* have the capacity for valuing things. So, while it may not be for them, it seems entirely plausible to me that it can be a mistake *for us* not to be affected by such things as values. For someone with the capacity to appreciate the value of art, for instance, the destruction of all remaining art would not merely be a reason to conclude that the world now contained less value. It would be a reason to *feel sad*.

Something similar is true of the relation between blame and our judgments concerning obligations. The difference, of course, is that what we react to in holding attitudes like blame is not merely events or objects, but the behaviour of people. Blame is a negative reaction *to them*, and such attitudes are thus, in the sense coined by Peter Strawson, *reactive* attitudes of a negative sort. And, in introducing that notion, Strawson made some further observations about these kinds of attitudes that are relevant to understanding both why they should involve these practical and emotional elements, and why they should seem similar to responses to value.

First, he noted the distinction discussed above, between 'personal' reactive attitudes, such as resentment, and 'impartial' reactive attitudes, such as indignation.[30] Second, he proposed that the source of normative support for these attitudes was connected to 'the great extent to which our personal feelings and reactions depend upon, or involve, our beliefs about' other people's attitudes, in particular those with whom we share meaningful relationships. And third, he specifically suggested that the impartial reactive attitudes were 'vicarious analogues' of the personal ones, to be

---

[29] I owe the analogy itself to Wallace (2011), though Wallace phrases it somewhat differently.
[30] See page 25.

understood in terms of their relation to seeing and reacting to someone as a fellow member of a 'moral community'.[31]

These observations seem to me relevant for two different, though related, reasons. On the one hand, they raise the question of how precisely the normative grounds for the impartial and the personal reactive attitudes are related. On the other hand, they actually suggest a hypothesis about this relation. For if the impartial reactive attitudes are analogues of the personal ones, and if Strawson's own proposal is correct, the explanation for the 'importance' of these attitudes should be understood, quite generally, in terms of their role in personal relationships and analogues thereof.

For our purposes, it is not essential whether this was, in fact, the idea that Strawson had in mind. The important point is that, if something like that hypothesis were *true*, this would provide further support for Pragmatism's central idea; that the normative source of our obligations can be located in the kind of normative relation that is central to interpersonal agency. I shall argue that something like this hypothesis *is* true. And, as I shall now suggest, we can identify this central normative relation by considering a type of attitude that disposes to such personal forms of blame.

5. Trust

Because personal reactive attitudes such as resentment are blaming attitudes, they form part of the negative, 'sanctioning' aspect of our moral psychology that involves

---

[31] Cf. Strawson (2003, in particular pp. 75, 83, 90). Precisely *how* Strawson thinks of the normative connection here is a difficult question that I will not try to settle. I take it as relatively uncontroversial, however, that Strawson took himself to be making observations about these attitudes with normative implications, as when he remarks that 'Inside the general structure or web of human attitudes and feelings of which I have been speaking, there is endless room for modification, redirection, criticism, and justification. But questions of *justification* are internal to the structure or relate to modifications internal to it' Strawson (ibid., p. 91), my emphasis.

thinking about obligations. But, by focusing on these attitudes, we can also look more carefully at the role that thoughts about obligations play in a more positive aspect of our moral psychology. My contention in this section will be that they do so in particular in the attitude of *trust*.

Claims about 'the attitude of trust' raise difficult questions, because the term 'trust' itself is unusually contested.[32] However, as with the attitude of blame, my intention here is not to defend any particular theory about what trust *is*. In this context too, it is doubtful whether any such account can succeed. Consequently, my aim here is to present the notion I take to be important, explain how it differs from related notions, and present some arguments for focusing on it.

The notion of trust I have in mind brings together two features which are at the core of a wide variety of views on the nature of trust.[33] The first is that trust is a kind of expectation which disposes to certain emotional reactions. Betrayal is most commonly mentioned here, but it hardly covers the whole territory. If a trusted friend fails to help you, it may be equally appropriate to feel let down, disappointed in him, hurt, or even resentful. I take it that trust thereby disposes to various forms of blame, specifically a variety of negative, personal reactive attitudes.

The second is that trust is a type of reliance relative to specific domains of interaction. To trust someone is to trust him with respect to *something*. You trust someone to ϕ only if you rely on him to ϕ. To add a bit more detail, let p be the

---

[32] Some examples of controversy: Some believe trust is a cognitive or belief-like state, others hold that it is a type of emotion, cf. Becker (1996); Faulkner (2011); Hardin (2002); Helm (2014); Holton (1994); Jones (1996); Lahno (2001); McGeer (2008) Some believe that it is possible to trust at will, others reject this view. Some hold that trust is as much an action as it is an attitude, others reject it, cf. Faulkner (2011) Some hold that it is a two-place relation between one agent and another, others that it is a three-place relation between one agent, another agent, and an act-type, cf. Faulkner (2015) and Domenicucci and Holton (2017). And some hold that any analysis of trust is bound to fail, cf. Simpson (2012) I accept this latter view.

[33] See Baier (1986); Darwall (2017b); Domenicucci and Holton (2017); Hawley (2013); Hieronymi (2008); Hinchman (2017); Holton (1994); Jones (2017); O'Neill (2017).

proposition that this person ϕs. To rely on p is then to be disposed to use p as a premise in practical reasoning when p is relevant to the choice of some option ψ. Specifically, p is relevant to a choice of ψ-ing just in case the normative status of ψ-ing conditional on p is different from the normative status of ψ-ing conditional on $\sim p$.[34] Putting this point in terms of reasons, we can say that the normative status of some option is the total set of reasons counting for and against that option. In other words, p is relevant when the truth or falsity of p makes a difference to how you have most reason to respond. As I am using this term, then, reliance is an attitude. It is thus to be contrasted with the sense of 'reliance' in which a comatose patient might be said to 'rely' on hospital care. I shall use 'dependence' in place of this other notion.

It is no coincidence that these two claims are widely accepted. Essentially, reliance is an attitude to perceived dependence on something beyond one's own intentional control.[35] This attitude can involve the further assumption that another person is in a position to respond to this fact; that success in one's reasonable aims is dependent on *him*. When it does, it is an attitude to perceived vulnerability to another in this fundamentally normative sense. And in holding him accountable for responding to this fact, one is liable to hold specifically *personal* reactive attitudes. The notion of trust here merely combines these two claims:

Trust: For x to trust y to ϕ is for x to (i) rely on y to ϕ and (ii) be disposed thereby to hold negative personal reactive attitudes towards y for failing to ϕ.

---

[34] Alonso (2014); M. N. Smith (2010); Hawthorne and Stanley (2008).

[35] Matthew Noah Smith employs this idea to identify objects of reliance by what he calls '*The Remainder Operation*: Given my plan, given my beliefs about the world, given my intentions, and given all my other commitments, what are the remaining conditions that are out of my control but are nonetheless necessary for the successful realization of this plan?' M. N. Smith (2010, p. 146).

Unless otherwise noted, I shall use 'trust' to refer to this attitude. The central question, then, is why we should focus on *it* rather than other attitudes the term 'trust' might pick out. And my answer is that, given the account of obligations and blame defended above, it captures many of the intuitions that motivate philosophers who think trust somehow plays an important normative role.

Consider first the point that trust disposes to blame. As I argued above, blame involves the judgment that another person *was* obligated to respond in some way. But since trusting someone is not itself to blame them, it does not involve that particular judgment. Rather, trust disposes one to make that judgment in case this person fails to respond as one trusts him to. Thus, the judgment trust actually involves is that blaming a person *would* be appropriate, if he failed to do as one trusts him to do, because there is *now* an obligation on him to respond in this way.[36]

Additionally, trust involves the judgment that the trusted will *not* fail to do what he is trusted to do. If Jones relies on Johnson to bring wine, Jones not only thinks that Johnson will do this, he employs the proposition 'Johnson will bring wine' as a premise in his reasoning. Part of the explanation for why such reliance disposes to blame is that we often care about whether things turn out as we expect them to. But this is not all we care about when trusting people in these ways. If Jones' trust involves the judgment that it *would* be appropriate to blame Johnson for failing to bring wine, it further presupposes that Johnson is in a position to respond to reasons to do so and that is therefore *responsible* for bringing wine. The disposition to blame would be inappropriate if this condition were not satisfied, for instance if Jones thought Johnson was under hypnosis, or merely caused by some funny virus to haul wine around. In trusting him, Jones supposes that Johnson will come to

---

[36] Nickel (2007) also defends this view.

do so by responding to reasons, that is, that Johnson will do what he trusts him

to *for reasons*, either by spontaneously acting on them, or after deliberating more

explicitly about them. Does it matter for the attitude of trust in question here *which*

reasons people act for? I believe it does, and I think there is a reasonably strong

argument to that effect.

On one view, an action has moral worth (or is praiseworthy) just in case it is

done for the reasons that make it right.[37] This view is plausible for a number of

reasons. For instance, imagine a long list with descriptions of actions, each of which

are marked as being either 'right' or 'wrong'. Andrew, who happens to have such a

list, is very concerned with doing what he is obligated to. Luckily, Andrew knows

that his list is correct, so he does all and only those actions which are marked 'right'.

Because giving to charity is marked 'right', Andrew does that, and so on for other

right actions. But if someone were to ask him whether he thinks that poverty is bad,

or that people have reasons not to want to suffer, etc. Andrew would be nonplussed.

He does not think such facts are reasons to give to charity. When Andrew gives

to charity, his reason is something like 'I am obligated to' or that 'it's on the list'.

And that, I think we should agree, makes Andrew's actions seem odd and slightly

obsessive, rather than praiseworthy.

One diagnosis of Andrew's oddity is that he seems not to understand *why* the

things on the list are obligatory. As we can also say, it is because Andrew does

not understand the reasons that *make* the actions on his list right or wrong. In

contrast, consider Andrea who realises that Andrew really is in a terrible state and

feels compelled by that thought to help him see a psychiatrist. Andrea has no list like

Andrew's, and she is not too concerned with checking that 'helping moral fetichists

---

[37] Markovits (2010) and Arpaly (2002) defend different versions of this view.

get psychiatric care' is on Andrew's list of right actions, even though she could. Nevertheless, it seems like there is at least *some* worth to her actions. In helping Andrew for the reason she does, she seems like a decent person. And this makes it plausible that, if an action is done for the kind of reason that makes it right, it is worthy of an attitude like praise.

Can actions be worthy of praise if they are done for other reasons? Reasons like Andrew's do not seem like good candidates, but perhaps there are others. For instance, Bernard Williams argued that for a man whose wife was in distress, a good reason to save her would surely be a fact like 'She is my wife'.[38] But that, Williams claimed, is not the kind of reason a theory of what it is morally right to do would give someone for saving their wife. Rather, it would be something like 'this would increase happiness overall' or 'this would comply with the moral law' or 'this is my wife and in situations like this it is permissible to save one's wife'. And it just cannot be true that it *never* has any worth for people to act for reasons that have to do with their deep, personal relations with other people. So it is not true that actions are worthy of praise *only* if they are done for reasons that make them right.

This argument can be challenged on two points. First, it seems to assume that a moral theory is incompatible with the idea that facts about one's relationship with other people can make it the case that one has obligations to them. As the above characterisation of Pragmatism makes clear, this is not true. Nothing about the claim that obligations obtain in virtue of particular kinds of reasons rules this out. Second, the argument seems to assume that if a theory posits an explanation of *why* something is a reason, that explanation itself must be 'part of' the reasons it explains. For if that were not true, it is difficult to see why a moral theory would

---

[38] Williams (1981).

have to claim that people's reasons to save their spouses include facts about overall happiness, or moral laws, or permissibility. Of course, this view is intelligible, and some do accept it.[39] But it is not clear why a moral theory *must* adopt this view, and it seems to me more plausible simply to deny it.

Taking both of these points into consideration, it is reasonable to hold that actions done for the kinds of reasons Williams has in mind are worthy of praise on the view above. And if we take into consideration cases like Andrew and Andrea's, all of these points together make it plausible that actions are worthy of praise just in case they are done for the reasons that make it right, i.e. the reasons constitutive of an obligation. Now, since blame and praise are intuitively each other's contraries, it is also reasonable to expect that the conditions under which such attitudes are appropriate mirror each other. Given the claims I have defended, this seems to be the case. For I argued above that blame involved the presupposition that the blamed party had inexcusably failed to respond to the considerations that obligated him. So, just as there are reasons to praise people who respond to the reasons that obligate them, there are also reasons to blame people who are in a position to respond to such reasons, but who inexcusably fail.

Of course, trust is not just an attitude to people and their thoughts about reasons. In trusting someone to act, for instance, our concern is also with whether they will act, and we assume that they will. Combined with the further facts that (i) trust also presupposes that the trusted party is in a position to act on the basis of the facts that obligate him, and (ii) that he will act for *some* reasons, and (iii)

---

[39] Consider, for instance, Thomas Nagel's claim that reasons to X are modally sufficient conditions for the truth of propositions of the form 'There is a reason to X' Nagel (1970, pp. 90–5). Similarly, Joseph Raz claims that 'A complete reason consists of all the facts stated by the non-redundant premises of a sound, deductive argument entailing as its conclusion a proposition of the form "There is a reason for P to V"' Raz (2000, p. 228).

that these reasons are *not* of the kind that would call for blame, then (iv) the class of reasons that trust presupposes the trusted party will act on narrows down considerably. In fact, it narrows down to the very class of reasons that Pragmatism is concerned with; namely the reasons that make it the case that he is obligated to act as one trusts him to. So trust not only ascribes obligations to people, it also ascribes motivating reasons to them.

Together, these two ideas seem to me to explain many of the intuitions that philosophical accounts of trust are supposed to capture. They explain, for instance, why it is so appealing to think that trust involves a kind of 'optimism' in another person. It is often claimed that trust is therefore a kind of reliance on another person's 'goodwill'.[40] But this idea is subject to the well-known 'trickster problem': A deceitful trickster might rely on your goodwill towards him to exploit you, but it hardly seems to follow that he thereby trusts you at all. The present account easily avoids this problem. For insofar as the trickster's reliance on your goodwill does not ascribe any obligation to you, it follows on this account that he does not trust you.[41]

It also explains the motivation for resisting what is sometimes referred to as the 'moralisation' of trust.[42] It is not clear, for instance, that one criminal could not trust another to do something they both know to be morally wrong.[43] And this criminal might blame his accomplice for failing to. In doing so, he would, on the present account, be ascribing an obligation to him to do this morally wrong action. But since it is a substantive question which kinds of reasons can constitute obligations,

---

[40] See Jones (1996) and Baier (1986).

[41] Cogley (2012) defends a similar version of this solution to the trickster problem.

[42] See Jones (2012); Jones (2017) and Nickel (2007) for discussion of this point.

[43] Liberto (2016) defends a similar claim that trusting a person to keep promise cannot necessarily involve the ascription of moral character to him.

we should not consider this conceptually confused.[44] In principle, one can ascribe such an obligation to people in trusting them, if one thinks they have *bona fide* reasons to do it. Doing so would clearly involve a mistake, but the mistake might just be factual, not necessarily a conceptual or rational one.

Most importantly, however, it explains the widely agreed connection between trust and ascriptions of reliability. As I argued above, obligations are distinctive in that they present themselves in deliberation as presumptive constraints that are not easily outweighed by ordinary practical reasons. There may be cases in which the facts require one to break a promise, but these are standardly assumed to be special cases, not similar to how things normally are. To say that they are not easily outweighed is thus to say that they reliably win out in competition with other reasons. So, in ascribing obligations, trust involves the judgment that the reasons to act as trusted are not easily outweighed, indeed that they will win out, other things being equal. Because it involves the judgment that the trusted party will be responsible for acting, it also involves a judgment that he has the capacity and competence to do so. And because it further involves the judgment that the trusted party will act *for these reasons* that reliably obtain, it further ascribes to him reliable motivation for doing so.

The fact that this notion of trust thereby captures so many intuitively plausible claims seems to me a good reason to think that it carves this part of our moral psychology at its joints. It also seems to me a natural fit with the reasons based account of obligation defended above and the corresponding account of blame. This is important. For if we put these claims together with those discussed above, they seem to lead naturally to some interesting unifying hypotheses.

---

[44] These criminals might hold the view considered by Foot (1972), that the constraints we refer to as moral obligations merely purport to be supported by reasons in the way that rules of etiquette do. This view is not conceptually confused either.

We began in Section 2 with the idea that obligations obtain in virtue of certain kinds of reasons. In Section 3, we noted the distinction between bipolar obligations and monadic obligations, and the idea that the normative grounds of obligations can be identified via some relation presupposed by the kinds of blame that they make appropriate. We can add to this the 'Strawsonian' hypothesis from Section 4, that the justification for blame is somehow internal to a way of relating to other people, and that 'impartial' forms of blame are analogues of more basic 'personal' forms of blame that are internal to personal relationships. But if that is right, then the kind of attitude that disposes people to these kinds of blame ought also to be internal to such relationships. And that this is the attitude of trust is, of course, the idea here from Section 5. So trust ought to be internal to relationships too.

Now, given that trust and blame both constitutively presuppose obligations, those presuppositions have to be internal to relationships as well. And if relationships involve people relating *to* each other, it is not too implausible that the kinds of obligations involved here are the ones people would have *to* each other. But if this correct, then there would have to be a kind of normative nexus between bipolar obligations, trust, personal forms of blame, and relationships. And if the further Strawsonion hypothesis is correct, that impartial forms of blame are really just vicarious analogues of the personal ones, then we should think that there was some kind of analogue of that normative nexus that played across the so-called 'moral community'. There should be an analogue of a relationship between everyone that constitutively involved an analogue of personal forms of trust that disposed to analogues of personal forms of blame.

This might seem like a slightly radical idea, but it packs a powerful theoretical punch. It makes good on the suggestion from Section 3, that bipolar obligations are

somehow normatively fundamental. It also goes some way towards explaining why blame should involve the emotional and practical elements discussed in Section 4. Blame could involve revisions of intentions to interact with people, because what they did somehow affected our relationships with them, relationships that involve various intentions to interact with people. And it could involve negative emotions, because such relationships involve something worth caring about in a kind of valuing way, and that blameworthy behaviour somehow causes the valuable thing to break, or become lost.

Finally, it makes Pragmatism's job of identifying a specific 'kind' of reasons that might ground obligations much easier. For if bipolar obligations are normatively fundamental, the kinds of reasons Pragmatism is ultimately after have to be those that ground them. Those reasons are supposed to obtain in virtue of some kind of relation between the person who is obligated and the person to whom he is obligated. And if all of the above is correct, there is at least one type of relation like that we can identify whenever it is the case that such an obligation obtains. This is the relation of reliance, the attitude one person holds towards another, when he takes his own agency to be dependent on that of the other's. Let us call this the

RELIANCE VIEW: For R to be the kind of reason that can contribute to making x obligated to y is for R to be a fact about y's reliance on x.

If we can put all the ideas above together to defend something like the Reliance View, I think Pragmatism starts to look promising. But whether we *can* defend the Reliance View is obviously a substantive question since, like all views, it has substantive implications. For instance, it implies that if Greg relies on Gemma to build him a spacecraft, that very fact is reason for Gemma to do so. However, while

this may seem counterintuitive, it remains to be seen whether it also constitutes an objection. For the Reliance View, as I have just stated it, does not imply that it would be a *good* reason to do so, or a *weighty* one, or one that amounts to an *obligation* all by itself. After all, just as Greg plausibly has good reasons not to rely on Gemma to build him a spacecraft, Gemma plausibly has good reasons not to build it. These reasons might help explain why Gemma's reason to build Greg a spacecraft is an insignificant reason when we evaluate all of her reasons together. So while the Reliance View implies that it would be a reason, and that such a reason *can* contribute to making it the case that Gemma is obligated to Greg, it does not follow from this that Gemma *is* obligated to Greg to do anything, and so it does not imply that this reason *does* contribute to obligating her in this way.

What this point illustrates, however, is that the Reliance View needs to offer *some* explanation of how such reliance-based reasons come to ground obligations between people. And however it does this, it will have to earn its keep in the way other accounts of obligations do, by providing some plausible predictions about the circumstances under which people are subject to specific kinds of obligations. So in the next chapter I explore whether the Reliance View can do this for obligations that at least appear to be bipolar: promissory obligations and obligations within special relationships. I will argue that it is actually sufficiently plausible for us to explore in chapter 3 whether the Reliance View might apply to all obligations, and that, I hope you will now agree, would be a really interesting thing to find out.

Chapter 2

The Normativity of
Bipolarity

1. Where We Are

I suggested above that Pragmatism might fruitfully develop the idea of a norma-
tive nexus between bipolar obligations, trust, blame, and relationships. I further
suggested that the type of relation which provides the reasons in virtue of which of
such obligations obtain is the one involved in trust: reliance. This Reliance View
is interesting to explore for our purposes. For consider: If Jane relies on Jimmy
to help her out, then Jane presupposes that her agency is somehow dependent on
Jimmy's. And the interdependence of agency is precisely what Pragmatism takes to
be fundamental to the normativity of obligations. If Pragmatism is going to ground
the normativity of bipolar obligations, defending the Reliance View would seem
like a good place to start. So the aim in this chapter is to show that the Reliance
View also gives us an independently plausible account of paradigmatically bipolar
obligations. I will do this by discussing two kinds of obligations that seem to have
a bipolar structure, promissory obligations and obligations of special relationships.

   After setting out some reasons for focusing on promissory obligations in par-
ticular, I go on in Section 2 to discuss a variety of views on how giving a promise
obligates one to keep it. Each of these views is attractive in its own way, but all of

them face some difficult problems. I will be arguing that dealing with these problems leads us naturally to something like the Reliance View, and I therefore go on in Section 3 to illustrate in further detail precisely how it works. Central to this view is the attitude of care or concern presupposed by those who give and accept sincere promises. Then, in Section 4, I discuss a variety of views of the obligations of special relationships. Again, each of these views is attractive, but they also raise some difficult questions. Here, I argue that the Reliance View offers a simple way to answer such questions. On the independently plausible view that relationships involve attitudes of care and concern, the account of promissory obligations can easily be extended to these special obligations too.

The general point in this discussion will be that the Reliance View is particularly well placed to capture the characteristic bipolar features of these kinds of obligations. Because they are perhaps most evident in the case of promissory obligation, I will start by discussing some features of these obligations which I believe a satisfactory account ought to capture.

Most agree that if Prometheus gives a promise to Priam, then Prometheus is obligated to keep it, other things being equal.[1] So if we are going to explain why all obligations are normative, we had better be able to explain why this kind of obligation is. Now, consider the view that promissory obligations are not bipolar. On this view, if we ask whom Prometheus' promise obligates him *to*, then the answer is: no one. That sounds pretty odd. After all, the obligation obtains *because* Prometheus made the promise in the first place, and at least the *promise* could not have been made if it were not made to someone. A simple explanation for why this

---

[1] Even act-consequentialists who deny that we *always* have an obligation to keep our promises tend to agree that, on balance, consequentialist considerations *almost always* make it morally obligatory for people to keep their promises. See e.g. Singer (1972).

view seems odd is that the obligation itself is necessarily directed in precisely the way the promise is; towards someone.

A further interesting feature of promises is that the obligation to keep them is usually at someone's discretion. For Prometheus' promise to obligate him at all, it has to be *accepted.* Imagine that Prometheus promises to get rid of Priam's bow and does so. If someone were to point out that no one ever actually consented to Prometheus' getting rid of Priam's bow, it would be bizarre for Prometheus to reply that his promise to do so was nevertheless binding. Similarly, the obligation to keep a promise is something from which Prometheus can be *released.* For instance, if someone were to regret accepting a promise from Prometheus, it would be equally bizarre for Prometheus to reply 'Too late!' He cannot on his own decide whether his promises are binding. The crucial point here is that when Prometheus gives a promise to, say, Priam, there is precisely one person who can do this sort of accepting and releasing: Priam, of course. The obligation to keep the promise thus goes together with some kind of normatively significant relation between them.

A final reason to start with promises is that it provides perhaps the best illustration of the kind of reliance I have argued is central to trust. We are normally disposed to blame those who break their promises to us by holding the types of attitudes that are central to trust: feeling betrayed, let down, hurt, or resentful. We think that there are obligations on people to keep their promises. And we expect them to comply with such promises, not because it may be convenient or expedient for them to do so, but because we take them to understand that their giving a promise puts them in some kind of normatively significant relation to us which makes trust and the kinds of blame to which it disposes appropriate.

So that was three important features of promissory obligations: Their apparent directedness from promisor to promisee, the standing of the promisee to accept and release, and a normatively significant relation which makes attitudes like trust and personal forms of blame appropriate for the promisee. Let us now look closer at some views about how we should understand these various features and their normative grounds.

2.   Promises

Until fairly recently, one of the most widely accepted accounts of promissory obligation was Rawls'. Rawls' account takes its lead from the observation that promises are generally communicated and accepted by *conventional* means. But this, he further observed, is not particular to promising. In fact, promising is just one of many kinds of conventions around which the incurrence and discharging of obligations is organised. On Rawls' view, such conventions generate genuinely normative obligations just in case (i) they are just, and (ii) people voluntarily benefit from them. According to this *Principle of Fairness*, when *you* are obligated to keep your promises, that is because *you* have voluntarily benefitted from other people keeping their promises, and because it is not unfair that people should keep their promises.[2]

As Rawls was aware, however, this analysis is subject to some difficulties, two of which have been widely discussed. The first has to do with the way in which the Rawlsian analysis construes such obligations. If the obligation to keep a promise is to be explained in terms of the Principle of Fairness, this seems to suggest that violating the obligation is a matter of being unfair *to those who contribute to the practice*. That claim may be plausible. But the Rawlsian analysis further seems to

---

[2] Rawls (1971, pp. 18, 52).

imply that *all* obligations to comply with just conventions are *equally* owed to *everyone* who voluntarily benefits from them.[3] This is less plausible.

If Prometheus breaks a promise to Priam, most will agree that this makes it appropriate for Priam to hold personal reactive attitudes like resentment. It may also be appropriate for other people to hold other kinds of reactive attitudes, like indignation, but not resentment. That difference would be easy to explain by pointing out that Priam is the only person to whom Prometheus was obligated. But that is precisely the kind of explanation that the Rawlsian analysis seems to prevent, since it implies that Priam is *not* the only person to whom Prometheus was obligated by giving him this promise. So it does not explain why only Priam has a reason to feel resentment, or why only he could forgive Prometheus, or accept an apology from him.

The other difficulty with the Rawlsian analysis is that its explanation for the normativity of the obligation explicitly assumes the existence of something like an established social convention or practice. As Rawls points out, it does not matter how the practice comes about. Once it is there, the Principle of Fairness explains under which conditions one must comply with it. The problem is that it seems like there could be cases where one might incur a similar, if not identical, type of obligation in the absence of such a practice. Consider the following kind of case conceived by T.M. Scanlon.

> I am stranded in a strange land. In an attempt to get myself something to eat,
> I make a spear. I am not very good at using it, however, and when I hurl it

---

[3] As Rawls puts it 'obligations are normally owed to definite individuals, namely, those who are cooperating together to maintain the arrangement in question' Rawls (1971, p. 113). The 'normally' clause is somewhat confusing here, since Rawls goes on to say on the same page that he believes this actually applies to all the obligations he discusses. That he does not consider promising an exception is apparent from his subsequent claims that in making a promise 'one invokes a social practice' that is 'obviously in the common interest' cf. Rawls (ibid., pp. 346–8).

at a deer it goes wide of the mark and sails across a narrow but fast-running river. As I stand there gazing forlornly at my spear, lodged on the opposite bank, a boomerang comes sailing across and lands near me. Soon a strange person appears on the opposite bank, picks up my spear, and looks around in a puzzled way, evidently searching for the boomerang. It now occurs to me that I might regain my spear without getting wet by getting this person to believe that if he throws my spear across the river I will return his boomerang. Suppose that I am successful in this: I get him to form this belief; he returns the spear; and I walk off into the woods with it, leaving the boomerang where it fell.[4]

As Scanlon goes on to observe, this seems objectionable in pretty much the same way that breaking a promise does. It seems one would have done enough to incur an obligation to throw back the boomerang, and to make it appropriate for one's disappointed counterpart to blame one for failing to. Yet, since the example more or less rules out the existence of any institutionalised practice, it seems the Rawlsian account does not have the resources to explain why.

The example suggests that we can make sense of the idea that someone manages to give a promise without making any assumptions about *how* this is done. So even if a proponent of the convention based account could find a convention or practice here, it would seem to be an idle wheel. And this suggests that what matters to the existence of a promissory obligation is not the *means* by which the promise is made, but rather the *effect* of providing the promise itself. So, according to a more recent account, when Prometheus makes a promise to Priam, what happens is that Priam comes to expect that Prometheus will do what he promised. And, this account holds, we have good reasons to want people to do the things they lead us to expect they will do.

---

[4] Scanlon (1998, pp. 296–7).

This is precisely the type of account Scanlon takes such examples to support. Scanlon suggests that the conditions under which the inducement of such an expectation generates an obligation to perform are captured by his *Principle F*:

> If (1) A voluntarily and intentionally leads B to expect that A will do X (unless B consents to A's not doing so); (2) A knows that B wants to be assured of this; (3) A acts with the aim of providing this assurance, and has good reason to believe that he or she has done so; (4) B knows that A has the beliefs and intentions just described; (5) A intends for B to know this, and knows that B does know it; and (6) B knows that A has this knowledge and intent; then, in the absence of special justification, A must do X unless B consents to X's not being done.[5]

Scanlon further provides an extensive defence of the claim that Principle F is merely one of a class of similar principles. What those principles have in common is that they restrict permissible ways in which we may affect other people's beliefs about what we intend to do, and specify what we are obligated to do *if* we lead them to form these expectations. Particular to promising is that the obligation is to *perform*, rather than reciprocate, prevent a potential loss, or compensate it.

Like the convention based account, this account underwrites an explanation for why promissory obligations would be directed at all, as well as a rationale for generalising that explanation to other sorts of obligations. However, its more specific account of the bipolar features involved seems more plausible. If Prometheus becomes obligated to Priam to respond in any of these ways by leading him to form an expectation, this obligation obtains, Scanlon holds, in virtue of reasons for Priam to want Prometheus to respond in these ways, and to object to him if he fails to do so. And the account here can point to a salient relation between them to explain why Priam would have this reason: Prometheus has, by his actions, lead *Priam* to form a

---

[5] Scanlon (1998, p. 304).

particular kind of expectation, not anybody else. Nevertheless, grounding the obligation in the relation that an expectation (in the purely descriptive sense) has been generated, however, leads to a pair of related problems for Scanlon's account.

One of these has to do with the attitude of expectation itself. For instance, Prometheus might make a very reliable, public prediction that he is going to do something Priam wants him to do, thereby causing Priam to expect him to. It seems he can do that without thereby promising to do it, or incurring any obligation to. Moreover, suppose Prometheus and Priam are both aware that Prometheus gets all jittery and nervous about promises; making them often causes him to forget to keep them. This is unfortunate, in particular because it makes it difficult for Priam to expect him to keep it (in the purely descriptive sense). In this case, Prometheus would be *less* likely to generate an expectation of performance in Priam when he incurs an obligation by making a promise, than when he avoids doing so by making a prediction. But it does not seem that this would exempt him from any obligation to keep his promises.[6]

The other difficulty concerns the explanation for why people can reasonably expect that promises will be kept in the first place. According to Scanlon, promises are intended to indicate that one takes seriously the fact that one is obligated to keep it, and *this* is what gives the promisor a reason to expect that it will actually be kept.[7] The problem is that, on the expectation-based account, that obligation is supposed to obtain *because* one has generated this expectation, which, on this line of reasoning, one can do only if the obligation is already there. So, either it must be false that the obligation obtains as a consequence of generating an expectation in this

---

[6] Though Scanlon seems to have been aware of this problem, (cf. his discussion of the 'Profligate Pal' in Scanlon (ibid., p. 312f)) the problem posed by the latter type of case was first noted by Owens (2006, pp. 54–6).

[7] Scanlon (1998, pp. 306–7).

way, or this explanation simply helps itself to the very obligation it was supposed to explain.[8]

This result indicates a more general problem. As we have seen, the convention-based account and the expectation-based account both aim to explain the bipolar normativity of promissory obligations by appealing to some kind of relation between promisor and promisee. But neither is able to specify any relation between promisor and promisee that qualifies as a necessary condition for promissory obligation. Where the relation which grounds the obligation on the convention-based account can not make sense of the promissory obligation's bipolar features, the relation identified by the expectation-based account makes sense of those features only by bootstrapping the obligation into place. So perhaps the solution is to accept that the bipolar normativity is simply built into the promissory obligation itself, not derived from some other kind of relation.

This is the thought behind the more recent *authority*-based account. On this type of view, in giving Priam a promise, Prometheus recognises a certain kind of authority on Priam's part to hold Prometheus to an obligation that his promise creates. Essentially, the idea is that merely causing someone to form an expectation is different from undertaking a promise, because promises involve a normative element of assurance. It is in virtue of this element of assurance that a promisee comes to have authority to demand performance from the promisor. But what is it to have such authority?

According to Darwall, it is (at least partly) to have a warrant for personal reactive attitudes when such a demand is disregarded.[9] As Darwall argues, it is only by appealing to the transfer of a warrant for such reactive, blaming attitudes that we

---

[8] Cf. Mason (2005).

[9] Darwall (2011, p. 262).

can understand why promissory obligations are normative at all, and why they have the bipolar structure they do. The idea that warrants for such attitudes are crucial to understanding the bipolar structure of promissory obligations seems to me plausible. As noted in the previous chapter, however, there is a general problem with identifying bipolar obligations with warrants for blame. For instance, incurable promise breakers might have no warrant to blame others for breaking promises, because that would be hypocritical of them. Yet such hypocrisy cases are not the only ones that cause trouble for this type of account as applied to promises.

Here is a case similar to the one discussed in chapter 1.[10] Suppose you will soon die from some mildly painful disease, if you are not treated for it. The only possible treatment has a very limited chance of success. If I tried to treat you, it is very likely that you would live for many years in agonising pain, unable to communicate with others. Even if you consented to this treatment, it would be permissible for me to refuse to risk causing anyone such suffering. But, if you consented to it, it would also be permissible for me to give you this treatment. Now, suppose I promise you to give you this treatment before you die, and that you accept my promise. It would be a consequence of violating this promissory obligation that you could have no warrant to resent me for it, since you would be dead. So it is not clear how, in giving you this promise, I could transfer a warrant to you to resent me, if I were to break it. And if there could be no such transfer, it could not explain the promissory obligation.[11]

It thus seems the authority-based account needs a different kind of explanation here. One alternative has been suggested by David Owens. On Owens' account,

---

[10] See page 25.

[11] This case is importantly different from the usual 'deathbed promise' case. The problem here is not to explain how I could have a promissory obligation after you die. The problem is to explain how I could have an obligation to give you this treatment before you die *given* that you will soon die, if I do not.

when Prometheus makes a promise to Priam, he thereby grants Priam the authority to decide whether he should perform as promised.[12] For Prometheus to break a promise, then, is for Prometheus to violate Priam's associated right to make this decision. Note that this view neatly sidesteps the problem generated by the example above. For if I break my promise by failing to treat you before you die, I have made it impossible for you to decide whether I should treat you or not, thereby violating your right to do so. But this account faces a different problem, namely that of explaining how promises may still be binding when such an authority is relinquished.

Suppose Prometheus has promised Priam to join him on a road trip.[13] Priam is *really* looking forward to it, but Prometheus likes the idea less and less as the trip approaches. Priam realises this and he feels bad about it. He wants Prometheus to join him on the trip because it would be fun, but it is not going to be fun if Prometheus grudgingly goes along just because he promised to. So he says to Prometheus 'Look, forget about the promise, I will go on my own'. At this point, Owens' authority account should hold that the promise is void: Priam has relinquished his authority to decide whether Prometheus joins him on the trip.

But now suppose Prometheus replies to him 'I am really sorry, Priam. I've just been really selfish. I'll pull myself together, and we should go have some fun. After all, I promised you that we would go'. On Owens' account, Prometheus' last assertion cannot refer to an actual promissory obligation, because it is not there. Nevertheless, it sure seems like Prometheus considers it to be, and it is not clear that he is just confused in doing so. After all, it does not appear that he has *broken* the promise. Perhaps he has not *kept* it either. But he is doing what he promised

---

[12] Owens (2006, pp. 53–4).

[13] My discussion of this case is based on Hallie Liberto's discussion of a similar case, see Liberto (2016, pp. 110–11).

he would. And it is not as if Priam no longer wants him to do it. He does. So it cannot be that the promise is no longer in effect because Priam no longer consents to Prometheus' going on the trip, since that is not true. Unless we assume the authority account, there is no reason to think that there is no promissory obligation left for Prometheus to comply with.

It may seem there is one reason to think so. For it may seem that if Prometheus were still obligated to go on the trip, then it should be appropriate for Priam to blame him for failing to go. And it seems not to be, since Priam suggested himself that Prometheus should not go. But this would imply the absence of a bipolar obligation only if we assume that it *necessarily* comes with a warrant for Priam to blame Prometheus, and this is just the other version of the authority account I have argued we should reject. Blaming others for failing to do what they have promised to, when one has urged them not to, may be a form of hypocrisy. But hypocrisy merely defeats the warrant for blame for the hypocrite. So even if offering 'release' makes it inappropriate for *Priam* to blame Prometheus for failure to perform, blame could still be appropriate for the rest of us. It seems plausible to me that it would be. And that further suggests that Prometheus *is* obligated to Priam to do what he promised to.

As I see it, the problem with these authority based accounts is not that it is implausible that people should have some kind of standing to hold others accountable for their doing what they promise to. The problem arises once we try to specify the directed nature of the obligation in terms of what authority *is*. And as Seanna Shiffrin has suggested, it is not clear that we need an account of what authority is to explain its role in promissory obligation. Instead, Shiffrin suggests that the obligation to keep a promise is derived from the *functional* role that authority plays in

holding people to their promises, a function central to what she calls 'minimally morally decent relationships'.[14]

Shiffrin's argument for this claim is strictly transcendental. The argument is that it would not be possible for people to share such relationships, if they did not have the power to 'bind' themselves to each other by giving promises. And this, Shiffrin claims, is because the point of promising *within* such relationships is to solicit and cultivate relations of trust.[15] This is an interesting idea, not least because it gestures at the type of normative nexus between bipolar obligations, trust, blame, and relationships discussed above. For her part, Shiffrin holds that this type of transcendental argument cannot establish *why* promissory obligations are normative, or what it *is* for such an obligation to obtain. But this point is actually disputed by the more recent *trust*-based account of promissory obligations.

According to Daniel Friedrich and Nicholas Southwood, promises are essentially invitations to trust in one's moral character. On this view, the bipolarity of such an obligation can be explained by the fact that inviting others to trust us puts us in normatively significant relation to them. By issuing such an invitation, one thereby signals one's recognition of the importance of the promised action for the person to whom the invitation is issued. The obligation to keep a promise thus obtains, because we are obligated to those we invite to trust us to deliver on that invitation.[16]

This idea may help explain why promissory obligations can obtain in the absence of the features picked out by the various accounts above. It seems to me plausible that there is sense in which one can 'invite trust' in the absence of established

---

[14] Shiffrin (2008).

[15] Shiffrin (ibid., p. 498).

[16] Friedrich and Southwood (2011).

social practices. As the example above suggests, we may also be able to trust people to do what they have promised to, even after we have granted them leave not to. And, as Friedrich and Southwood argue, one may also accept someone's invitation to trust them without expecting that they will do what one trusts them to do, for example because of weakness of will.[17] Let us grant these claims for the sake of the argument. Even so, the trust account seems to have a shortcoming of its own.

The one big problem for the trust account is its characterisation of the content of the trust that is supposedly invited. As Friedrich and Southwood phrase the idea, it is trust not just in a person's character, but in a person's *moral* character. This seems to me implausible. It is unclear why Priam could not accept a promise from Prometheus, even if he *knew* that Prometheus had a bad moral character. He might come to know this, for instance, by reflecting on the fact that Prometheus had promised to do something immoral, like covering up a crime. Or he might know from Prometheus' track record that he has a bad character, even though the promise he now makes is morally permissible. He might even accept a promise from Prometheus because he thought that this would help Prometheus *develop* a moral character.[18]

For these reasons, I believe none of the above accounts are fully satisfactory. But, as I shall now argue, many of the ideas that motivate them can be incorporated into an account that grounds promissory obligation in the normative significance of the type of relation built into the attitude of trust itself: reliance.

---

[17] Friedrich and Southwood (ibid., pp. 283–4).

[18] See Pettit and McGeer (2017) for an account of how trusting people may reinforce their disposition to do what one trusts them to.

3.   Reliance

According to what I shall call the Reliance View of promissory obligation, when Prometheus' promise to Priam obligates him to do something, this is because a certain reason giving relation obtains between them. This relation is Priam's reliance on Prometheus. On this view, promises indicate a kind of concern. When Prometheus says to Priam 'I promise to ϕ' (or conveys a message to this effect), what he conveys is that he would take any reliance by Priam on him to ϕ to provide him with a weighty reason to ϕ. For Priam to accept that promise is, at least partly, for him to rely on Prometheus to hold this type of attitude. So although people often do rely on the performance of a promise, relying on the content of a promise is not necessarily an effect of accepting it.[19] The Reliance View merely holds that one cannot accept a promise without relying on the promisor at all.

To defend this view, I will discuss some examples that are taken by its critics to offer counterexamples to it. One such example is Scanlon's 'Guilty Secret' case: You are visiting at a university where you know almost no one. At a party one night, you unexpectedly bump into an old acquaintance, Harold. Back in the day, Harold did something of which he is now greatly ashamed and would prefer for no one to know, so when he asks you to keep his secret, you promise that you will. As Scanlon describes the case:

> It seems that Harold cannot rely on the expectation you have created by taking or forgoing any action, because there is nothing he could do to shield himself from the embarrassment that would result from your telling the story. He can't leave town, and I am assuming that murder and bribery are ruled out. He's not that kind of person. So there are no options that he is passing up as

---

[19] This view is thus slightly different from that defended by Neil MacCormick, who argues that promises generate obligations because they *induce* others to rely on its performance. See Mac-Cormick (1972).

the days of your visit go by. But even if there can be no reliance of either of these kinds, Harold has reason to care about whether you act as you have told him you will.[20]

As I understand Scanlon's claim here, the obligation to keep the promise is supposed to be explained by Harold's reason not to want to allow you to break it. But, since it seems Harold has no plans that involve relying on you to keep it, reliance cannot be the source of Harold's reason. Scanlon does not tell us what this other reason might be, but I think the Reliance View has an answer. As the example assumes, Harold has a reason not to want the story to be told. It is plausible that he has this reason *because* he has a further reason not to want to be embarrassed. We all have such reasons. The acute pain of embarrassment is the thought that other people hold certain attitudes towards one; attitudes like ridicule, contempt, and disregard. Such thoughts are painful, because those kinds of attitudes are incompatible with the kinds of relationships we have reasons to want to share with people. These include relationships with friends and partners, but also colleagues, acquaintances. And one can even be too embarrassed to walk down the street for fear of being recognised as 'that person' by strangers.

For most of us, many of our plans and intentions concern these kinds of relationships. Whether we have reasons to revise those plans depends crucially on whether we can maintain the corresponding relationships. For our purposes, it does not matter which type of relationship might be the source of Harold's concern here. What matters is that, *if* Harold thinks that the reasonableness of *any* of his choices depends on whether or not *any* such relationship would be impaired by the telling of the story, then Harold *does* rely on the story not being told. Since this depends on whether you do as you promised, your doing as you promised is

---

[20] Scanlon (1998, p. 303).

also something Harold relies on, on the notion of reliance introduced in chapter 1. Given that reasonable concern with other people's attitudes can make a difference to any of Harold's choices, reliance *can* explain Harold's reason not to want to be embarrassed.

In this way, the Reliance View holds that, whenever we can find a reason for a promisee to want performance of a promise, we can find some minimal bit of reliance on the promisor to perform. In many cases, such as Harold's, we can find such reliance derivatively, by looking at its connection to reliance at the level of attitudes and relationships with other people. And in many such cases, these kinds of reliance relations will obtain independently of the promise itself. What the Reliance View adds is that such reliance does not generate an obligation, unless the promisee also relies on the attitudes of the promisor. This last point is illustrated by another putative counterexample to the reliance account formulated by Hallie Liberto:

> Budd and Bianca are two colleagues who regularly host a visitor in their department. Bianca is planning to pick up the visitor from the airport. However, she is upset that she is always tasked with these jobs, and believes that Budd is not contributing his fair share to departmental service. Budd observes Bianca's distress and its source, and promises Bianca that he will pick up the visitor himself. Bianca happily accepts. She has no faith that he will do as he says. But she hopes that he will, and is prepared to go herself if Budd doesn't come through. Budd knows all of this. Budd successfully recruits a reliable grad student to pick up the visitor from the airport.[21]

Liberto claims that these facts establish a promissory obligation on Budd. Let me list two further points that Liberto cites in support of this claim: (i) Budd has invited Bianca to trust in his good moral character, and (ii) Budd may not delegate the job to a graduate student *because* Budd has invited Bianca to trust that he is

---

[21] Liberto (2016, p. 116).

fair, that he will do his fair share. These facts are important, Liberto further claims, because they determine that, in the context, Budd and Bianca have a 'shared understanding' that the importance of the performance is not due to the fact that Bianca relies on the performance itself.[22]

As we noted above, the Reliance View can grant this point. Even if it is committed to the claim that there is always some reliance on the performance of the promise, it is not committed to the claim that such reliance is *sufficient* to explain why it is 'important' that people do what they promise to. Whence the importance? As Liberto describes the case:

> Something is at stake – Bianca is giving Budd a chance to boost her opinion of him. He knows that Bianca thinks that he is not contributing his fair share, and he is promising to do something that would enhance his otherwise minimal contributions to the department.[23]

We can reasonably assume that this is why Bianca reacts emotionally by being 'frustrated' in response to hearing of Budd's actions.[24] According to Liberto, this shows that the obligation here is explained by the fact that it was important to Bianca that he did not let down the *trust* he had invited her to hold in him: to 'be fair' and 'have a good moral character'. So the shared understanding here seems to be that the point of keeping the promise is to demonstrate that Budd has not forsaken this invited trust. Now I take it that one cannot accept an invitation to trust without trusting. Insofar as accepting the promise amounts to accepting this invitation, it then follows that Bianca trusted Budd to 'be fair' and 'have a good moral character'. Does it follow that she thereby relied on him to?

---

[22] Liberto (ibid., p. 117).
[23] Liberto (ibid., p. 118).
[24] Liberto (ibid., p. 117).

It does on the notion of trust I set out in the previous chapter. Indeed, the fact that Bianca trusts Budd to 'be fair' and 'have a good moral character' implies that she relies on him in these respects on any view according to which trust involves reliance, and, as we noted above, that includes most views. So I hold that Liberto's case actually supports the very claim it is invoked to deny: Budd is obligated to Bianca to do what he promised to do, because her acceptance of the promise involves reliance on him. This is just what the Reliance View says. Accepting a promise constitutively involves reliance on another person at the level of *attitudes*. That is why, to use Liberto's phrase, it matters to Bianca whether Budd does what he promised to.

This point can be illustrated, I believe, by distinguishing between different ways that people might do what they promise to. Consider a kind of case discussed by Judith Thomson.[25] Suppose Prometheus promises Priam to meet him at the store at 6pm and Priam accepts. As we are assuming, Prometheus is now obligated to meet Priam at the store at 6pm. Now, suppose Prometheus never intended to be at the store at 6pm, and was rather hoping to disappoint Priam. However, Prometheus is a fairly incompetent villain. He forgets the whole thing and, setting off for his usual 6pm shopping, he bumps into Priam at the store. While Prometheus did not exactly *break* his promise in this case, it seems he did not exactly *keep* it either.

This distinction between 'doing what one has promised to' and 'keeping a promise' can easily seem laboured or artificial, and the same applies to the supposed distinction between 'failing to do what one has promised to' and 'breaking a promise'. But the point here is not linguistic or conceptual. It seems to me that we should draw this distinction in the same way that we should distinguish between acting wrongly and doing something worthy of blame, or between doing the right thing and doing

---

[25] Thomson (1990, p. 305).

something praiseworthy. I defended these distinctions in chapter 1, and they seem to me equally apt here. To break a promise is not merely to fail to do what one has promised to, but to *inexcusably* fail.

In the case above, since Prometheus did not fail to do what he promised to do, Priam cannot blame him *for breaking the promise*. Nevertheless, it seems that he would have just as much reason as Bianca for holding the types of attitudes that one is disposed to hold in trusting people: feeling let down, disappointed, or hurt. If this is correct, it suggests that what matters in *accepting* a promise is not merely that it be performed. Rather, in accepting Prometheus' promise, Priam relies on Prometheus to hold the kinds of attitudes on the basis of which he *could* keep it.

It may be difficult to see how such attitudes could make any difference to the possibility of keeping a promise. So let me explain how this might work given the claims I have defended. I have argued that one can succeed in giving a promise to someone only if they accept it. They accept it only if they rely on one to hold certain attitudes. On the Reliance View, such reliance provides one with a reason to hold those attitudes. And those attitudes are important, because they affect the status of one's *other* reasons. They affect those reasons precisely in the way that is needed for them to constitute an obligation, on account of the relation between reasons and obligations defended above. So, on this view, it is only by actually taking up those attitudes that one can act *for* the reasons that make it the case that one is obligated to a person. Here is a toy example for illustration.

Suppose only three people exist: Prometheus, Priam, and Paris. Priam relies on Prometheus to make fire, Paris relies on Prometheus not to. On the Reliance View, these facts each provide Prometheus with *a* reason to do each of these things. But because Priam and Paris rely on Prometheus to precisely the same extent, Prometheus

does not have *most* reason to do either thing. But now suppose Priam accepts a promise from Prometheus to make fire, and that Prometheus actually takes up the attitude Priam thereby relies on him to. According to the Reliance View, this attitude *modifies* the reason he already had to do what Priam relied on him to: make fire. It makes this reason *weightier*. It thus makes it the case that Prometheus now has *most* reason, of the reliance-based kind, to do this thing. And those are the kinds of reasons that, according to the Reliance View, make it the case that Prometheus is *obligated* to Priam to do it.

This sort of explanation easily illustrates how it could make sense to trust people to do what they promise to do. On the account of trust I defended in chapter 1, the presuppositions of trust are necessarily satisfied in any case of promissory obligation. If Prometheus has a sincere promise accepted by Priam, it will be true that Priam relies on Prometheus to do what he promised to do, that there is an obligation on Prometheus to do it, and that Prometheus will act on the very reasons that obligate him to do it. For what the promise conveys, this account holds, is that he holds the kinds of attitudes that involve a concern for such reasons; that is, reliance-based reasons. If the promise is sincere, he does hold those attitudes. Of course, the promise may not be sincere. But if Priam has accepted the promise, this would imply that Prometheus would be failing to respond to *other* reasons of the reliance-based kind, e.g. the fact that Priam now relies on him to hold these attitudes.

Like the conventions-based account, this account emphasises the content that a promise conveys, but it does a better job of explaining why the promissory obligation is directed purely to the promisee. Like the expectation-based account, it emphasises the reasons that people have to want to be able to count on its perfor-

mance, but it avoids the problematic claim that acceptance of a promise presupposes the existence of an obligation to keep it. Like the trust-based account, it emphasises the relation between accepting a promise and trusting the promisor, but it avoids the implication that we cannot accept promises from people we take to have a dubious character. Lastly, I believe it can explain what Shiffrin referred to above as the functional role of accountability relations, without relying on problematic claims about what authority is.

This last point can be appreciated by considering two questions that arise for the view I have just outlined. The first is precisely what kind of attitude is doing the modifying work here. The second is how this account is consistent with the intuition that people cannot blamelessly fail to do what they promise to simply by failing to care about their promises. But both of these questions can be answered by considering the type of attitude that might be involved in what Shiffrin referred to above as 'decent' relationships. The most appropriate label for these sorts of attitudes, I believe, is 'care'.

The idea that caring about something modifies one's reason to engage with it in certain ways is intuitive. It seems plausible to me that we all have some reason to preserve valuable works of art, whether we actually care about them or not. But whether or not *you* care about any particular piece of art also seems to make a difference to the weight of *your* reasons to take care of it. It seems intelligible to me that it could make your reasons to do such things weightier than other people's reasons.[26] It also seems to me to explain why the possibility of engaging in promissory obligation is a condition for the possibility of sharing such relationships. It is only in virtue of having such a caring concern for each other's reliance that people can

---

[26] Maguire (2017) provides an extensive defence of the idea that caring about something modifies the weight of one's reasons.

share them. Conversely, breaking a promise, or giving an insincere one, expresses precisely the lack of such concern.

Together, these claims support the idea of a normative nexus between promissory obligations, trust, and relationships, one that in turn helps explain why personal forms of blame should be an appropriate response to a broken promise. At the most general level, care is perhaps best thought of as determinable in the way that 'red' is. Just as there are many determinate shades of red, there are many determinate ways of caring about other people. Still, part of what it is to care about other people, in any way, is to be sensitive to the ways in which they rely on one. The kind of care promises convey is thus of a highly localised sort; it modifies the normative significance of reliance relations in particular domains relevant to the content of the promise. However, the fact that it is more localised is no obstacle to the claim that it has the same normative function as other kinds of care. And, as I shall now argue, this is itself an independently plausible characterisation of the role that care plays within relationships.

4.   Relationships

As in the case of promising, I will begin by setting out some observations about special relationships and the kinds of obligations they involve that I believe a satisfactory account ought to capture. It is relatively uncontroversial that relationships such as friendship, marriage, and other kinds of family relations involve obligations in some sense. Considering one's options to be constrained in the way they are by obligations seems to be a minimal condition on considering oneself party to such a relationship. For instance, it seems that a person who considered himself under no obligation to help you when he could easily do so would be a defective friend at

best. Conversely, if someone with whom you are actually friends refused to help in such cases, it could be appropriate for you to think he had violated an obligation of friendship and was a bad friend for that reason.

These obligations also seem to have the kind of bipolar structure we have been discussing. One's spousal obligations, if one has any, concern one's *own* treatment of one's *own* spouse, not the treatment of spouses as such. One could not, for example, discharge one's obligations of marital fidelity by ensuring that other people are faithful to their spouses. And even if one's own infidelity *would* prevent many others from being unfaithful to their spouses, few would accept that this justifies the violation of one's own obligation. This is part of the explanation for why one's own spouse would have particularly strong reasons to blame such infidelity, and for why this would be precisely the person to whom one would owe an apology, or from whom one could request forgiveness.

This is further reflected in the fact that the appropriate response to violations of obligations within these kinds of relationships can both fall short of, and go beyond, outright moral indignation or resentment.[27] Even when the failure to help a friend does not call for resentment, some kind of reactive response, like feeling hurt or let down, will often be appropriate. And even when more egregious errors call for outright indignation and resentment, these sorts of responses will often *also* be called for. They are not merely optional 'extras' or emotional salt that can add flavour to the experience if one likes. Indeed, it is difficult to understand what it would mean for two people to be friends or lovers, if they were not vulnerable to such attitudes in these cases.[28]

---

[27] Similar claims are defended in Macnamara (2013).

[28] Cf. Kolodny (2010a); Scheffler (2001); Scheffler (2004); Wallace (2012).

It is also widely accepted that trust is central to special relationships.[29] Friends, for instance, tend to rely on each other for various purposes. Indeed, it is difficult to understand what it would mean for two people to be 'friends' if they never relied on each other in any respect. They would never agree to meet anywhere, nor be willing depend on each other for help, advice, or enjoyable conversation. In all but the emptiest sense, friendly relations surely involve these kinds of things.

Lastly, these kinds of relationships tend to come with a requirement of partiality. The extent to which we can reasonably give preferential treatment to those with whom we share relationships is limited, of course. But the limit is there precisely because, within it, such partiality can be both permissible and required.[30] In spending my resources, I have an obligation to be at least somewhat partial to my own children. Conversely, you have an obligation to be at least somewhat partial to your children. And that, it seems, is somehow related to the fact that we each have reasons to care about our own children in ways that we do not have reasons to care about other children.

On what I have called Reliance View, the normative features of these obligations are explained in the same way that those of promissory obligations are. This idea is familiar from extensions of the Rawlsian account discussed above.[31] As an example, consider the fact that the relation of marriage is widely taken to be governed by conventions requiring faithfulness, conscientiousness, etc. and that it is normally institutionalised by the exchange of vows to observe them. The explanation for why those conventions are normative, on these accounts, can thus be considered the

---

[29] This is almost universally agreed in the literature on trust. Cf. Baier (1986); Baker (1987); Cogley (2012); Darwall (2017b); Frost-Arnold (2014); Hawley (2014); Hieronymi (2008); Holton (1994); Jones (1996); Jones (2012); Keren (2014); Lahno (2001); McGeer (2008); Pettit (1995)

[30] This point is a general theme of Samuel Scheffler's work. See Scheffler (2001); Scheffler (2004); Scheffler (2010).

[31] Cf. e.g. the account of filial obligation in Sommers (1986).

same; namely that it would be unfair not to contribute to a conventional scheme from which one has voluntary chosen to benefit.[32]

Precisely for this reason, however, the voluntarist account of these obligations is difficult to sustain. Samuel Scheffler argues that it can be sustained only if one subscribes to the 'fantasy' that we each have supreme control over all of our relations to other people.[33] On the contrary, many of our relations to other people we become aware of only once we realise that their normative requirements are already in force. Scheffler himself argues that this supports the view that there are relationship-dependent reasons, the distinctive strength and content of which depend on our relationship with people. And he defends this claim by appealing to the idea that relationships can be the objects of *valuing* attitudes. It is only insofar as they are valued for their own sake that relationships can be a source of *these* kinds of reasons, Scheffler argues. Thus, he claims that the normative properties of these reasons are similar to reasons that are dependent on our valuing other kinds of objects. Personal projects that shape our lives are another such source, for instance, a source of project-dependent reasons.[34]

While this analogy is interesting, it also raises a difficult question: Why do relationship-dependent reasons seem to establish normative requirements with the surface structure of obligations, when project-dependent reasons do not? People can be criticisable in some sense for wasting their talents, thwarting personal projects they have devoted themselves to, or abandoning them altogether. This is

---

[32] Rawls himself was careful not to overstate the point. In 'A Theory of Justice', for instance, he merely extends the model to obligations of public office, marriage and 'the obligation to play by the rules and be a good sport' in games, cf. Rawls (1971, p. 113).

[33] Scheffler (2001, p. 106).

[34] See Scheffler (2004); Scheffler (2010) for different statements of similarities between valuing projects and relationships.

often unreasonable or regrettable, sometimes even sad. But, as Scheffler himself points out, the type of criticism involved here is not a kind of *complaint*.[35]

It would be a shame, for instance, had Gauguin decided not pursue a career as a painter, but it could not be reason for him to feel remorseful, or for others to blame him. The fact that Gauguin so carelessly neglected his family in pursuit of that career, however, was a different kind of mistake. It seems to be something that *could* give his family a cause for complaint and blame, and Gauguin would not have been confused in feeling remorseful about it. Without determining what Gauguin ultimately had most reason to choose, we can imagine that abandoning his family would have been a substantive mistake, and that abandoning his career would. Either way, the idea that he valued both painting and his personal relations does not explain why the criticism involved in one case should take the character of blame when it should not in the other.

R. Jay Wallace has proposed a view similar to Scheffler's that avoids this type of objection. On this view, we can think of the relationship as a being a source of reasons for us to respond with valuing attitudes to the person we share it with. And, Wallace suggests, once it is granted that relationships can be source of these reasons, it is arbitrary to deny that they can be a source of other reasons, such as those that might ground an obligation to that person.[36] The relationship itself is thus a distinctive normative source of obligations because it gives rise to reasons that are different in kind to those provided by engagement with personal projects, such as reasons to blame people. Nevertheless, saying that relationships are a distinctive source of reasons *because* it gives rise to reasons that have a distinctive character puts the onus on explaining what exactly is distinctive of those reasons. Pointing

---

[35] Scheffler (2010, p. 111).
[36] Wallace (2012, pp. 187–8).

out that it is appropriate to blame people for disregarding them will not do, since this merely returns us to the question: What is it about those reasons that makes this appropriate?

The natural answer is that this is because the reasons in question somehow ground obligations, but, as we saw in chapter 1, it makes a difference precisely how we think of this relation. Wallace suggests that the difference between reasons and obligations is merely a difference in degree.[37] But degree of what? Not simply in weight or strength. They are not simply reasons which dominate or reliably outweigh competing considerations in deliberation.[38] Rather, Wallace suggests, the distinctive feature of these reasons is what he calls their 'deontic structure'. They characteristically shape the deliberation of those who recognise their normative force, and they do this by making other kinds of reasons irrelevant for the purposes of deliberation. In this way, they come to have the deliberative character that obligations do on the account I defended in chapter 1.

When it comes to explaining *why* they have these features, however, Wallace suggests that they are grounded in the value of the relationship.[39] This, it seems to me, simply brings us back to the problem that faces Scheffler's account. Other things than relationships can be valuable, even valuable for their own sake, like some personal projects are. Why should the value of relationships ground reasons that have the character of obligations, when personal projects do not? This question is not answered simply by denying that normativity is fundamentally teleological, or about maximising value. One might argue, as Wallace also does, that it is because such reasons are unlike other reasons in being 'parts of a relational normative structure'.[40]

---

[37] Wallace (ibid., p. 189).
[38] Wallace (2013, 142ff.).
[39] Wallace (2012, pp. 187, 194); Wallace (2013, pp. 157, 163).
[40] Wallace (ibid., p. 156).

But it is not clear how saying this is different from saying that they are reasons that obtain in virtue of the value of relationships.

Simon Keller suggests a view that seems to me to provide a more substantive answer. As Keller points out, the core idea behind Scheffler's proposal is that some reasons come to have the character of obligations for people *because* of the relationships they are party to. We may note that this seems to apply to Wallace's claims as well. This, Keller points out, is nevertheless compatible with the claim that the reason itself obtains independently of the relationship. For instance, the fact that a child is drowning is a reason to help it, for anyone who can. But if that is true, then it follows that to explain why that would be a reason for *you* to help this child, we do not *also* need to appeal to the fact that it is *your* child. Keller thus proposes that we might instead think of relationships as 'intensifiers' that modify the weight of reasons that obtain independently of the relationship itself.[41]

Keller suggests that this modificatory function of relationships might simply be a primitive truth about them. But note that if the account of promissory obligation above is correct, there is no need for an appeal to primitive truths here. If attitudes of care and concern can modify the weight of our reasons in the way that the Reliance View says they can, it is not difficult to see how relationships could modify the weight of independently existing reasons for those who are party to them. Relationships will give rise to obligations, if the following is true.

Relationships: Relationships are constituted by attitudes of care and concern.

---

[41] In fact, Keller seems to be more in favour of thinking of them as 'enablers', as when he suggests that '[t]he fact that you share a relationship with someone is not itself a reason of partiality, but it enables other facts to be reasons of partiality' Keller (2013, pp. 135–6). As a parenthetical point, Keller then mentions the possibility that they might also work as intensifiers, Keller (ibid., p. 136). Because it seems to me misleading to speak of 'reasons of partiality' or 'reasons to be partial', I find the parenthetical point the more plausible version of Keller's proposal.

I believe this is a plausible account of what relationships are, but not in virtue of being a conceptual truth about our notion of a 'relationship'. As with our notions of trust and blame, people can intelligibly use this term in different ways. For instance, like Wallace does, Niko Kolodny holds that rather than being *constituted* by attitudes of care and concern, relationships are precisely what give people *reasons* for these attitudes.[42]

On Kolodny's view, when the actions, attitudes, or reasons of one person are about or involve another, those two people have what he calls an 'encounter'. Each encounter provides reasons to respond to people in certain ways. If one is helped by a stranger, that is a reason to feel gratitude, if the stranger needs one's help, that is a reason to see to his needs, and so on. When their encounters evolve into a temporally extended pattern, two people share a 'history of encounters'. And, Kolodny argues, our attitudes towards such histories of encounters should *resonate* with our attitudes towards the discrete encounters of which they are composed. In other words, the more extensive our patterns of encounters with people are, the more reason we have to hold attitudes towards them that involve recognising the force of reasons to be concerned with their particular interests.

The terminological choice of labelling histories of encounters as relationships, however, strikes me as unhelpful. For one thing, it seems to me fairly common to think of relationships as something that can come and go with time, but the latter point is hard to square with Kolodny's view. Attitudes can come and go, but *histories* of encounters remain. No doubt, the emphasis on the history of encounters can help explain why we cannot easily release ourselves from our obligations. It would be implausible to hold that we could do that simply by walking away from

---

[42] Kolodny (2010a); Kolodny (2010b).

our relationships, mentally or literally. Yet the idea that it should be as difficult for friends to let a friendship go, with all the requirements it involves, as it would be to alter history does not seem to me to illuminate the phenomenon.

Another concern is that 'resonance' is too ephemeral to provide answers to the central normative questions here. As Kolodny acknowledges, it is unlikely to be given any deeper explanation.[43] Thus, even if we grant that something like a history of encounters is the source of reasons for *attitudes* like care and concern, there is still a question about how special obligations get off the ground. Kolodny goes to great lengths to illustrate that the idea of resonance can explain why some relationships provide what he calls 'reasons for partiality' while others do not. Essentially, there are reasons for partiality within a relationship just in case it is constituted by encounters that attitudes of care and concern can resonate with. But even if that claim is true,[44] it is neutral on whether *these* reasons derive from histories of encounters, the attitudes that resonate with them, or something else.

Kolodny also argues that loving or caring about another person, in the way that is characteristic of a valuing relationship, just *is* to see it as a source of such special reasons.[45] But saying that love involves *seeing* something as a reason, or as a source of reasons, does not answer the question of whether love itself actually makes it the case that there is such a reason. Moreover, such appeals to value and valuing merely bring us back to the concerns raised about Scheffler and Wallace's views above. For these reasons, it seems to me that Kolodny's account of relationships leaves open the question concerning the precise relation between reasons, relationships, and the obligations they involve. I believe it is a virtue of RELATIONSHIP that it does

---

[43] Kolodny (2010b, p. 48).
[44] See Keller (2013, pp. 67–9) for criticism on this point.
[45] Kolodny (2003).

not. On this view, we can answer these questions by saying that attitudes of care and concern modify the weight of our reasons to respond to some people's projects and concerns rather than others.

The above account of promissory obligation provides a model for showing how this works. In the case I discussed above, Prometheus' promissory obligation obtains as a consequence of the fact that a sincere promise involves concern for another person's reliance. A promise to help, for instance, expresses a particularly attenuated type of concern for his reliance on one to help, one that modifies the weight of the reason to help. Already at this stage, the Reliance View accommodates a measure of partiality. For it implies that one can thereby have a weightier reason to help one person rather than another, even when they rely on one to help them to the same extent. And that, I suggested, explains why the reliance-based reasons ground an obligation to help one person rather than the other.

It is consequence of this view that, if people do not actually hold such attitudes of care and concern, their reasons are not modified in these ways. But the Reliance View does not thereby imply that it is permissible for those who give insincere promises to break them, or that those who are party to relationships can blamelessly relieve themselves of their special obligations simply by ceasing to care. As I argued, in accepting a promise, one relies on the promisor to have this kind of concern. That, I argued, explains why it can be appropriate to feel saddened or let down when others discharge their promissory obligations for extraneous reasons. Because these attitudes are kinds of blame, they presuppose an obligation to care, and the Reliance View can accommodate this by appealing to additional reliance-based reasons *to care*.

It is because those who give such promises have reasons to care about our reliance on them that there are reasons for us to revise our relationships with them, when they break promises to us. At the practical level, most relationships involve reliance on people to do things. So, if breaking a promise indicates a lack of concern for another person's reliance, it is clear why one would have reasons not to share relationships that involve reliance on a promise breaker, or at least to revise it by modifying one's willingness to rely on him. Doing so is another type of blaming response on the account I defended in chapter 1.

Besides such practical reliance, relationships also involve reliance at the level of attitudes. One can decide to be friendly with people and care about them whatever they think. But one cannot actually be friends with them, I believe, unless they reciprocate those kinds of attitudes.[46] While a decision to develop friendship can be made individually, doing so involves reliance on other people to hold such attitudes towards oneself. They can be encouraged with kind words and gestures, but there is a limit to the extent to which we can reasonably *make* other people care about us in these ways, if we aim to share a decent relationship with them.

On the Reliance View, then, we should expect that such relationships give rise to bipolar obligations because the attitudes they involve modify the weight of people's reliance-based reasons. But this, you will recall, was also the Reliance View's explanation for how it could make sense to trust people. So we should expect that such relationships constitutively involve trust in others to do what one relies on them to do in virtue of that relationship. And then we should also expect that the parties to such relationships are disposed to hold the kinds of personal blaming attitudes towards each other for proving unreliable in those respects. Because that seems to be

---

[46] Cf. Aristotle: 'For goodwill is said to count as friendship only when it is reciprocated' in Aristotle (2005, 1155b).

exactly what we find, the Reliance View thus provides a unified explanation of the independently plausible nexus between these bipolar obligations, trust, blame, and relationships. In the next chapter, we will explore whether this idea can be extended to all other obligations.

Chapter 3

Universality

1.   Challenges Ahead

Pragmatism holds that all obligations are normative in virtue of a particular kind of reason. In chapter 1, I suggested that we could explain the nature of those reasons by looking at the normative presuppositions of a particular kind of 'personal' blame and the type of trusting attitude that disposes to it. A central part of the argument here was that there is a difference between failing to comply with an obligation (or complying with it) and it being appropriate for people to blame one for such a failure (or trusting them to comply). In blaming or trusting a person, we assume that the person towards whom these attitudes are directed is somehow responsible for his actions and this, I argued, is, at least partly, a matter of the attitudes which lead him to act in that way.[1]

The more specific version of Pragmatism I defended in chapter 2, the Reliance View, explained why such attitudes matter. They do so in two distinct, but related, ways. On the one hand, people's attitudes help settle what their obligations are. On the other hand, those attitudes help settle what relationships they share. The

---

[1] I emphasise 'at least partly' here, since claims about the nature of responsibility naturally raise questions about free will which, while important, are outside the scope of the present thesis.

general idea was that when, for instance, John cares about Jane, this modifies the weight of John's reasons to do what Jane relies on him to. Insofar as John is generally responsive to his reasons, he will reliably be motivated do what Jane relies on him to, *because* he will reliably take the fact that she so relies on him to be a reason that outweighs competing reasons; he will take himself to be *obligated*, that is. That is what makes it reasonable for Jane to trust him. But if John inexcusably fails to do this, that is an indication that he did not take those reasons to be weighty, in the way they would be, if he cared about Jane. And that makes it reasonable for Jane to blame him.

My argument so far has been that this provides us with an appealing account of promissory and special obligations. But you will recall that Pragmatism has greater ambitions. It claims that the normative properties of *all* obligations are ultimately explained in the same way. So it claims that if, for instance, you are obligated to help relieve the suffering of someone in front of you, or someone in your neighbourhood, or *anyone anywhere*, the reasons that make it so are also going to be explained by way of the Reliance View. So the question is whether, and how, the Reliance View can extend to these others kinds of obligations too. This is the question we will be addressing in the present chapter.

To do so, it will help to distinguish some of the more specific challenges that this raises. On the Reliance View, we are obligated to treat others in certain ways because we are obligated *to* them, and we are obligated to them partly in virtue of their reliance on us. One problem for this view is that it seems there are obligations concerning the treatment of people whom we cannot be obligated *to*, at least not because of such reliance relations. Coming across a stranger who is drowning in a pond, one plausibly has an obligation to help him out. Yet, if this stranger has no

prior causal or epistemic acquaintance with one, it is unclear how he could hold an attitude like reliance towards one. So the Reliance View needs to explain how we can be obligated *to* people who seem to be in no position to hold attitudes like reliance towards us, in the way that friends and promisees are.

A second challenge arises in cases where it is questionable which individuals, if any, correspond to the respective 'poles' of an obligation. For instance, there are currently more poor people deserving of help than any one person could possibly help. Plausibly, many of us have a so-called 'imperfect' obligation to help at least some of these people. But now consider a person who, while able to help some of these poor people, helps none of them. It is unclear whether any particular person among the poor would be correct to think that *they* (as opposed to others among them) had been wronged by this person. Similarly, it is unclear whether any of these poor people would be relying on *this* particular person (as opposed to others like him) to help them. It may thus seem doubtful that obligations can always be explained in virtue of reliance relations between individual people.

A third challenge concerns the standing of third parties to blame us for violating our obligations. As we have seen, the Reliance View can give a substantive explanation for why Priam, for instance, has reasons to blame Prometheus for flouting a promissory obligation to him. That is how it explains that it is Priam's business what Prometheus does. But it is deeply implausible to hold that *only* Priam has this reason. Surely all of us have *some* reason to blame Prometheus for breaking his promises, even if he only ever gives them to Priam. However, it is unclear how this could be explained on anything like the Reliance View. For since it holds that Prometheus' obligation to Priam is explained in virtue of some kind of relation between *them*, not in virtue of any other relation to other people, it is unclear how

it might account for *other* people's reasons to respond to violations of such obligations.

These are significant challenges to the project of applying the Reliance View to obligations generally. Accordingly, I will take a moment to explain how I think they can be met. As I see it, the central issue here is that it seems the Reliance View has the resources to explain the existence of obligations only *within* existing relationships. And each of these challenges arises, because it is unclear whether our notion of a 'relationship' can be extended as it needs to be for this type of explanation to get going. Nevertheless, I believe there is a way to do so, and my main aim going forward will be to identify and discuss what seem to me the primary sources of resistance to this idea. The claims I shall be defending are not uncontroversial, yet I think there are sound reasons to accept them, and that they leave us with a general account of obligations which is both appealing in its own right and compatible with the Reliance View.

I will first discuss an intuitive way of stating a relationship-based approach to obligations and argue that we should accept a slightly revised version of this view. This revised view is compatible both with the claim that one can impair a relationship simply by holding attitudes, and that one can impair a relationship with *everyone* by doing so.[2] I will then be defending a more specific account of the nature of these attitudes, according to which they are essentially 'second-personal' to use a phrase coined by Stephen Darwall. It falls out of this account that there is a respect in which moral agency constitutively involves reliance on everyone.

---

[2] I am assuming here that 'everyone' includes *at least* all persons. Though I briefly suggest in chapter 5 (page 180, fn. 26) how the present account can be extended to include animals, young children, and those with severely limited cognitive abilities, I shall generally set aside these questions. I do so, not because they seem to me unimportant, but rather because they seem important enough to merit a more thorough treatment than is possible here.

If these arguments are correct, we can extend the explanation offered by the Reliance View to the kinds of obligations that concern our treatment of all people. That gives us an answer to the first challenge. I suggest that if we then combine this insight with those of the previous chapter, we can coherently posit asymmetries between groups of people within the moral relationship. In turn, those asymmetries allow us to connect individuals across these groups to each other via certain reliance relations and thereby explain how even 'imperfect' obligations have bipolar features. That gives us an answer to the second challenge. Finally, I discuss how this account of the moral relationship can explain why there are reasons for us to blame others for violating obligations of any kind. In doing so, they impair the moral relationship itself, a relationship that they have sufficient reason to share with us. That answers the third challenge. These answers do not come cheap, however, so I conclude by defending some of their implications. As we shall see, it is a consequence of it that all of the argumentative pressure for Pragmatism is compressed to a single point: It must demonstrate that there is sufficient reason for everyone to hold the kinds of attitudes that, to use Darwall's phrase, constitute the 'second-person standpoint'.

2.   Relationalism

We can all imagine people who do not share relationships with other people in the standard sense of that term: people who have no friendships, no family, no colleagues, and so on. We can imagine someone being brought up without forming any such standard relationships, who lives in complete solitude, and of whose existence no one else is aware. Such a person would not necessarily be outside the domain of moral evaluation. It would be wrong, for instance, for such a person to kill someone just because he felt like it. Of course, some people are exempted

from blame due to their upbringing, mental disorders, or other circumstances. But being lonely does not entail that one is a psychopath, or subject to any other mental disorder that exempts one from moral responsibility. Such a person could be blameworthy.

The possibility of such cases poses a problem for the Reliance View. For the Reliance View says that, when it is appropriate to blame people for violating an obligation, this is because their attitudes somehow negatively affect a relationship they are party to. But if people who seem to share no relationships with anyone at all can be blameworthy, it seems this explanation cannot be correct. How should the Reliance View address this problem? One option is to point out that there might be other types of relationships than those we normally think of. The Reliance View might hold that there is a notion of 'relationship' according to which:

EVERYONE: There is a relationship everyone shares.

In fact, this view does have its proponents, notably T.M. Scanlon. On Scanlon's view, the blameworthiness of moral wrongs consists in the impairment of a 'moral relationship' that obtains between everyone. According to Scanlon, this relationship obtains between people in virtue of the fact that they stand in the relation of 'fellow rational beings.' And an impairment of it occurs when a person's actions are based on attitudes that are incompatible with the kind of mutual concern for the justifiability of one's actions that this relationship ideally involves. That is how, Scanlon thinks, anyone can be blameworthy for their actions.[3]

On this account, the fact that even strangers can violate obligations is no objection to the claim that obligations are wholly internal to relationships. Thus, it

---

[3] Scanlon (2008, pp. 133–4, 140–1).

is tempting to think that this is the kind of view the Reliance View will need to adopt if the type of explanation of obligations it offers is supposed to generalise appropriately. But Everyone is a controversial claim, and this point is not lost on Scanlon. He thinks it is likely to seem controversial, however, because we are inclined to accept two false claims about relationships. The first is that relationships are constituted by attitudes towards *specific* people. The second is that relationships are constituted by *attitudes* towards specific people.[4]

It seems to me correct to deny the first claim, but it may be hard to see why. It does seem to be a condition for standard relationships like friendship that two people care specifically about each other. And even if we can share relationships with groups of people, it can seem that this could not be a case of holding similar attitudes towards specific people. Football fans, for instance, may have a 'relationship' with their club, but that can seem more like having attitudes towards the club itself, rather than each of the players that make it up. After all, individual players come and go, but, for true fans, the sense of unity with the club remains. Nevertheless, it seems to me that one *can* share a relationship with each member of a group, even if there are no attitudes involved towards anything like a 'club'.

Imagine, for instance, that I come to know via testimony that a group of people have assaulted my friend. 'What a bunch of thugs!' I think to myself. Before I can ask him to elaborate, my messenger disappears. In this case, there is some group of people towards whom I now hold certain attitudes. But it is not clear that this is true only because my attitudes are directed towards some abstract entity that these people constitute. All I know is that this is a bunch of people. So it seems equally apt to say that, for some group of individual people, I now hold certain negative attitudes towards each of them.

---

[4] Scanlon (2008, pp. 139–140).

Now, suppose that my club wins the championship. Thinking of the squad, I find myself thinking 'These guys are just great!' Realising that they could not have done it without the support of fans like me, the players find themselves thinking of us 'These guys are just great!' We all feel a special bond in this moment, one of having 'done it' together, and one we subsequently feel we do not share in quite the same way with other players and fans of the club who were not around at the time. Such cases suggest to me that the attitudes of players and fans towards each other do not have to be mediated via attitudes towards anything like the club itself, although they no doubt often are. It thus seems intelligible that attitudes expressed *de dicto* towards something like a club can be attitudes *de re* towards each individual member of a group of people. The same could be true, I believe, of the kind of attitudes expressed *de dicto* as attitudes towards 'everyone'.

I believe other attitudes could constitute a relationship that everyone would share in this sense, one that would obtain in virtue of every single person holding certain attitudes towards everyone *de re*. And I argued in the previous chapter that more standard kinds of relationships are also constituted by attitudes. I thus believe we should accept:

ATTITUDE: Relationships are constituted by attitudes and dispositions.

Scanlon's second claim above denies this, however. And if we *should* deny ATTITUDE, that would pose a problem for the Reliance View. For then the fact that everyone shared a relationship would not show that they all hold attitudes towards each other. And attitudes are, of course, precisely what the Reliance View needs in order to explain the normative features of obligations in the way it does. So it

is worth looking more closely at the line of reasoning that might lead one to deny Attitude, as Scanlon does.

Here is one such piece of reasoning: It might be conceded that *I* can hold attitudes towards the players of my club, and that some such attitudes *would* partly constitute a relationship with them *if* they held corresponding attitudes towards me. But whereas I can entertain thoughts about each of the players that make up my club at any given time, it is not clear that the converse is true. After all, I have seen these players, and I know their names and many other things about them. But they do not know my name and they have never seen or heard about me. So it seems they are not in any position to think about me at all. And then, it seems, they cannot hold the kinds of attitudes towards me that are required for us to share a relationship.[5]

The intuition that drives this line of reasoning is that one can think *about* an individual or object only if one is somehow *acquainted* with it, either causally or epistemically. It can easily seem curious to think that one can have a meaningful thought about an object if one could not identify any of its features, or if one has never had any kind of causal relation to it, such as having seen it, or touched it, or heard about it from someone else. But both causal and epistemic constraints on reference of these kinds can be questioned.

John Hawthorne and David Manley discuss a number of examples that put pressure on them.[6] One example against causal constraints is based on Urban Leverrier's discovery of Neptune. As they observe, Leverrier had never seen, or touched,

---

[5] This idea also seems to be driving Aristotle's intuitions about friendship: '[M]any have goodwill towards people they have not seen, but suppose to be good or useful; and the same feeling may exist in the other direction. They appear, then, to have goodwill to each other, but how could anyone call them friends when are unaware of their attitude to one another?' Aristotle (2005, p. 1155b).

[6] My discussion of both examples below is based on arguments in Hawthorne and Manley (2012, pp. 32–5, 74–8). The Leverrier example is originally from Kripke (1980, p. 79 fn. 33).

or had any other type of direct causal relation to Neptune, when he postulated its existence on the basis of perturbations in the orbits of other planets, nor had anyone else. Should we really insist on these grounds that Leverrier had no genuine thoughts about Neptune? Suppose he pointed to a mark on a celestial map and said 'This is where the perturber is!' Should we say the he did not really know what he was talking about, or that he was not really talking about anything? This certainly rings false to me.

A different example against epistemic constraints was conceived by Gareth Evans. In this case, a man observes two steel balls in succession. Soon after, he comes to have his memory of the first ball destroyed by amnesia. Evans then has us imagine that, years later, the man starts reminiscing and thinking about 'that shiny ball'. This mental episode is, we suppose, causally connected to his memory of the second ball. Yet (and this is the gist of Evans' example) he cannot bring anything to mind that would distinguish the two balls, and he is not thinking about 'that shiny ball' as being causally connected to the ball that he observed as the second one.[7] Should we say on these grounds that when this man has a thought he would express with 'that shiny ball' he is not really thinking about the second ball, or that he is not thinking about anything at all? Again, this seems to me to be false.

It is a difficult question in the philosophy of language precisely what one should say about these kinds of cases. But this merely shows that the assumptions motivating the objection above are less innocuous than they seem. Insofar as it is intelligible that people like Leverrier and the man in Evans' example can have the relevant thoughts, it is equally intelligible that they should be able to hold attitudes towards the objects of those thoughts. Thus, in the example I gave, the absence of such causal

---

[7] Evans (1982, p. 90).

or epistemic relations does not constitute any decisive objection to the claim that both me and the football players could hold attitudes about or towards each other *de re*.[8] So the idea that relationships must be constituted by attitudes about the people with whom they are shared is no principled objection that everyone could share a relationship in virtue of their attitudes.

A second source of resistance to ATTITUDE comes from a particular way of thinking about the relation between obligations, blame, and relationships. Like Scanlon's view, the Reliance View holds that people's inexcusable violations of their obligations somehow negatively affect their relationships. In a word, the attitudes on the basis of which such violations occur *impair* those relationships. And both of these views hold that such impairment is what explains the propriety of blame. These ideas invite a specific view of what it is to be *blameworthy*.

RELATIONAL: For x to be blameworthy for ϕ-ing is for x's ϕ-ing to be based on attitudes that impair a relationship with some person y.

This type of view can be defended on various grounds.[9] But it is a natural one to hold if one thinks, as Scanlon does, of impairment as a matter of acting on the basis of attitudes which are incompatible with an *ideal* version of some relationship. For that idea seems to presuppose the existence of a relationship that obtains, in its impaired state, independently of such an ideal. A relationship's ideal form also seems to easily provide the type of standard that people can reasonably hold each

---

[8] Admittedly, the examples above do nothing to rule out a disjunctive account of reference, according which one must *either* have an epistemic relation *or* a causal relation to some object to have genuine thoughts about it. Leverrier plausibly had some epistemic acquaintance with Neptune, and Evans' example specifically assumes that the man has a causal relation to the shiny ball. But in the absence of independent motivation for disjunctive accounts of reference, this reply would appear somewhat *ad hoc*.

[9] Ciocchetti (2003); Radzik (2009) both defend views that, in different ways, are very similar to RELATIONAL. See Sher (2013) for discussion of these views.

other to when they do share some relationship. Finally, it seems to sit naturally with the Reliance View's account of the relation between trust and blame.

Scanlon himself emphasises that relation in claiming that a withdrawal of trust is an aspect of blame's practical dimension.[10] More generally, he refers to that dimension as involving a 'revision' in one's relationship with people. It is sometimes objected that this cannot be true because one can blame people while intending to *maintain* a relationship with them.[11] This objection is well-taken, I believe, if we understand the relevant revision as something over and above the withdrawal of trust itself; for instance as a firm decision to end the relationship. But we might also understand 'revision' as a matter of withdrawing trust with respect to more local domains of interactions; ceasing to rely on someone to turn up on time or ask how one feels, for instance. On the notion of trust I defended in chapter 1, there would be no limit to how fine-grained such revisions of relationships could be, because the practical respects in which we can trust people can be as minimal as one likes.

However, if RELATIONAL is correct, this is another source of resistance to ATTITUDE. For suppose the Scanlonian 'moral relationship' *was* constituted by attitudes. Then, if one did not hold these attitudes, one would not share the relationship. If so, there would be no relationship for one to impair, and thus no basis for blameworthiness according to RELATIONAL. So together, RELATIONAL and ATTITUDE seem to imply that moral impunity is objectionably cheap: It is surely implausible that one's actions would be blameless *merely* because one fails to hold certain attitudes. The truth of RELATIONAL would thus count decisively against the plausibility of ATTITUDE. And, as explained, ATTITUDE seems to be precisely what the Reliance View needs for its account of obligations to generalise.

---

[10] See Scanlon (2008, Chap.4); Scanlon (2012); Scanlon (2013).
[11] Wallace (2011); Wolf (2011).

However, as I shall now suggest, Relational is itself open to objections. To see this, note first that it is an identity view. It says that blameworthiness *is* a matter of impairing relationships. This can be true only if the corresponding biconditional is; namely that one is blameworthy if and only if one impairs a relationship. And that claim is vulnerable to counterexamples in both directions.

To begin, consider first a counterexample to the left-right direction. Suppose that, because you do not love your children, you never show them any affection. Still, you might have overwhelming reason to have a loving relationship with your children. But since you do not love your children, you do not have a loving relationship with them. So your attitudes cannot impair *that* kind of relationship. Suppose instead we say that e.g. your children's financial dependence on you constitutes a relationship with them. Even so, your failure to love them is not impairing or incompatible with *that* relationship either. And if your attitudes do not impair any actual relationship with them, then the left-right direction of Relational implies that you are not blameworthy for failing to show your children any affection in this case. This seems to be false.

Here is a counterexample to the right-left direction: Suppose you and a bunch of criminals come to share a relationship by swearing an oath of unconditional loyalty to each other.[12] As you later realise, however, the gang's reprehensible behaviour gives you sufficient reason to tell the police. But to do so would be disloyal, even if it is the right thing to do. You would be acting on the basis of attitudes that are incompatible with this relationship of loyalty. In this case, the right-left direction of Relational implies that you would be blameworthy for telling the police. Again, this seems to be false.

---

[12] Scanlon actually discusses this type of example, but he seems not to think that it constitutes an objection to his view, cf. Scanlon (2012, p. 88).

Now the easy reply to both of these supposed counterexamples is that they simply overlook the distinction between actual and ideal relationships that motivates RELATIONAL. But this reply seems to me *too* easy. The notion of an 'ideal relationship' is helpful only if we have an independent grasp of the type of relationship it is supposed to be an ideal *of*. This defence of RELATIONAL thus needs some independently plausible way to characterise the non-ideal relationships above. One might appeal here to the notion of 'histories of encounters' we discussed in the previous chapter. But that conception of relationships is not helpful either if RELATIONAL is going to be compatible with the notion of a relationship between everyone, because it is not true that all people share a history of encounters. And it is no use if it can do so only in terms of attitudes, since, as we have just seen, the combination of RELATIONAL and ATTITUDE is itself open to serious objections.

Of course, if there were other reasons to reject ATTITUDE, the examples above might simply be said to show that RELATIONAL should not be hostage to any particular account of what relationships are. I have argued, however, that other apparent reasons to reject ATTITUDE are less than compelling. And there seems to me to be an alternative diagnosis of these kinds of cases available: Each direction of RELATIONAL fails because it leaves out a 'sufficient reason' clause. You are blameworthy for failing to love your children, because you have sufficient reason to share a loving relationship with them. And you are not blameworthy for ratting on your fellow gangsters, because you do not have sufficient reason to share a relationship of unconditional loyalty with them. So we might instead hold:

RELATIONAL*: For x to be blameworthy for ϕ-ing is for x's ϕ-ing to be based on attitudes that impair a [possible] relationship with some person y [that there is sufficient reason for x and y to have].

This is only a minor modification of Relational, tying the notion of impairment directly to relationships there are reasons to share, rather than to ideal versions of relationships that are actually shared. The modification is thus slight enough to retain the attractive nexus between trust, blame, and relationships. On this revised view, for others to fail to do what you trust them to will seem to be worthy of blame, insofar as it violates obligations of the kind of relationship that, in trusting them, you took yourself to have *reasons* to share with them. It is also substantive enough to avoid the problems for Relational discussed so far.

The 'sufficient reason' clause allows Relational* to deal with the kinds of cases just discussed. And the 'possible' clause easily allows that anyone can be blameworthy for their treatment of anyone. For if it is possible both to (i) impair relationships one does not actually have and (ii) be blameworthy for doing so, then (iii) not having the attitudes that constitute a relationship will not imply that one cannot impair it. If Relational* is correct, one does not need to reject Attitude to allow that anyone can be blameworthy.[13]

One potential concern with this view, however, is that it may seem to leave the notion of impairment too abstract. It may seem odd to hold that attitudes can impair merely *possible* relationships. Yet I maintain that it makes perfectly good sense, on the Scanlonian idea we started out with; that impairment is a kind of *incompatibility*. Precisely what it is for attitudes to be incompatible will depend on further issues, such as whether they are cognitive or non-cognitive. But whatever it is, incompatibility is plausibly an abstract relation, one that can obtain between actual and possible attitudes. Nothing about this notion of impairment restricts it to relationships that actually obtain.

---

[13] The fact that Scanlon explicitly rejects this claim seems to me strong (though defeasible) evidence that his considered view is Relational, though he occasionally makes claims that suggest a view very close to Relational*. See e.g. Scanlon (2008, p. 128) and Scanlon (2012, p. 88, fn. 5).

Another concern is this: Suppose Alice has sufficient reason to have a relationship with either Bo or Beau, but that she can only have a relationship with one of them. RELATIONAL* seems to imply that no matter who Alice asks on a date, she will be blameworthy. And that is surely implausible.[14] I reply that this seems so only if one overlooks the normative significance of choice in these kinds of situations. For instance, when you have sufficient reason to form one of two incompatible intentions, forming one of them can give you a further reason not to revise it. That is how your choices can make a difference to what you ought to do, and why intentions have a specific kind of normative 'inertia'.[15] Similarly, when you have sufficient reason to have two incompatible relationships, choosing to form one of them can mean that you no longer have sufficient reason to have both of them, because you immediately acquire an extra reason not revise your choice. Indeed, that is particularly plausible if, as RELATIONAL* allows, relationships may be constituted by attitudes, like intentions. So it is compatible with this view that Alice would not be blameworthy in virtue of her choice.

If we accept RELATIONAL*, we can thus defend the claim that anyone can be blameworthy by arguing that there is a relationship everyone has sufficient reason to share. We need not argue that there is any relationship they do, in fact, share. And, unlike the proponent of RELATIONAL, it is open to us to appeal to attitudes in giving an account of what sharing this relationship might consist in.

3. Moral Attitudes

If blameworthiness is a matter of holding attitudes that are incompatible with attitudes that constitute a relationship one has sufficient reason to share, the obvious

---

[14] Thanks to Michael Smith for pressing me on this point.
[15] See Kolodny (2011) for a detailed defence of this claim.

question is what those attitudes might be. What I want to suggest now is that we can identify those attitudes by looking once again at the presuppositions of blaming attitudes.

As we saw in chapter 1, blame involves a number of presuppositions. I mentioned the presupposition that the blamed party had reasons not to act in the way he did, and that he was in a position to respond to these reasons. In this way, it is a kind of rational criticism. But blame is also different from other kinds of rational criticism. Unlike attitudes such as disgust and contempt, blaming attitudes like resentment and indignation involve a kind of complaint or claim that the subject of one's blame could understand and take into account, a claim that he *should* take into account, and one that it therefore makes sense to address *to* him. To hold such attitudes towards another person is to hold him responsible to oneself by addressing him in this way, if only in thought.

As Stephen Darwall puts it, blaming attitudes thereby have a distinctively *second-personal* character, because they presuppose that one has the standing to address reasons, claims, complaints, and demands to the person towards whom they are directed.[16] But the idea that another person has the capacity to understand and respond to such an address seems to presuppose in turn that he has a corresponding standing to return such addresses. It thus appears to be implicit in holding such attitudes that one shares with other people a recognition of the force of such normative considerations and the standing of all persons to address them to each other.

---

[16] Cf. Darwall claims that '[r]esentment is felt as if in response to a violation of a legitimate claim or expectation, and not simply as directed toward the violator, but as implicitly addressing her' Darwall (2006, p. 72) and 'I claim that reactive attitudes are always implicitly second-personal and that they therefore invariably carry presuppositions of second-personal address about the competence and authority of the individuals who are their targets, as well as about those who have them' Darwall (ibid., p. 67).

In taking that perspective on one's own agency and that of others, one takes up a *second-person standpoint*, as Darwall calls it.[17]

I shall have more to say on how precisely I understand this notion of the second-personal below. For now, let me note that this line of thought can be extended to relationships. When the kinds of attitudes through which one recognises the force of other people's reasons are also mutual, they can ground a moral relationship with others in the way that reciprocal care and concern can ground other kinds of relationships, like friendship. This idea can be expressed as follows:

> **MORAL:** For x and y to share the moral relationship is for x and y to hold the attitudes that constitute the second-person standpoint.

To be clear, MORAL does not say that these attitudes must be mutual in the sense that two people must be mutually aware of each other's attitudes, as they typically are in friendship. Admittedly, this makes the moral relationship unusual. But, as I argued above, other kinds of relationships can have this feature too. The moral relationship involves mutuality in a slightly weaker sense. It obtains just in case both parties hold the same kinds of attitudes.

The specifically *moral* feature of this relationship is that these attitudes involve the kind of normative phenomenology that is distinctive of the motivation to do what one thinks is *right*. This is an extension of Scanlon's familiar Contractualist idea: that those of us who are motivated by morality take ourselves to have reasons to share a relationship with others that is constituted by attitudes of recognition. And the link between this thought and the one above is that such attitudes are second-personal. The idea that there should be such a link is not novel.[18] My

---

[17] 'Call *the second-person standpoint* the perspective you and I take up when we make and acknowledge claims on one another's conduct and will' Darwall (ibid., p. 3), emphasis in org.

[18] It is explicitly developed in Darwall (ibid., Chap. 12).

argument here is rather that the aim of sharing that relationship itself involves a specific form of reliance. But we still need an argument for this claim. This is the task I now turn to. The basic idea here is to start from the point raised above, that attitudes like blame are second-personal only insofar as they posit the kind of inter-action that we can summarily call an 'address'. And because such second-personal interactions require reciprocity, they further involve a kind of reliance at the level of attitudes.

Let me first point out a few background assumptions about the notion of an address I am working with here. Addresses are intentional, directed, and successful instances of agency. They are intentional in the sense that they cannot be performed unintentionally. They are directed in the sense that they have intentional objects, i.e. persons. And they are successful in the sense that they obtain only when their 'felicity conditions' do. These felicity conditions involve mutual recognition. Because an address is necessarily intentionally directed towards an individual one considers a person, any attempted address involves a kind of recognition of its object. And to amount to a successful address, the attempted address must itself be recognised by its intentional object. One might, like Xerxes, address a body of water with one's complaints, but this does not suffice for one's address to succeed as a second-personal one. A second-personal address is thus a sort of joint agency to which two parties must contribute by an intentional performance that involves recognition of the other.

'Recognition' is a term of art, and it raises some subtle questions that I have to set aside for now.[19] But doing so will allow us to focus on a more important struc-tural point. Whatever the specific nature of recognition, second-personal address

---

[19] On some views, it involves distinctively second-personal *thoughts*. See Heal (2014), Haase (2014), Longworth (2014), Rödl (2014), and Salje (2017) for this debate. I shall return to the nature of recognition in chapter 5.

requires that it be exchanged, not just that it is mutual or reciprocated. As Darwall emphasises, for a second-personal address to occur with another person

> she and I must be mutually aware of our second-personal *relating*. We must have a second-personal perspective on one another. I must be able to see her as responding (more or less rationally) to my address, which she also regards as an intelligible response or address to her. This is what brings our respective perspectives into the interaction and requires that we be able to take each other's up.[20]

I take it that 'seeing' and 'awareness' are factive terms.[21] One cannot see or be aware of p unless p obtains. So an actual second-personal address requires not merely that two subjects *think* in second-personal terms, or that they *try* to relate to each other. It requires that their thoughts are actually related to each other by going through an interaction with a specific kind of reciprocal structure, one for which the two parties' intentional recognition of each other is mutually dependent. That, I suggest, is what a second-personal address *is*.

An example might help illustrate the point. Suppose I try to address you second-personally, for instance by waving and saying 'Hello'. We can assume that, in doing so, I recognise you as a person with a standing to be greeted by me (whatever we should take 'recognition' to mean here). Now suppose you recognise that this is what I am doing, and you try to address me back, also second-personally. In this case we have mutual recognition. I recognise you as a person, and you recognise me as a person. Yet, so far, the situation still involves an asymmetry. You have recognised *that I recognise you*, but the converse is not true: I have not yet recognised that you recognise me.

---

[20] Darwall (2006, p. 44, emphasis in org.).

[21] On a knowledge-first conception of such states (e.g. Williamson (2000, Chap. 1)) this line of thought further vindicates Thomas Reid's claim about 'social operations of the mind' that such acts 'cannot exist. . . without being known to the other party', Reid (2010, p. 330).

Incidentally, Darwall's preface to the 'Second-Person Standpoint' begins with a mundane example of how your attempted address to me might fail. Suppose you suffer from strabismus, being unable to direct your eyes in the same direction. As a consequence, your attempt to reciprocate my address causes me to believe you are really greeting someone else. In the normal case it will take little to correct the misunderstanding, but we can imagine cases in which the environment somehow ensures that the problem generalises, perhaps because everyone suffers from an extreme strabismus-like condition. In any case, we can imagine cases in which mutual recognition never amounts to an *exchange* of recognition: all attempted reciprocations of addresses misfire. In such cases, should the parties be understood to be *relating* in the way required for a second-personal address to obtain?

I believe the answer to this question is 'No'. For note that a person in this scenario could not be aware that anyone ever recognised him in response to his recognition of them. This matters here, because it brings out that what we care about in engaging in second-personal addresses with others is not merely whether they recognise us. What we care about is whether we can interact with them on this basis, and it is precisely this kind of *interaction* which seems to go missing here. And this seems to me to be driving the further intuition that people in such cases would be missing out on something important; that their aims would go unfulfilled, and that we have reasons to wish to avoid being in their position.[22]

If this is correct, the second-personal disposition we have to blame people and hold them responsible for their conduct is merely one aspect of a much more general disposition; one that we have, because we take ourselves to have reasons to intend to interact with others in this way. There is room for disagreement as to

---

[22] Thomas Nagel similarly argues that sexual interactions lacking a complex of reciprocal awareness of arousal are 'significantly incomplete' in Nagel (1969, pp. 10–12).

what those reasons might be. But this point concerning the structural nature of second-personal address is independent of these issues. However one wishes to cash out the notion of recognition, and whatever the reasons to hold such attitudes might be, the following characterisation should thus be plausible: To 'take up' the second-person standpoint is essentially a matter of intending and being disposed to engage in this kind of recognitional interaction; to do something that constitutively involves reliance on an 'other' to participate.

4. The Argument

This account of the moral relationship and the attitudes that constitute it provides us with the resources to establish an argument for why the kinds of reliance relations which ground obligations can extend to everyone. The main premise of this argument is the claim that moral agents take the fact that an act is morally wrong to imply at least two things. First, that there is sufficient reason for them not to do it. Second, that they would be *prima facie* blameworthy for doing it.[23] The first claim is precisely what a moral theory is standardly supposed to explain. And as I argued in chapter 1, the second is a consequence of the idea that morality is somehow worth caring about, with the emotional and practical dispositions that this kind of attitude involves.

On the view I have defended, these two claims come together in an appealing way. Together, they imply that moral agents take themselves to be blameworthy for committing a moral wrong, precisely because, and only insofar as, their attitudes would thereby impair a relationship they see themselves as having sufficient reason to have. So, on this view, if you are a moral agent, then you take yourself to have

---

[23] The *prima facie* clause is to allow for the possibility of excuses.

sufficient reason to have this relationship. In intending to *be* a moral agent, having this relationship with other people is an *aim* of yours.[24]

According to Moral, this relationship obtains between two people only if *both* of them have these attitudes. This is because the attitudes that constitute it include intentions to do something which requires the intentional participation of both parties. In letting one's agency be governed by the aim of sharing such a relationship, one necessarily relies on other people to have formed an intention to regulate their agency similarly. One relies on *everyone*, everyone who counts in the morally relevant sense, to do so. Thus, one relies on everyone to have taken up the second-person standpoint.

We can use this claim to show how the first challenge I discussed at the beginning of this chapter might be met. The claim here is that each person who has the aim of sharing the moral relationship relies on each other person to recognise each other person. But the kind of recognition central to this aim constitutively involves reliance on the person recognised, Moreover, it is itself a kind of responsiveness to the fact that this person relies on one. It follows from this that acting out of a lack of responsiveness to *any* person's reliance is itself incompatible with an attitude of recognition towards all persons. Accordingly, insofar as there is *anything* that everyone has sufficient reason to rely on everyone else for, it follows that taking up the second-person standpoint commits one to respond to this kind of maximally joint reliance – by letting one's agency be guided whatever principles specify its precise content.

It is not my brief here to defend any particular account of what those principles are. For all I have said so far, these principles may be accounted for within a Con-

---

[24] As I explain below, there may be *persons* who have no such aim. While such people would not qualify as moral agents in the technical sense I am using the term here, this does not imply that we cannot have obligations to them.

tractualist, Consequentialist, or Kantian framework. The crucial point is that this account allows for an explanation of obligations towards all persons that is structurally analogous to the Reliance View's explanation of promissory and special obligations. The basic idea is simple: For those who take up the second-person standpoint, the fact that another person relies on them to do something *is* a reason to respond in that way. So the set of reliance-based reasons that determines which obligations they have *to* all persons is the set of reasons that *everyone* has for relying on each other in certain respects. And precisely *what* they are thereby obligated to do is going to be determined by what everyone's reliance-based reasons weigh in favour of. That is the Reliance View's account of what each of us owes to each other.

Again, the idea here is that second-personal recognition of others involves a type of baseline care or concern for their reliance on one. In particular, it is by being modified or intensified forms of this kind of baseline recognition that the kinds of care and concern at play in promissory and special obligations come to have their modificatory function. This is why, when people rely on us *to care* for them in these modified ways, as those who accept our promises or share our friendships do, this basic kind of second-personal recognition of them provides us with reasons to respond to such reliance. The moral relationship and the obligations it involves to others *as persons* thus permeates the kinds of relationships that give rise to more specific obligations to people. And the Reliance View uniformly cashes this out in terms of a particular species of reasons, the kinds of reliance-based reasons to which attitudes of recognition and concern are constitutively responsive. This, I believe, answers the first challenge mentioned above.

The second challenge concerned obligations which do not easily conform to the Reliance View's bipolar conception of obligations, in particular so-called 'imperfect'

obligations. I gave the example above of the obligation that the wealthy have to alleviate the suffering of the poor. As I noted then, it is hardly clear that there are any reliance relations connecting each and every person in these two categories of people. It is doubtful, for instance, whether any of these poor people would be correct to address any of these rich people with the words '*You* (as opposed to others like you) are obligated to help *me* (as opposed to others like me)'. But we can appeal to the claims just defended to explain how there might be other kinds of reliance relations.

On the present view, the obligations that characterise the moral relationship are those that arise off the back of an aim of interacting, on the basis of certain attitudes, with others who are relevantly like oneself. As far as the moral relationship goes, the group of people who are relevantly 'like oneself' is maximally extensive. Some obligations, however, connect people within much more restricted relationships. The kinds of 'special obligations' we discussed in chapter 2 are a case in point. Nevertheless, it follows from the remarks above that these more restricted relationships, and the obligations they ground, always obtain *within* the moral relationship. And some of these relationships involve normative asymmetries. Parents, for instance, do not rely on their children in the ways that their children rely on them.

Putting these thoughts together, it is possible that there might be obligations that connect *different* restricted groups of people *within* the maximally extensive relationship *in virtue* of such asymmetric reliance relations.[25] It thus seems more apt, I believe, to think of each of the poor as having a standing to address each of the rich by saying that '*You* (and others like you) are obligated to help *me* (or others

---

[25] This possibility was brought to my attention by R. Jay Wallace (forthcoming, Sect. 6.2) whose discussion of imperfect duties I am much indebted to. Wallace, however, takes the notion of a bipolar relation to be basic, not one that can be cashed out in terms of reliance relations, as I am suggesting.

like me)'. That is, we can think of each of the poor as relying on each of the rich to do their individual bit to contribute to a scheme through which the rich (as a group) aid the poor (as a group). This would explain why each of the poor have a particularly good standing to blame each of those among the rich who refuse to contribute such a scheme.

It would also explain why imperfect obligations have bipolar features similar to promissory and special obligations. For example, a lawyer representing the poor might similarly rely on each of the rich to contribute to such a scheme, because he relied on the rich to ensure that the poor have the necessary means to retain his services. But this person would not have the same standing as the poor to blame the rich for failing to do their part in providing aid to them. He would not have the reasons that the poor do to rely on the rich to aid people *like him*.

This brings us to the third challenge, that of explaining why there are reasons for us to blame those whose actions do not affect us directly. To answer it, we can start by noting that we can, at least sometimes, be accountable for our attitudes. For instance, we can be criticisable for having incompatible beliefs, or failing to base them on the available evidence. For an example of this phenomenon closer to our discussion, we can also consider Derek Parfit's case of a gangster who buys a cup of coffee from a coffee seller. I will call this gangster 'Ego' and the coffee seller 'Alter'. To Ego, Alter differs little from a vending machine, something he would be happy to rob or destroy, if that was worth the trouble. But it turns out not to be worth the trouble and Ego simply buys his coffee. Parfit claims that 'though this gangster treats the coffee seller merely as a means, what is wrong is merely his attitude towards this person. In buying his cup of coffee, he does not act wrongly'.[26]

---

[26] Parfit (2011a, p. 216).

Parfit's last claim seems uncontroversially true. Buying a cup of coffee is morally permissible, no matter one's attitudes. The former claim, that Ego's *attitudes* are wrong, is more controversial. Still, it seems to me to gesture at an important truth, namely that we all believe there is sufficient reason for everyone not to have such attitudes. Insofar as this belief is reasonable, it has implications for how it is appropriate for us to respond. It makes it appropriate for us to hold Ego *accountable* for having (or failing to revise) these attitudes, at least insofar as he is not merely ignorant of such reasons, but has in fact *disregarded* them. Discovering that he is like this would call on a range of distinctively *negative* feelings towards him, and rule out many positive ones. And it makes it appropriate to *revise intentions* to interact with him. This suggests that it would be appropriate for all of us to blame someone like Ego.

Of course, precisely what kind of blame is at issue here is a further question. Since their interaction places Alter at greater risk than the rest of us, Ego's attitudes plausibly take on a special significance for Alter. But it is important to recall here that, because they are strangers, Ego's disregard for Alter is merely a consequence of his general disregard for other people. This seems to be something for which it would be correct for anyone to blame him. And *this* fact cannot be explained simply by appeal to the significance of Ego's attitudes for Alter. But, I have argued, it can be explained by the fact that his attitudes impair a relationship he has sufficient reason to have with everyone.

Moreover, if my account of the nature of the attitudes that constitute this relationship is correct, the kind of blaming emotion that corresponds to this impairment is no less a personal reactive attitude than those we refer to as feeling let down, hurt, or betrayed. This claim may seem strange. Strawson, who coined the term 're-

active attitude', clearly distinguished between what he called 'personal' reactive attitudes (such as resentment) and their 'vicarious analogues' (such as indignation). That distinction is often cashed out in terms of a distinction between 'points of view' of the kind often associated with the views of Thomas Nagel.[27]

It seems to me, however, that the distinction between such 'points of view' is an accretion to Strawson's own distinction between the various kinds reactive attitudes, one that may suggest a more fundamental difference between them than he had in mind. I suggest that we may instead see 'personal' and 'vicarious' reactive attitudes as picking out opposite poles of a continuum of reactive attitudes that presuppose more or less abstract forms of reliance. So, while I agree that we are apt to refer to our attitude towards someone like Ego as 'indignation', my aim has merely been to illustrate that this does not fundamentally set it apart from other kinds of blaming attitudes. It presupposes reliance all the same.

5.   An Objection

These claims demonstrate that the Reliance View can meet the three challenges discussed at the beginning of this chapter. There is, however, one further aspect of this view which remains to be discussed. For I have argued that there is a relationship that everyone has reason to share, that everyone has reasons to blame others for impairing it, and that this is because it constitutively involves reliance. Yet I also argued in chapter 1 that such reliance is constitutive of our relationships only insofar as it amounts to the rich kind that disposes us to blame: trust. Taken together, these claims imply that the type of relationship we have sufficient reason to share

---

[27] Darwall, for instance, readily distinguishes between the point of view of 'the transagents' and the point of view of 'the moral community', taking the two classes of reactive attitudes to map onto this distinction. See Darwall (2006, pp. 66–67).

with everyone constitutively involves trust. They seem to imply that it is reasonable for us to trust everyone. And this may seem incredible.

In fact, I believe they imply a slightly different view. As I argued above, even those who are genuinely party to a relationship may have reasons to revise their attitudes on account of each other's behaviour, for instance by ceasing to trust each other in certain respects. Thus understood, the account merely implies that trust in everyone is reasonable *unless* such additional reasons come into play. In other words, the claim implied here is:

> **Universal:** For moral agents, trust is a reasonable default attitude towards everyone.

However, Universal is still open to a fairly fundamental objection. Consider a type of 'state of nature' case. Let us assume that you have taken up the second-person standpoint, but that you are oblivious to the existence of anyone else. Now you come across Ego, and you immediately discover that he is the kind of person who would kill someone on a whim. It is plausible that your attitudes towards him would, quite reasonably, not include trust. But then, the objection goes, trust cannot not be your 'default' attitude towards him, let alone a reasonable one. So Universal must be false. If it is, then any view that implies it must be too. And then Pragmatism is in trouble.

This seems to me a powerful objection insofar as it shows that it is reasonable to *withhold* trust from Ego, by refraining from extending it in the first place. I suspect this thought tracks the intuitions of many. But there is a good explanation for why that is. When we say that something is 'reasonable', we usually mean to say that the reasons 'for' outweigh the reasons 'against'. In the present case, it seems clear that

you have a reason for not trusting: you know Ego is the kind of person who would kill someone on a whim. But it does not seem like there is *any* reason pulling in the opposite direction. And since you knew nothing about Ego before, it can seem that you never had any such reason.

So it is not surprising why it seems that the reasonable response is to withhold trust. If we cannot appeal to any reasons to explain why you would already trust Ego, it is not very charitable to think that you trusted him anyway. The charitable thought is that you did not trust him. And if you go from this state to having a reason not to trust, then the reasonable response must be to withhold trust. Right? Of course it is! But the crux of the argument here is the intuition that there is no reason pulling in the other direction, indeed that there never was one. And intuitions about reasons can be misleading.

The phenomenon of defeat provides a particularly interesting example of this. Here is one: That Tweety is a bird is a reason to think Tweety flies. That Tweety is a penguin is a reason to think Tweety does not fly. Given both these facts, the reasonable conclusion for you to draw is that Tweety does not fly.[28] Now, the standard explanation for how one reason wins out against another to make some conclusion reasonable is that one 'outweighs' the other. In the present case, however, this explanation fails to accommodate another fact: Once you know that Tweety is a penguin, the fact that she is bird does not seem to count *at all* against the conclusion that she does not fly.

A better way to explain this is in terms of defeat: Separately, each of these facts is a reason; each counts in favour of a conclusion. But *given* that Tweety is a penguin, that she is also a bird is no reason to think that she flies. The former fact defeats the

---

[28] I take this example from Horty (2012) who provides a formal framework for defeat.

latter by undercutting its status as a reason. This phenomenon is important. For it explains why, sometimes, it seems there is no reason. It is because something which otherwise is a reason has been defeated by another reason.

Something similar could explain our intuitions about Ego. That he is the kind of person who would kill someone on a whim is a reason not to trust him at all. If it were also a defeater for something which otherwise *is* a reason, it would seem intuitive that there is no reason pulling in the other direction, and that there never was one. And this would make it seem intuitive that the reasonable response is to withhold trust rather than withdraw it. But there would be a default reason to trust all the same.

On some views, such a reason can be derived directly from the fact that we have reasons to share the moral relationship with everyone. On one view, this is because the fact that reliance on a person is a means to a reasonable end is itself sufficient for such reliance to be reasonable, at least in the absence of conflicting evidence.[29] On a different view, it is because the fact that reliance on a person is a means to a reasonable end can make such reliance reasonable, irrespective of any evidential considerations.[30] But this way of deriving our reasons to hold default attitudes of trust seems to me to put the phenomenon of defeat in the wrong place.

When we discover that a reason to pursue some end is defeated, what we discover is usually that this end is no longer worth aiming for in the way we thought it was. This does not seem to me to be the appropriate response upon learning about Ego's attitudes. On the contrary, the aim of sharing the moral relationship with him remains valuable, in spite of his attitudes. This, it seems to me, is why the withdrawal of trust that is appropriate should be accompanied by a sense of *regret*.

---

[29] Longworth (2017).
[30] Alonso (2016).

If the reason defeated here were a reason to share the moral relationship, it is not clear why this should be our reaction. So let me illustrate how the account I have defended allows for a better explanation on this point.

On the broadly second-personal framework I have advocated, the concept of a person may itself be conceived of second-personally.[31] A person is someone with whom one can engage in second-personal address. This claim gives the second-person theorist a simple story to tell about the reason we are looking for. On this type of view, the fact that an individual is a person is a reason to think that they take up the second-person standpoint. If that is all you know about them, it is reasonable to conclude that they do. If they do, and if you do yourself, then they are one of the people with whom you share the moral relationship. And if they are, then they are one of the people that you rely on to govern their agency on the basis of an aim to share that relationship, and whom you are disposed to blame if they do not. So they are someone that you trust to do something.

Of course, we need some sort of argument as to *why* the fact that someone is a person would be a reason to think that they take up the second-person standpoint. I shall suggest one momentarily. Note first, however, that *if* such an argument can be supplied, the conclusion it supports would help explain why the phenomenon of undercutting defeat works in the way it does here. Consider first an analogy with our original case. If I *first* tell you that Tweety is a penguin, what I have told you is a reason to think that Tweety does not fly. But if you know what penguins are, then you also know that Tweety is a bird. That fact is a reason to think Tweety flies. However, since the former reason undercuts the latter, the reasonable conclusion is

---

[31] Cf. Darwall (2006, p. 126), Fichte (2000, p. 45), Haase (2014), and Locke (1975, p. 346) for different versions of this view.

that Tweety does not fly.[32]

Similarly, suppose I tell you that Ego is the kind of person who would kill someone on a whim. Plausibly, what I have told you is a reason not to trust him at all. But, on the present view, if you know what persons are, then you can infer from what I have just told you that Ego has the capacity for second-personal interaction. That fact is a reason to think that he takes up the second-person standpoint. However, since the former reason defeats the latter, the reasonable response is to withdraw trust. And this should be accompanied by a sense of regret, not because a moral relationship with this person is not valuable, but rather because it is not possible given the available facts: that Ego is clearly failing to exercise these capacities.

Of course, if this reason is immediately undercut, it can seem there is no rationale for describing the process here as a *withdrawal* of trust. It seems one could never respond to the reason to trust before it is undercut. But here one should recall that it is not a restriction on the attitudes which constitute the moral relationship that one must have some prior *acquaintance* with people for those attitudes to concern them. Even before you become *acquainted* with Ego, he is one of those generic 'others' that you have reason to direct your second-personal attitudes towards, some of which it may thus be appropriate to revise or withdraw.

It may also seem like there is some kind of illicit bootstrapping going on. The argument here asserts that the fact that something is a person is a reason to think that it takes up the second-person standpoint. But it also asserts that to be a person

---

[32] I am not claiming that this is the only way to make sense of the reasons in play here. For instance, one might wish to build the absence of the defeating condition into the reason itself. The problem with this idea is that it quickly seems to lead to the view that the absence of *any* possible defeating condition must be built into the reason itself. That view is intelligible (cf. Nagel's conception of reasons as modally sufficient conditions at fn. 39 on p. 40), but it surely is to abandon the idea that reasons are facts with separable normative force. See Bader (2016) for a defence of the separability of reasons in light of the possibility of defeaters and modifiers.

is to have the capacity to take up the second-person standpoint. So it may seem to assert that the fact that something has a capacity is a reason to think that it exercises it. This claim would clearly be implausible. Human beings also have the capacity to kill each other on a whim, so this claim would imply that there would be an equally good reason to think that they do that. How, then, could it be reasonable to think that other people exercise their second-personal capacities *rather* than killing each other on a whim?

This seems to me a plausible answer: Given that moral agents take themselves to have sufficient reason to share the moral relationship with everyone, they take everyone else to have those reasons too. As we can plausibly assume here, they do not take everyone to have sufficient reason to kill other people on a whim. If being a person is to have the capacity to respond to reasons to hold second-personal attitudes, these facts together seem to provide at least *a* reason to think that persons hold those attitudes rather than not. I see no reason to claim that those facts together do not count in favour *at all* of thinking that an individual holds second-personal attitudes. And if they do count in favour of it, then they do amount to a reason to think so.

Of course, one cannot deductively *infer* from these facts that people hold those attitudes, but the reason relation is much weaker than deductive inference. To defend this argument, it does not matter how weak that relation is. All the argument needs is that there is *some* reason. For if you only know about an individual that it is a person, you have *a* reason to think that they take up the second-person standpoint. If that is all you know about them, you have *no* reason to think they do not. So it is reasonable for you to think that they do. And then, by the argument above, it is reasonable for you to trust them to do so. Of course, knowledge which can

defeat this reason may be easy to come by. It may be true that most people cannot, in actual fact, reasonably trust everyone. But the argument here is compatible with this claim.

As these claims illustrate, Pragmatism can extend all aspects of the Reliance View to any single individual who is a person. As you will have noted, however, it can do so only by consistently appealing to a particular claim. This is the claim that there is sufficient reason, indeed most reason, for everyone who is a person to take up the second-person standpoint. So I conclude that *if* there are such reasons, Pragmatism can successfully explain the normative properties of all obligations in virtue of the bipolar obligations that are grounded in reliance-based reasons. It remains to be seen, however, whether there are any such reasons.

Chapter 4

Reasons and Recognition

1. Directions

In the previous chapter, I argued that the Reliance View can generalise to all obligations. That is, for any obligation, we can explain why that obligation obtains in virtue of a certain species of normative reasons – facts about how other people rely on us. But that conclusion was strictly conditional. It relied on the assumption that *everyone* has sufficient reason to hold the attitudes that amount to 'taking up' the second-person standpoint, the attitudes in virtue of which a certain moral relationship between people obtains. Now, the Reliance View's account of the moral relationship could provide a simple explanation of this: For any person, they are relied on to do so by those who are party to the moral relationship. The problem, of course, is that this explanation simply assumes that facts about other people's reliance are reason-giving in the first place. How might we defend this claim?

In this chapter, I argue that answering this question poses a particularly intricate problem for those who subscribe to the more basic normative commitments of Pragmatism. As you will recall from chapter 1, Pragmatism subscribes to a programmatic view in philosophy that is sometimes referred to as the *reasons-first* view.

According to this view, *all* normative properties obtain in virtue of reasons. Pragmatism's claim that the property of being obligated to do something is a normative property *because* there are reasons not to do it is an instance of this more general schema. As we have seen, Pragmatism defends this view by giving a more detailed analysis of this property, in terms of reasons to have specific kinds of second-personal attitudes. But there is a well-known problem about analysing normative properties in terms of reasons for attitudes. Philosophers call this the Wrong Kind of Reason problem.

In Section 2 below, I explain in further detail what this problem is about, and discuss three standard responses to it. As we shall see, one of these responses is difficult to motivate on any view, and two of them are unavailable to those who subscribe to the reasons-first view, such as proponents of Pragmatism. This has lead many to believe that the Wrong Kind of Reason problem cannot be solved if one assumes the reasons-first view. So I move on to explore how we might account for reasons to hold second-personal attitudes in the absence of such a solution. I do this by discussing argumentative strategies familiar from three different moral theories: Contractualism in Section 3, Kantianism in Section 4, and Consequentialism in Section 5. And my argument will be that each of them lead to some problematic implications, in particular for proponents of the Reliance View.

If we cannot solve the Wrong Kind of Reason problem, we may have to accept at least some of these conclusions. Conversely, if we find these conclusions unacceptable, we need to show that the Wrong Kind of Reason problem can be solved. This latter option is the one we will be exploring in chapter 5. There, I argue that we can solve the Wrong Kind of Reason problem without giving up the reasons-first view. And that solution provides some strong independent support for the claim that the

Reliance View has been assuming: It is correct to take the fact that another person relies on you to be a reason to do it. If the arguments of chapter 5 are sound, it is correct because there are non-second-personal reasons to hold second-personal attitudes. And in chapter 6 I defend an account of what those reasons are.

2. The Problem

I hope this provides you with some sense of where these arguments are going. Nevertheless, it will be easier to see how they lead us there if we say a bit more about the reasons-first view. This view combines two key claims. The first we can call *reasons fundamentalism*. Fundamentalists hold that all normative properties obtain in virtue of reasons. For instance, suppose miserliness is prudent. According to reasons fundamentalism, if the property of being prudent is also a *normative* property of miserliness, this is because there are reasons to be prudent. The second claim we can call *reasons unity*. This is the idea that there is something it is to be a reason. Reasons are thus to be contrasted with, say, pieces of jade which can be what they are either in virtue of being made of the mineral 'nephrite', or in virtue of being made of the fundamentally different mineral 'jadeite'. In contrast, unionists hold that there is just a single 'counting in favour of' relation in virtue of which all reasons are what they are.[1]

Many proponents of the reasons-first view analyse normative properties in terms of reasons for attitudes. An example of this is the much discussed 'buck-passing' view of value, according to which the property of being valuable is a normative

---

[1] A picky point: It is important not to confuse reasons unity with its related cousin that we might call *reasons primitivism*. Primitivists hold that 'reason' cannot be analysed in other normative terms, nor reduced to any other non-normative property. On this view, reasons are both normatively *and* metaphysically fundamental.

property, but not a normatively fundamental property. Rather, the fact that something is valuable implies that it has other properties that provide reasons for people to hold valuing attitudes towards it. On the buck-passing view, having such properties just *is* what it is for something to be valuable.[2]

By now, it is generally recognised that the combination of *reasons fundamentalism* and *reasons unity* presents a serious difficulty for this type of analysis. It has been widely discussed with respect to the buck-passing account of value,[3] but one can equally demonstrate this problem with the property of being laughable, or enjoyable, or blameworthy, or praiseworthy, or any other putatively normative property corresponding to an attitude verb of English.[4] To show this, I will take the property of being admirable as my example, assuming for the sake of the argument that it is a normative property.

Suppose Mary is admirable. A buck-passer about admirability must now say that this is because there are reasons to admire Mary. Let us suppose that Mary does in fact have properties that count in favour of admiring her. She is both steadfast, considerate, and wise. And since she is neither cruel, dishonest, or the like, she has no properties that count against admiring her. So the buck-passing conception of admirability now predicts that Mary is admirable. That sounds right. But now suppose an evil demon threatens to destroy the whole world unless people admire him. That counts in favour of admiring him. Moreover, we can easily imagine that there are no reasons not to admire him that outweigh those reasons. Then buck-passing about admirability predicts that the evil demon is admirable too. That does not sound right.

---

[2] Cf. Scanlon (1998, p. 95f).

[3] Rabinowicz and Rønnow-Rasmussen (2004); Rabinowicz and Rønnow-Rasmussen (2006); Olson (2004); Olson (2009); Lang (2008); Brunero (2010); Samuelsson (2013); Rowland (2013).

[4] Indeed, one of the earliest discussions of this type of problem was applied to the properties like being funny, shameful, pitiable, etc. in D'Arms and Jacobson (2000b).

Intuitively, this seems to come down to a difference between the reason to admire him and the reason to admire Mary. Unfortunately, the combination of reasons fundamentalism and reasons unity makes it difficult to explain what the difference is. If you accept reasons fundamentalism, you cannot explain the difference in normative terms, since you accept that all normative terms are ultimately cashed out in terms of reasons. If you accept reasons unity, you cannot explain it by saying that there are two different ways something can 'count in favour of' admiration. For you accept that there is only one way that it can, the way that reasons do. And if you have nothing else to say, it seems you are stuck with the claim that the evil demon is admirable. Since that appears to be false, it appears to be a problem for anyone who thinks that the reasons-first view can explain what it is for admirability to be normative in this way. It is what has come to be known as the Wrong Kind of Reasons problem.

That problem can easily be generated for the view I defended in chapter 3. On this view, the property of being 'obligatory' can be analysed in terms of reasons for second-personal attitudes like care, concern, and recognition. Of course, there does not seem to be any attitude-term in English that corresponds to the property of being 'obligatory' in the way that 'envy' corresponds to 'enviable', but this should not distract us. Just as we can imagine an evil demon who threatens us to value something, we can imagine an evil demon who threatens us to hold the kinds of attitude that modifies the weight of our reliance-based reasons in the way that grounds obligations. By parity of reasoning, such threats could provide reasons for admiration only if they could also provide reasons for these other kinds of attitudes. But just as an evil demon cannot make admirable that which is not, he cannot make obligatory that which is not. So it seems a proponent of Pragmatism needs a distinction

between right and wrong kinds of reasons for such attitudes, and it is not clear how he can get it if he accepts the reasons-first view.

The standard approaches to dealing with the Wrong Kind of Reason problem can be divided into three types. One holds that the difference between between the reason to admire Mary and the reason to admire the evil demon is this: The fact that counts in favour of admiring Mary makes it *fitting* or *appropriate* to admire Mary, while the fact that counts in favour of admiring the evil demon does not make it fitting or appropriate to admire him.[5] But this view solves the problem about accounting for the normative property of admirability only if we accept that the property of being 'fitting' or 'appropriate' cannot itself be analysed in terms of reasons, otherwise the problem simply recurs. And since accepting that is to give up reasons fundamentalism, it is to give up the reasons-first view.

A second type of response is to hold that there are simply two ways that facts can provide normative support for attitudes. As Christopher Howard has suggested, it is open to a proponent of the fittingness view to hold that the evil demon's threat makes some *other* attitude fitting. For instance, it might make it fitting to *want* to admire the evil demon, even though it does not make it fitting to *admire* him. In principle, that conception of normative support can be cashed out in terms of reasons. But it would nevertheless seem to imply that the correct account of reasons is disjunctive: Something is a reason for an attitude either if it counts in favour of having that attitude, *or* if it counts in favour of wanting to have that attitude.[6] And since accepting such disjunctivism is to give up reasons unity, it is to give up the reasons-first view.

---

[5] See McHugh and Way (2016); Chappell (2012) for some recent statements of this idea.
[6] Howard (forthcoming).

Because Pragmatism is committed to the reasons-first view, it cannot rely on these responses. There is, however, a third type of response. One may simply deny that the evil demon's threat provides a reason to admire him in the first place. If the reason is not there, there cannot be a problem of explaining why it does not ground a normative property. Thus, some argue that the fact that the evil demon will destroy the whole world unless you admire him is not a really to admire him, but rather a reason to *want* to admire him or to *try* to admire him.[7] The only real reason for admiration is the one you have to admire Mary, and then there is no problem in saying that Mary is admirable because there are reasons to admire her.

This reply does not require a commitment to disjunctivism about reasons. It does, however, require a commitment to the claim that there is not really a reason to admire the evil demon. This claim strikes many as implausible,[8] but I will not press the point here. In any case, I believe we can better appreciate why this reply is unpersuasive once we appreciate the more general scope of the problem here.

Consider the normative property of being 'wrong'. There is a group of moral views according to which an action is wrong just when it violates a principle that no one could reasonably reject. How do we decide which principles are 'reasonably rejectable'? Well, we look at the reasons for rejecting it and the reasons for accepting it and see how they weigh out. Which reasons? On one view, all the reasons there are. For instance, suppose John has a reason to reject some principle, whereas Mary

---

[7] Skorupski (2010, pp. 87–8); Parfit (2011a, p. 51, Appendix A).

[8] Suppose you had a reason to want to admire the evil demon, but not a reason to admire him. As noted in the literature, this raises three problematic questions. First, why would you have a reason in the first place to want to have something that you do not have a reason to have? (Rabinowicz and Rønnow-Rasmussen (2004, p. 412)). Second, how could wanting to have an attitude without having it be the most reasonable configuration of your attitudes, as the view seems to imply in the present case? (D'Arms and Jacobson (2000a, pp. 743–4)). Third, why would you have a reason merely to want to have this attitude when merely *wanting* to have it will not avert the demon's threat? (Rabinowicz and Rønnow-Rasmussen (2004, p. 413); Danielsson and Olson (2007, p. 514)).

and Molly each have a reason to accept it. As it happens, neither Mary's nor Molly's reason is sufficiently weighty *individually* to outweigh John's reason. But according to so-called *aggregationists*, we can aggregate Mary and Molly's reasons. And it just so happens in this particular case that Mary and Molly's individual reasons aggregate such that they *together* outweigh John's reason. According to aggregationists, that makes it the case that this particular principle is not 'reasonably rejectable'.

According to proponents of a different view, however, this is false. These *non-aggregationists* hold that the only reasons which can make it reasonable to reject some principle are the reasons of individuals. Thus, non-aggregationists hold that Mary's reason and Molly's reason cannot aggregate such that they together outweigh John's reason. This is not because non-aggregationists reject the notion of aggregation altogether. Some of them are happy to allow, for instance, that reasons can aggregate for the individual person whose reasons they are.[9] They merely hold that reasons cannot aggregate across persons. But it is widely debated whether non-aggregationists can provide any plausible explanation for restricting the set of reasons that make a principle reasonably rejectable to the reasons of individuals.[10] If John's reasons can add up to make it reasonable for him to reject a certain principle, why can Mary and Molly's reasons not add up to make it reasonable *for them together* to reject it?

This discussion between aggregationists and non-aggregationists seems to me to be another instance of the more general problem that arises for those who wish to analyse normative properties in terms of reasons. It arises when it seems like, of all the many reasons there are, only some of these are relevant to determining whether

---

[9] Scanlon (1998, p. 237).
[10] See e.g. Reibetanz (1998); Parfit (2011b, Chap. 21); Horton (2017) for various points of critique.

that property obtains. And the reasons-first view therefore needs a general rationale for determining the kinds of reasons that help settle different correctness standards.

Think about it like this: Who would be admirable on a reasons-based analysis of admiration? Those there are reasons (of the right kind) to admire, of course. But there might also be reasons (of the right kind) not to admire them. And it would clearly be a mistake to admire them if the former outweighed the latter. So the admirable must be those whom it would be *correct* to admire, and these would be the people whom there is more reason (of the right kind) to admire than not. The general schema here is simple: Standards of correctness for ϕ-ing are settled by weighing reasons of the right kind for and against ϕ-ing.[11]

That schema generalises to other putatively normative properties. Something is valuable just when it is correct for people to value it, and that is correct just in case there is more reason of the right kind to value it than not. Something is credible just in case it is correct to believe it, and that is correct just in case there is more reason of the right kind to believe it than not. And someone is blameworthy just in case it is correct for people to blame them, and that is correct just in case there is more reason of the right kind to blame them than not. In general:

CORRECT: It is correct for x to ϕ if and only if reasons of the right kind for x to ϕ outweigh reasons of the right kind for x to not-ϕ.

As should be clear, the correctness standards specified by CORRECT are all relative to specific sets of reasons. Correctness standards for blame are relative to reasons of the right kind for blame, while correctness standards for value are relative

---

[11] This idea was first developed by Mark Schroeder in Schroeder (2007, Chap. 7); Schroeder (2010).

to reasons of the right kind for value, and so on. The point here is just that the problem arises for other correctness standards than those which apply to attitudes. The standard of being morally wrong is at the very least *also* a correctness standard for actions.

Now, consider how a reasons-first non-aggregationist might respond to the charge that he has no independent rationale for holding that aggregative reasons do not help determine which acts are wrong.[12] Suppose he takes the line of those who deny that there is a reason to admire the evil demon. He holds that the reasons of groups are not *really* reasons to reject principles. At best, they are reasons for them to *want* to reject such principles. Indeed, he might argue that since groups can have attitudes, they cannot have reasons to want anything. He might even deny that groups can have or respond to reasons of any kind.[13] But none of these responses would help him.

As the aggregationist can plausibly claim, even if *groups* cannot have reasons to reject principles, an individual can have reasons to reject a principle, because it would be worse for some group if everyone accepted it.[14] Surely individual people

---

[12] Scanlon himself describes the discussion between his own non-aggregationist version of Contractualism and Parfit's alternative aggregationist version in these terms. These views are in agreement that an action is wrong just in case it is disallowed by a principle that no one could reasonably reject. But they disagree which reasons are of the right kind to make this the case. As Scanlon notes, the idea that reasons deriving from benefits to groups or the value of objects factor into the 'reasonable rejection' property is the fundamental difference between these two views. Scanlon says 'These two differences may be seen as improvements over the view stated in [Scanlon (1998)], which seemed implausible to many because it excluded aggregative arguments and because it gave no weight to impersonal values in determining what is right or wrong. These objections could be dealt with by allowing *reasons of these two kinds to be considered in determining whether a principle could be reasonably rejected*'. He then goes on to note that 'Many people may be drawn to consequentialism because they see that there are some situations in which the morally correct way to decide what to do is to figure out what would produce the best consequences overall... This seems to me to be a mistake: producing the best consequences might be *the correct standard* in these cases not because it is the basis of morality but because it is what is owed to people in situations of that kind, by agents who stand in a certain relation to them'. See Scanlon (2011, pp. 133–4, 139), my emphasis.

[13] For arguments against this view, see Dietz (2016).

[14] Cf. Parfit (2011b, pp. 210–11).

can have reasons to *want* groups of people to be better off. It would thus be a mistake for the aggregationist to accept that reasons aggregate in the sense that groups can 'have' or 'own' or 'act' on them. Instead, he should hold that they aggregate merely in the sense that people can come to correct conclusions about their aggregate weight. He should hold that these reasons aggregate *for the purpose of determining what it is right or wrong to do.* In other words, he should hold that they aggregate in the way that CORRECT says they do. And then the question for the non-aggregationist remains: Why should these reasons together not help determine what it is right or wrong to do?

This debate seems to me to illustrate three interrelated points. First, that the problem facing proponents of the reasons-first view is much more general than it seems. Second, that this generality makes it much less plausible to deny that reasons of the wrong kind are really reasons. Finally, it illustrates that the Wrong Kind of Reason problem is basically a problem about distinguishing between the reasons that factor into specific standards of correctness from the reasons that do not. The correctness standard which fixes the normative property of being 'wrong' or 'obligatory' is one of these. And this suggests that the problem arises for *any* reasons-first approach to obligation, not just Pragmatism.

In one way, the generality of this problem is good news. For if there were a solution to it, this would be likely to help solve many other problems than just those that arise for buck-passing accounts of value or admirability. But, in a different way, this generality is bad news. For it also suggests that a failure to deal with this problem would commit a proponent of the reasons-first view to many more false claims than previously thought. In this case, there would be an even better reason for abandoning such views. So in the following sections, I will explore whether

there is any way to account for the reasons that the Reliance View presupposes, in the absence of a solution to the Wrong Kind of Reason problem.

3.   Contractualism

The problem about aggregation just discussed is particularly clear on views that conceive of moral standards in terms of principles that there are reasons for people to reject or accept. One such view is T.M. Scanlon's version of Contractualism. According to Scanlon, the set of obligations which specifies how each person ought to treat each other person – the domain of what we owe to each other – corresponds to the set of principles that no one could reasonably reject. And while aggregative considerations may be relevant to determining what is right or wrong, Scanlon holds that they are not fundamental to determining what it is morally correct to do. Rather, the extent to which they are relevant depends on further facts about the relations between the persons involved in any given case.[15]

In general, Scanlon holds, our reasons for wanting others to obey moral principles, and for letting them guide our own actions, both derive from a more general kind of reason 'to live with others on terms that they could not reasonably reject insofar as they also are motivated by this ideal'.[16] To be motivated in that way is essential to the recognition of the distinctive value of human life, a recognition which involves an attitude of respect towards each individual person and appreciation of his reasons to reject or accept moral principles.[17]

This idea overlaps considerably with the Reliance View's account of obligations, in particular in its emphasis on the normative significance of attitudes like recognition. It may thus appear well-placed to explain why there are reasons for everyone

---

[15] See fn. 12 page 124 above.
[16] Scanlon (1998, p. 154).
[17] Scanlon (ibid., pp. 103–6).

to hold such attitudes. Scanlon appeals to a variety of arguments for thinking that there are. My argument here will be that all of the considerations Scanlon marshals in support of these arguments lead us back to the same basic question about the nature of these reasons.

One of Scanlon's arguments appeals to the idea that relating to other people on the basis of mutually acceptable terms is somehow valuable and desirable in its own right. As Scanlon is aware, appeals to what is 'valuable' or 'desirable' are apt to convey the impression that these reasons can be derived from teleological or Humean conceptions of practical reasoning. But Scanlon argues that we should rejects these views in favour of the buck-passing account of value discussed above: Something is 'valuable' or 'desirable' just in case there are reasons for people to value or desire it.[18] Couched in these terms, Scanlon's claim is that people have reasons to relate to each other on the basis of this ideal because they have *other* reasons to value or desire it. But why should we think everyone has *those* reasons?

Scanlon also sometimes argues that this kind of ideal needs to be understood in terms of a different kind of relation that it is the ideal *of*. Unlike violating the rules of a game that people might have any number of different reasons for deciding to play together, violating the 'rules' of morality reflects the fact that one has failed to appreciate the importance of a relation that all persons stand in, regardless of their attitudes. This is the relation of being 'fellow rational beings' that we discussed in the previous chapter and what, in Scanlon's terminology, counts as the moral relationship. As I argued then, the related notion of 'impairment' is central to this conception of how violations of moral principles are blameworthy. On Scan-

---

[18] Scanlon (ibid., p. 95f).

lon's model, such impairment is a matter of acting on the basis of attitudes that are incompatible with the ideal form of *this* relationship.[19]

Scanlon further defends this idea by arguing that blaming another person is appropriate just in case the attitudes on the basis of which he acted do not reflect the appropriate way of relating to other people.[20] Such appeals to the notion of what is 'appropriate' might be thought to issue in yet an additional normative property to be accounted for. But Scanlon issues a corrective to this line of thought, by pointing out that he thinks of appropriateness in terms of reasons that are 'internal to certain relationships'.[21]

This point raises a question about what it means for such reasons to be 'internal' to relationships, and which relationships they are supposedly internal to. On the one hand, we might understand them in terms of the revised account of blameworthiness I advocated in the previous chapter, as reasons that are somehow shared by people within relationships that they have sufficient reason to have. But this would only seem to push back the question once more, since we would need a separate account of the nature of these *other* reasons to share such relationships.

Alternatively, we might think of them as reasons that somehow obtain in virtue of the relation of 'fellow rational beings' itself.[22] This position raises a number of different concerns. For one thing, the idea that there are reasons to hold attitudes of respect or recognition towards other people in virtue of standing in such a relation to them is compatible with the claim that there are other reasons to hold such attitudes, such as those that an evil demon might provide. Presumably, those are

---

[19] Scanlon (2008, pp. 139–41).

[20] Cp. e.g. Scanlon (1998, pp. 98, 271–2) and Scanlon (2008, p. 124).

[21] Scanlon (ibid., p. 189, p. 234, fn. 58).

[22] This position would be akin to Wallace's relationship-based conception of obligations we discussed in chapter 2, see page 72.

not the kinds of reasons that make it 'appropriate' to respect people, but *why* are they not that kind of reason?

Insofar as the explanation for this requires appeal to a normative property of 'being rational' that is not itself to be understood in terms of reasons, this position forces us to abandon reasons fundamentalism and thus the reasons-first view. Alternatively one may argue, as Scanlon also does, that a reason internal to a relationships is not a reason simply to bring about or promote the existence of people who live on such terms.[23] On the contrary, it is a reason which bears directly on what one's own attitudes should be. In the philosophers' phrase, the reason is agent-relative rather than agent-neutral. But if that distinction between the ways that reasons can favour attitudes is supposed to be primitive, resorting to this reply is to give up reasons unity and thus the reasons-first view.

There is yet another sense in which these reasons might be said to be 'internal' to the moral relationship. Scanlon also maintains that claims about moral reasons are justified via a process that aims to approximate to what John Rawls called 'reflective equilibrium'; a process by which we continuously revise our conception of the basic moral principles and their normative grounds by comparing the implications of these principles with our considered normative judgments about particular cases, carefully assessing at every point whether a revision of the relevant principle or our considered judgment is best supported by the other judgments and principles we hold fixed at that point, and striving at every step to attain the point where all of our considered judgments and principles agree.[24]

We might think of the relation of 'fellow rational beings' simply as the relation that all normally functioning human beings share in virtue of having the capacity

---

[23] Scanlon (ibid., p. 234, fn. 58).
[24] Scanlon (2014, p. 96).

to engage in such reasoning. But the appeal to the character of specific modes of reasoning raises two related concerns. The first concern is that, insofar as this kind of reflective equilibrium reasoning is supposed to proceed from considered normative judgments, it will deliver Scanlon's conclusion about our fundamental reasons to recognise other people only for those who appreciate the normative force of at least some moral considerations in the first place.

It seems possible to imagine someone with the capacity for such reasoning who rejects *all* of the judgments we would consider morally acceptable under reflection. It is wholly unclear, I believe, whether reflective equilibrium reasoning might be able to deliver for such a person the conclusion that he has most reason to live with others on terms that they could reasonably accept. This does not imply, on Scanlon's view, that there are no reasons of this kind for such a person. The question is whether such a person would be in a position to appreciate what those reasons are. It is not clear to me that he would be, on Scanlon's view.

That concern is exacerbated by the fact that, on Scanlon's view, *we* seem not to be in a position to appreciate what those reasons are. When we ask why certain kinds of reasons make actions wrong, Scanlon's claim is that it is because these reasons derive from a certain kind of relationship. But, as we have seen, any attempt to specify those reasons leads us back to claims about what appears to be reasons for those who recognise other reasons to share it. The rationale for why there should be reasons to share the moral relationship in its ideal form thus seems to be wholly internal to the type of reasoning which characterises that very relationship itself. For recognising the force of other people's reasons to reject or accept principles that constrain one's own reasoning just *is*, on Scanlon's view, what characterises

that relationship. So we seem to have come full circle, but none the wiser about the reasons we were asking for.

It thus seems to me that each of the different arguments Scanlon provides raise the same difficulty. At each point, we can either characterise these reasons to which the argument appeals in terms of other normative properties or concepts, thereby abandoning the reasons-first view. Or we can appeal to yet other reasons, whereby we either go around in circles, or repeatedly face new questions about what distinguishes the reasons to which we have now appealed from other putative reasons. It is a bit like saying that the difference between the reason to admire Mary and the reason to admire the evil demon is that if you admire Mary, then it will seem reasonable to you to think there is a difference.

4.   Kantianism

Let us try a different tack. I suggested above that the attitudes of recognition which partly constitute the moral relationship are second-personal. This was supported, I suggested, by the fact that the kind of blame that responds to the impairment of this relationship involves a specific kind of criticism or complaint which one assumes could be addressed *to* another person. Blame thereby involves recognition, because it further presupposes that the person to whom it is addressed has the capacity to govern himself in light of reasons and the standing to make such addresses in return. According to Stephen Darwall, each of us have reasons to respond to each other person with such attitudes of recognition and respect. Moreover, Darwall argues, these reasons provide the normative foundation for the Contractualist account of morality that Scanlon offers.[25] Because the Reliance View incorporates

---

[25] Darwall (2006, Chap. 12).

elements of both these views, this would suggest that the reasons it presupposes are of this second-personal kind.

Darwall holds that these reasons are unlike other kinds of practical reasons in having a distinctive kind of normative basis, one that can be understood only from within a circle of inter-definable normative concepts. A *second-personal reason*, Darwall tell us, is 'one whose validity depends on presupposed authority and accountability relations between persons and, therefore, on the possibility of being addressed person-to-person within these relations'.[26] Such a reason is authoritative, Darwall holds, only because it can be addressed by someone with the *authority* to demand compliance with it. The authority to make such a *demand* itself presupposes a warrant for holding another person accountable for his conduct. And this, in turn, must be a warrant to hold him *accountable to* oneself for responding to such a second-personal reason. And, Darwall further claims, there is no way into this circle of normative concepts from outside.[27]

This aspect of Darwall's position is avowedly Kantian, because he traces its source back to Kant's claim about the 'Fact of Reason'. In the Critique of Practical Reason, Kant claimed that the encounter with the moral law in thought is itself a mode of reasoning, through which we can come to realise that we have the capacity to freely comply with it, independently of any antecedent desires or inclinations.[28] In Kant's famous example, someone might be threatened to lie while realising all the same that it would be wrong to do so. Whether that realisation would eventually lead him to refuse to lie is not decisive. The decisive point is that someone in this position could come to the conclusion that he should, and therefore *could*, do what would

---

[26] Darwall (2006, p. 8).

[27] Darwall (ibid., pp. 11–5).

[28] As Kant puts it '[T]he moral law is the *ratio cognoscendi* of freedom', see Kant (2015, 4n).

be right, no matter his desire to avoid punishment for it. And the content of that judgment is, Kant held, a 'Fact of Reason' in that 'one cannot reason it out from antecedent data of reason'.[29]

According to Darwall, however, this Fact of Reason demonstrates that our obligations have the normative source Kant claims for them only when understood in second-personal terms. He argues for this by pointing out that it does not follow from the mere fact that an agent has most reason, e.g. not to lie, that he is thereby in a position to know about the reasons that make it so, or respond to them in reasoning.[30] If we consider Kant's example again, now from our own perspective rather than that of the agent who is threatened to lie, we might grant the claim that he really is obligated not to lie, and that this really does imply that he has most reason not to do it. But it is consistent with this to deny that *he* is thereby rationally committed not to do it, because he might be rationally ignorant of those reasons. Consequently, the fact that he is obligated not to lie does not show that he could not rationally will to lie.

The connection between the fact of obligation and the agent's will, Darwall argues, derives from the relation between obligation and accountability, in its distinctively second-personal guise of blame. As I did in chapter 1, Darwall defends the claim that blame presupposes that the agent was somehow in a position to respond to the reasons he had to comply with the obligation. In turn, Darwall holds, this implies that those reasons must have been accessible to the agent via some reasoning process.[31] For if they were not, one would lack the authority to demand his compliance with those reasons, an authority which, according to Darwall, is built into the

---

[29] Kant (ibid., p. 31).
[30] Darwall (2006, p. 240).
[31] Darwall (ibid., p. 241).

accountability relation that blame further presupposes. What does not follow from the Fact of Reason itself does follow from the normative features built into its connection with the irreducible circle of concepts identified above: second-personal reason, authority, demand, and accountability.

This relation to accountability Darwall further traces back to what he calls 'Strawson's Point': the idea that the fact that it would have desirable or valuable consequences that we take such attitudes towards other people is the 'wrong kind of reason' for doing so.[32] A demand that another person get off one's foot, for instance, addresses a reason that he does so, but the reason cannot be that it would have desirable consequences that one so addresses him. That is not the kind of reason the agent could appropriately hold himself responsible for complying with, for instance by feeling remorse for disregarding it. A reason to hold second-personal attitudes towards someone, Darwall thus holds, must itself have a second-personal content.

There is a concern, however, about the arguments that lead to this conclusion, one that is well-known from discussions of Strawson's own view. Gary Watson has raised a particular worry about the Strawsonian idea that blame presupposes the possibility of being addressed. For addressing blame to people seems to presuppose that they can understand why we are blaming them, and this seems to presuppose in turn that they share an appreciation of basic moral claims. But that presupposition does not seem to be satisfied at what Watson calls 'the limit of evil'; people who simply reject all such claims. Since one could not address moral claims or demands to such people in the expectation that they would comply with them, it can that seem one could not coherently blame them for holding such attitudes. And this would suggest, paradoxically, that extreme evil would somehow be its own excuse.[33]

---

[32] Darwall (2006, pp. 15–7, 65–70).
[33] Watson (2004).

Christine Korsgaard has raised a similar concern about Darwall's position.[34] As Korsgaard points out, Darwall's position does not seem to rule out the possibility that a person could fail to hold second-personal attitudes altogether. And it is somewhat unclear how a person who does not *already* hold second-personal attitudes could reason his way to the conclusion that he has some kind of *second-personal reason* to do so. For to recognise that reason, he would have to recognise the authority of other people's warranted demands and, by assumption, he does not. So if such a person does have most reason to take up the second-person standpoint by forming such attitudes, it seems the reasons that make it so must be of some *other* kind. In reply, Darwall argues that:

> The way we get into the second-person standpoint is not by seeing a non-second-personal reason for doing so and then taking it up. These would not give us reasons on which we could genuinely take it up anyway. They would be 'reasons of the wrong kind.'[35]

And for a fact to be a reason of the right kind, Darwall tells us, it must be:

> a fact about or feature of some object, appropriate consideration of which could provide someone's reason for a warranted attitude of that kind toward it. It must be something on the basis of which someone could (and appropriately would) come to hold the attitude as a conclusion of a process of considering (deliberating about) whether to do so.[36]

A reason of the right kind to take up the second-person standpoint is thus a fact deliberation about which makes the attitude of respect appropriate or 'fitting' and that kind of fact is, Darwall holds, another person's *dignity*.[37] While Darwall thus

---

[34] Korsgaard (2007).

[35] Darwall (2007, pp. 59–60).

[36] Darwall (2006, p. 16).

[37] Darwall says: 'Respect is the fitting response to dignity, as esteem is to the estimable, desire is to the desirable, and so on. Just as reasons of the right kind for desire and esteem must be drawn from the distinctive ways these attitudes relate to their characteristic objects, the desirable and the estimable, respectively, so it is also with respect' Darwall (ibid., p. 120).

concedes that there may be non-second-personal reasons to take up the second-person standpoint, these can be distinguished from reasons of the right kind to do so only by appeal to normative concepts such as 'fittingness' and 'dignity'.

This position avoids the circularity concern raised by Scanlon's view by explicitly giving up reasons fundamentalism and thereby the reasons-first view. But it also raises a different concern. For it seems to imply that reasoning on the basis of second-personal concepts is the only means by which anyone could come to the conclusion that they had most reason (of the right kind) to hold second-personal attitudes. And that makes it difficult to understand how anyone could be *blameworthy* for failing to hold them in the first place. As I argued in the previous chapter, we can hold people accountable for their attitudes. And I specifically suggested the case of someone, Ego, who appeared to be blameworthy precisely for failing to hold such attitudes. Is this possible?

On this point, Darwall engages in an illuminating discussion with Derek Parfit about an Act-Consequentialist who is, as Darwall describes him, 'blind, conceptually and motivationally to the moral ought in the "ordinary" sense, that is, to moral obligation, demand, and blameworthiness'.[38] Darwall further describes the case as follows:

> As we are imagining it, the being would lack the Strawsonian reactive attitudes essential to "second-personal competence." In other words, although he forms desires for states of affairs from an impersonal, third-person's perspective, he lacks the capacity to take a second-personal perspective toward himself and others and form implicitly demanding, holding accountable attitudes.
>
> [l]acking reactive attitudes, he would be blind to others' blame and incapable of holding himself accountable through feeling guilt. So although there would be nothing incoherent or otherwise untoward about incarcerating him, it would

---

[38] Darwall (2017a, p. 275).

be incoherent to think that in doing so we were holding him accountable, that our actions could intelligibly be seen as expressing blame to him.[39]

Darwall thus assumes that two things are the case here. First, that this person lacks the relevant attitudes. Second, that he lacks the capacity to hold them. These are different claims, however, and the first does not imply the second. We can imagine people with the capacity to feel anger, for instance, but who do not actually feel anger. Moreover, we can imagine people who *inexcusably* fail to hold the attitudes they have reasons to. So even if we accept Darwall's claims above, it seems to me that we should also accept that one could, in principle, be blameworthy for failing to exercise one's capacity to hold second-personal attitudes.

To account for this fact on Darwall's view, we would need to establish that there was a sound deliberative route for such a person to the conclusion that there is an authoritative demand on him to take up the second-person standpoint. Yet *holding fixed the fact that he is not exercising his second-personal capacities*, it seems there is not *now* any second-personal reason on the basis of which he could arrive at that conclusion. To appreciate such a reason he would have to accept claims about the authority of other people's warranted demands and, by assumption, he does not. And according to Darwall, no other type of reason will do, since any such reason would be of the wrong kind. On Darwall's view, then, this person would not be able to take up the second-person standpoint *for* those reasons either. And this seems to imply he could not be blameworthy for failing to take up second-personal attitudes, because he seems not to be in a position to take them up *for* any of the reasons that are available to him.[40]

---

[39] Darwall (ibid., pp. 275–6).

[40] Darwall (2016) correctly points out that those of us who already hold second-personal attitudes may be blameworthy for coming to relinquish them, but this is a different point. First, from the

It is not obvious to me that this is true. Darwall and I both accept that blame presupposes that the blamed party was in a position to respond to the reasons that one is holding him accountable for disregarding. But it is not clear to me why one must actually hold second-personal attitudes for this to be so.[41] Here is an analogy to illustrate the concern: Suppose your eyes are closed. You are now told that there is a reason for you to open them. When you ask what the reason is, you are told that you will be able to see it, if you open your eyes. As we can assume, you have the capacity to do this. As we can also assume, the reason is actually there, and what you have been told is true: If you open your eyes, you will see what the reason is, and be motivated to respond to it.

Are you in a position to respond to this reason? This question may be ambiguous. You are in a position to respond to the reason in the sense that you can do what you have a reason to: opening your eyes. There is another sense in which you are not in a position to respond to this reason. You cannot open your eyes *for* this reason by *seeing* it, because you are not in a position to see it *given that your eyes are closed*. Does it follow that you are in no position at all to respond to this reason, by finding out what the reason is? It seems to follow only if we assume that the reason in question could not be available by any other means.

Of course, *if* that is true, it may also be true that the only way to distinguish the reasons that make it obligatory for people to open their eyes and those that do not is by appealing to other things that are visible only to the naked eye. If Darwall's

---

fact that one is in some condition C, it does not *follow* that one was responsible or blameworthy for ending up in C. Second, even if one *was* blameworthy for getting oneself into C, it does not follow that one is *now* blameworthy for failing to get out of C. My claim here is that, where C is a condition that involves the lack of second-personal attitudes, a proponent of Darwall's position needs an explanation for how one can blameworthy for failing to get out of C.

[41] McKenna (1998) similarly replies to Watson's concern above that moral accountability, on Strawson's view, merely presupposes that one has the *capacity* for moral address, not that one *exercises* it.

response to the Wrong Kind of Reason problem is correct, something like this may be true of our reasons to take up the second-person standpoint. The point here is just that the price of this conclusion seems to be a lack of a distinction between those who are wilfully blind and those without functioning eyesight. That may be a high price to pay.

5.   Consequentialism

I argued above that both Scanlon's and Darwall's strategies for specifying the nature of our reasons to take up the second-person standpoint are problematic. In different ways, these problems have to do with their attempts to specify those reasons in terms of the content of the very attitudes they are said to support. So we might consider instead specifying these reasons independently of any such attitudes. A certain type of Consequentialist argument may seem to offer us the resources for doing so.

Consequentialism can be stated as the view that everyone has most reason to respond, in their actions or attitudes, in whatever way will make things go best. Thus, Consequentialists can argue that, if it would make things go best that everyone take up the second-person standpoint, then everyone has most reason to take up the second-person standpoint.[42] Consequentialists *may* hold a buck-passing view of value, arguing that those states of affairs are best that provide most reason for, or reasons for the highest degree of, valuing attitudes. To avoid the Wrong Kind of Reason problem, however, Consequentialists may also simply give up the reasons-

---

[42] This way of putting things requires that we set aside non-ideal cases in which, because some people do not do what they have most reason to do, it would make things go *worse* if the rest of us do what we would *otherwise* have most reason to do. Because my discussion here concerns a question that arises, even when we assume that everyone responds as they have most reason to, I set aside these non-ideal cases here.

first view and claim that such reasons derive from the property of being impartially valuable, a property that states of affairs can have to a greater or lesser degree.

As we have seen, the idea that such 'state-of-the-world'-regarding reasons are not of the kind that provide support for second-personal attitudes can be motivated in a number of ways. The standard Consequentialist response to this sort of objection is that it confuses the Consequentialist *criterion of rightness* with a Consequentialist *decision procedure*. If it turns out that it would have better consequences on the whole that no one employs a Consequentialist theory in deciding what to do, Consequentialists can hold that they have most reason not to do so. The Consequentialist theory thus becomes *self-effacing*.[43]

Suppose, for instance, that the Consequentialist grants the claim that Consequentialist reasons are reasons of the wrong kind to blame someone for failing to hold an attitude like recognition towards others, *relative to the standards of second-person standpoint*. This Consequentialist concedes that, for anyone within the second-person standpoint, such a fact will seem to be irrelevant with respect to determining whether he is blameworthy. He may even concede that it is a conceptual truth about blame that it presupposes robustly second-personal reasons. He maintains, however, that there are no second-personal reasons to take up the second-person standpoint itself. This Consequentialist can accept that, if someone fails to take up the second-person standpoint, they are not blameworthy for it. Indeed, it seems he *should* accept this, if he accepts that blame conceptually presupposes second-personal reasons. Yet he can maintain, consistently with this concession, that they

---

[43] Parfit (1984, Chap. 1) discusses various implications of this possibility in detail. As Schafer (2016) points out, other moral theories can take a self-effacing form, even Kantian ones

have failed to do what they are morally obligated to do, on his account of obligation.[44]

The obvious reply is that such a Consequentialist view would fail to explain how we might be obligated *to* people to take up the second-person standpoint. It is not clear, however, that the Consequentialist should accept that this is something his account *should* explain, on his own terms. Moreover, it is not clear that the Consequentialist is any worse off on this score than a proponent of Darwall or Scanlon's views. For, as we saw, these views equally seem to have problems when it comes to explaining why the reasons that obligate people to take up the second-person standpoint are of a kind that makes it correct to blame them for failing to. The difference seems to be that the Consequentialist can specify our reasons to take up the second-person standpoint independently of concepts and properties that are internal to it. And this seems to give Consequentialism a leg up on these competing theories.

Even if this reply is sound, however, there is a deeper problem with this Consequentialist strategy. For note that the above reply merely concedes the point that Consequentialist reasons to hold second-personal attitudes are of the wrong kind *relative to the standards of the second-person standpoint*. In availing himself of the distinction between criteria of rightness a decision procedures, and relegating the standards of the second-person standpoint to the latter, the Consequentialist is committing to the idea that those standards are normative, in the sense that there are reasons to comply with them, only because of a more fundamental Consequentialist obligation to take up second-personal attitudes. And *that* obligation obtains in virtue of the kinds of reasons the Consequentialist theory identifies.

---

[44] Micah Lott appeals to these claims in arguing that such 'wrong kind of reason' objections really have no bite against a self-effacing Consequentialist theory. See Lott (2016).

Now suppose we hold this characterisation fixed: The kinds of facts that ground obligations are facts about the consequences of doing actions or holding attitudes. On this view, it will not be true that those who fully comply with their obligations do so *for* the reasons that obligate them. For instance, the fact that they have given a promise will not be a reason for anyone to do what they have promised to. Nor will the fact that those who have accepted the promise rely on one to keep it, or the fact that they can justifiably demand that one keep it. Since these are not facts about the consequences of doing an action or holding an attitude, they do not ground any obligations. To be clear, they may still be facts that everyone can justifiably *think* is a reason for people to do what they have promised to. And they may also help *explain* why one is obligated to. But they are not, on this view, the facts that *make* it the case that one is obligated to. Because only facts about consequences do so.[45]

On the account of trust I defended in chapter 1, this would be to say that those who fully comply with their obligations are not trustworthy. Similarly, it will not be true *by the lights of this Consequentialist view* that anyone is blameworthy for disregarding the reasons that obligate them. As we have just seen, the Consequentialist reply here concedes that those who fail to respond to their reasons to take up the second-person standpoint are not blameworthy for doing so. And it also concedes that those who *do* take up the second-person standpoint will not be blameworthy for violating the obligations that are internal to it. For instance, in disregarding the fact that they have given someone a promise, these people would not be disregarding a reason that grounds an obligation, though they would be disregarding what they could justifiably *think* is such a reason.

---

[45] Markovits (2010) makes this point in considering the implications of the view discussed in chapter 1, namely that actions have moral worth only if done for the reasons that make them right.

These claims seem to imply that, among those scenarios in which everyone complies with their fundamental Consequentialist obligation, no one is ever trustworthy, even in the best scenarios. They also seem to imply that, among those scenarios in which everyone comply with this fundamental obligation, no one is ever blameworthy, even in the worst of these scenarios. These implications are perhaps no less paradoxical than those we encountered in discussing the views above. But the most striking implication, I believe, is that the *only* people who are in a position to comply with their basic Consequentialist obligation, *for the reasons* that make it the case that they are so obligated, are those who do *not* already hold second-personal attitudes. For if they *did* hold these second-personal attitudes, then they would not, on this view, take those Consequentialist reasons to provide the appropriate kind of justification for second-personal attitudes in the first place.

For these reasons, it seems to me that each of these alternative approaches to specifying our reasons to take up the second-person standpoint come with significant costs. Scanlon's Contractualist account allows us to hold on to the reasons-first view only by seeming to commit us to an explanatory regress. Darwall's Kantian account gives up the reasons-first view and specifies these reasons directly by appeal to other normative concepts and properties. This approach seems to commit us to the claim that those who simply fail to take up the second-person standpoint *tout court* could not be blameworthy for doing so. And the Consequentialist account that gives up the reasons-first view by specifying reasons as facts about impartial value seems to concede that those who do take up the second-person standpoint do not act for the reasons that obligate them.

Each of these implications seems to me hard to accept. Of course, they may seem *more* acceptable than the implications of the reasons-first view, in the absence

of a solution to the Wrong Kind of Reason problem. If there is no such solution, proponents of the Reliance View may need to adopt one of the responses to this problem that these views offer. The problem is that each of these views threaten to undermine the motivation for the Reliance View itself.

The first view makes it exceedingly difficult to explain what the fundamental reasons to share second-personal relationships are. The second makes it difficult to explain how one might be blameworthy for failing to hold the attitudes that constitute them. And the third fails to explain how, in holding such attitudes, people might act for the reasons that obligate them. I argued that the Reliance View was plausible precisely because it seemed to offer some explanation on each of these points. So, if we should adopt either of these views, the Reliance View is less plausible than I have argued. Thus, it seems to me that the apparent plausibility of this view is motivation to see whether the Wrong Kind of Reason problem can be solved. That is what we will be doing in the next chapter where I will argue that it can be solved. And the solution I will defend supports the claim that the reasons presupposed by the Reliance View are grounded in interpersonal activities. And that, you will recall, is just what Pragmatism says.

Chapter 5

Reasons, Correctness, and Activities

In chapter 1, I presented Pragmatism as a basic account of the normativity of obligations and suggested a more particular view that might support it. That was the Reliance View which I defended in chapter 2 and chapter 3. That view is plausible, I argued, insofar as we can defend the claim that there are reasons for everyone to share second-personal relationships by holding attitudes of recognition towards each other. I argued in the previous chapter that this raises a general problem about distinguishing the reasons that determine correctness standards from the reasons that do not. That was the Wrong Kind of Reason problem. Because that problem seemed particularly difficult to solve on the basic assumptions of Pragmatism, I then explored some alternative ways of specifying what the right kind of reasons were with respect to obligations and argued that each of them threatened to undermine the plausibility of the Reliance View. So in this chapter I will explore whether we can provide a distinction between right and wrong kinds of reasons on the reasons-first view that Pragmatism assumes. To do so, we need to solve the Wrong Kind of Reason problem.

My solution to this problem is motivated by some observations about the elusive 'right kind' property that often go unnoticed. So I will start off in Section 1

by making some observations which I believe a solution to the problem should ac-
count for. In Section 2, I then set out the basis of my own solution, elaborate on
the different elements that make it up, and put the pieces together. In Section 3,
I note some appealing features of this view as a basic approach to the relation be-
tween reasons and correctness, and show how it applies to the specific instances of
the Wrong Kind of Reason problem we have encountered along the way. Finally, in
Section 4, I give a more detailed account of how this amounts to a theory about the
nature of obligations. On this view, the fact that another person relies on you is not
only a reason to do what they rely on you to, it is a reason of the right kind to do it.

1.   Background Conditions

When we ask whether some fact is a reason, it matters what the background con-
ditions are. Think of 'background conditions' simply as those features of a context
that help determine whether some fact is a reason for a person to respond in some
way. So for instance, whether the fact that there will be live fusion jazz at the party is
a reason for you to go depends on further background conditions: If you like fusion
jazz, it is a reason. If you do not like fusion jazz, it might not be. My aim in this
section is to argue that the property of being a reason of the right kind also varies
with facts about background conditions, and that the way we draw the distinction
between right and wrong kinds of reasons ought to account for this.

Let us start with a simple point. When the evil demon threatens to destroy
the whole world unless people admire him, that is a reason of the wrong kind to
admire him: it does not make him admirable. But it arguably *does* make him de-
plorable. If the properties of being admirable and deplorable are equally normative,
buck-passing about both of these properties implies that the very same fact can be a

reason of the wrong kind for one response (admiring) and a reason of the right kind for another response (deploring). So whether or not a reason has the 'right kind' property varies with *what* it is a reason for, and this further depends on background conditions of the case are. As we can all agree, there are at least *some* background conditions on a fact's being a reason e.g. for admiration. But if background conditions determine what facts are reasons for, and if what something is a reason for determines whether that fact is of the right or the wrong kind, background conditions help determine whether a reason has the 'right kind' property.

This point may seem innocuous, but is a substantive one, as evidenced by the fact that it seems to cause serious trouble one some ways of thinking about the distinction between right and wrong kinds of reasons. For instance, as Richard Rowland thinks of it, whether a reason is of the wrong kind *never* depends on what the background conditions are. On his view, a fact is a reason of the wrong kind for some response just in case it is provided or enabled by a fact about what he calls an 'additional consequence' of the response itself, that is, something which is not merely a logical or conceptual consequence of that response.[1] This view is intuitive. As we can all agree, there are many cases where facts about the consequences of having an attitude seem to provide reasons of the wrong kind to have them. The fact that the evil demon will destroy the whole world unless you intend to do something you have no reason to do would be one such case. But consider a slightly different type of case.

Suppose you know that you have most reason to ϕ, but remain are unsure whether intending to ϕ will help or hinder ϕ'ing. You know yourself well enough to know that, sometimes, you actually make it *less* likely that you do what you ought to, if you

---

[1] Rowland (2017).

intend to. In this case, it seems to me, it would be mistake to go ahead and intend to ϕ. But you now discover the following fact: In this particular case, intending to ϕ will make it more *more* likely that you ϕ. Does this discovery make a difference to whether or not you should intend to ϕ? It seems to me that it does. Could you rationally respond to this discovery by forming that intention? It seems to me that you could. Would it be unfitting for you to intend to ϕ in this case? It does not seem to me that it would. And all of this, it also seems to me, indicates that this fact would be a reason of the right kind to form that intention. Even so, it is a fact about an additional consequence of forming that intention.

This suggest that whether or not a fact about the consequence of having an intention constitutes a reason of the right kind to form it depends on further facts, not least whether the intention it is a reason for aims at an end that is supported by other reasons. And this is just to say that the distinction between right and wrong kinds of reasons should be sensitive to variations in background conditions. But just as background conditions can help determine *what* something is a reason for, they can also help determine *whom* they are reasons for. This too suggests that the 'right kind' property is dependent on variations in background conditions.

We can appreciate this point by considering a close cousin of the view for which the Wrong Kind of Reason Problem first arose. On the buck-passing account of value, the fact that something is valuable is a normative fact only because there are reasons for people to value that thing. But aside from G.E. Moore, most now agree that in addition to the property of being good or valuable *simpliciter*, there is also the property of being good *for* someone or having agent-*relative* value.

If buck-passing about normative properties in general is intelligible, a buck-passing account of the good-for relation should be so too. What would such a view

look like? Suppose there is exactly one apple that either Jim or John could have. Intuitively, it may be good for Jim that Jim has the apple, but not good for John that Jim has the apple. On a buck-passing view, that must be because some properties of Jim having the apple provide reasons (of the right kind) for valuing that, while some properties of John not having the apple provide reasons (of the right kind) for disvaluing that. However, if those reasons were all of the right kind for *everyone*, then the fact that Jim's having the apple is good for Jim would just imply that there are reasons for people in general to value that. But that was precisely how we stated the buck-passing account of good *simpliciter*. So if being good *simpliciter* is different from being good for someone, that cannot be the right way to state buck-passing about the good-for relation.

It is thus natural to think that the fact that something is good for Jim implies that *some* people have reasons to value that, while the fact that something is not good for John implies that *other* people have reasons to disvalue that. Which people? On a simple view, Jim and John: Jim has reasons to value Jim's having the apple, John has reasons to disvalue Jim's having the apple. And if that is right, then it seems the buck-passing account of agent-relative value implies that some facts are reasons of the right kind for some people to hold valuing attitudes, but not reasons of the right kind for other people to hold valuing attitudes.

Of course, the simple view may be mistaken. For instance, Darwall has argued that the fact that something is good for Jim does not necessarily imply that *Jim* has any reasons of any kind. Rather, for something to be good for Jim is for it to be rational for people who care about Jim to want it for him.[2] Or, as the buck-passer should say, it is for what people who care about Jim have reasons to want it for him.

---

[2] Darwall (2002).

Jim may or may not be one of these people. But note that, on this account of the good-for relation too, buck-passing implies that the distinction between right and wrong kinds of reasons is dependent on further facts, such as facts concerning who cares about whom.

I do not mean to suggest that either of these views must be true. They merely illustrate that on some substantive views about a property that many hold to be normative, the higher-order property of being a reason of the right kind is sensitive to variations in background conditions. And while these substantive views may be false, it seems to me that a satisfactory account of the distinction between right and wrong kinds of reasons should not prejudge the issue. Such an account should be compatible with the possibility that a reason may be of the right kind for some people to respond in some way, but not a reason of the right kind for other people to respond in that way, depending on further facts about the case.

Further evidence that the distinction between right and wrong kinds of reasons should be sensitive to variations in background conditions comes from the observation that the property of being a reason of the right kind is also a *defeasible* property. As I noted in chapter 3, it is widely agreed that the property of being a reason is itself a defeasible property. That is why reasons can be defeated. To see how something similar applies to the 'right kind' property, consider the case of Ronnie. Ronnie is threatened by an evil demon on pain of death to do something he has sincerely promised to do, before midnight. But Ronnie does not have any outstanding promises to deliver on. And the only person around to give a promise is Ruth, who will only accept a promise to help her out with the gardening (and she would really like some help with this).

Now, if evil demons can provide people with reasons to value things by threatening them with death, they can also provide people with reasons to give a promise. So there should be a reason for Ronnie here to sincerely promise Ruth to help with the gardening. But of course, merely *giving* the promise is not going to keep Ronnie from getting killed. So there should also be a reason for Ronnie to go ahead and actually help Ruth with the gardening before midnight. And Ronnie has that reason right now. But since helping Ruth *before* he has promised to will not avert the demon's threat, it is *inter alia* a reason for Ronnie right now to *intend* to help Ruth with the gardening before midnight.

It seems to me that there is nothing 'unfitting' about forming an intention to do something for the reason one has to do it.[3] It may be irrational, of course, if one has other, weightier reasons not to form that intention, but this does not show that the reason in question is of the wrong kind. Intending to eat nuts for the reason that they are savoury may be irrational for me if I know that I am allergic to nuts, but it hardly follow from this that the savouriness of nuts is a reason of the wrong kind for me to intend to eat them. So I see no reason to suspect that Ronnie's reason to intend to help Ruth with the gardening is of the wrong kind.

But now suppose Ronnie actually goes ahead and forms this intention, proceeding to give Ruth a sincere promise to help with the gardening. Now Ronnie's reason to intend to do what he has promised Ruth to do is no longer a reason of the right kind. After all, what was Ronnie's reason? It was the fact that the evil demon would

---

[3] The reason to form this intention thereby differs from the reasons that are usually taken to be of the wrong kind for an intention, such as the fact that an eccentric billionaire will give you a huge reward for intending to drink a vial of poison. This seems to be a reason of the wrong kind to intend to drink the poison, because it is much more difficult to accept that you have *any* reason to drink a vial of poison. After all, you will not get any reward for actually drinking it. See Kavka (1983) for the original 'toxin puzzle' case and Piller (2001); Piller (2006); Hieronymi (2005); Schroeder (2012) for discussion of its relation to the distinction between right and wrong kinds of reasons.

strike him dead if he did not do something he promised to. That fact is still a reason; it still counts in favour of doing it. But it is just a fact about Ronnie. And, as we saw in chapter 2, there is a wide range of views according to which the kind of fact which makes it fitting for Ronnie to intend to do what he promised Ruth to do is supposed to be a fact that involves *Ruth*. That is the sense in which promissory obligations are bipolar. And if the fact that the evil demon will kill Ronnie unless he does what he promised to is no longer a reason of the right kind, that must be because something about giving the promise now makes it a reason of the wrong kind. In other words, its status as a reason of the right kind must have been defeated.

If these claims are correct, a satisfactory account of the distinction between right and wrong kinds of reasons needs to explain how a reason's property of being of the right kind can depend on variations in background conditions. And the way it characterises those conditions will need to allow that they can vary from person to person at a time, overlap for the same person at a time, and vary for the same person across time. That seems like a tall order. To appreciate why, it will help to consider how we got to the Wrong Kind of Reason problem in the first case.

We came to it by realising the difficulties involved in explaining why there are reasons for everyone to hold second-personal attitudes, including those people who do not already hold them. What we need, then, is some way of explaining how reasons of the right kind can be, as Christine Korsgaard puts it, 'reasons we can share'.[4] The reason this is likely to seem difficult, I believe, is that it is both natural and common to think of shared reasons as reasons the validity or significance of which are recognised from some shared 'point of view'. But the problem with viewpoints is that it is difficult to hold more than one at a time. So if the kind of background condition that makes a reason 'right' is something validity from a certain standpoint,

---

[4] Korsgaard (1993).

then the 'right kind' property could work in the way the above arguments suggest only if we could occupy many different points of view at the same time, constantly switching back and forth between them. And that can seem difficult to make sense of.

Fortunately, I think there is a different way of making sense of how reasons could be shared, one that is actually quite amenable to the claims above. For instance, if you and I both love surprise parties thrown in our honour, then the fact that there is a surprise party for both of us behind the door is a reason for each of us to enter it. There is an intelligible sense in which it is a reason we do share, and so it is a reason we can share. But this is not because we can both recognise the validity of this reason from a shared point of view. For since neither of us can recognise the validity of that reason without making it go away, there is no such point of view. So there has to be something else in virtue of which reasons can be shared. And this brings me to a proposal by Mark Schroeder.

On Schroeder's view, reasons are of the right or wrong kind for people to respond in some way relative to their engagement in some kind of activity.[5] This idea seems to me better suited to make sense of the 'right kind' property's normative features. Since one can be engaged in different activities at the same time, a reason can be of the right kind relative to the one activity, but not the other. Since different people can be engaged in different activities, a reason may be of the right kind for some, but not for others. Since the nature of one's activities can change over time, a reason can be of the right kind at one instant, and of the wrong kind at the next. And since it is equally possible for people to be engaged in the same activities, they

---

[5] On Schroeder's official statement of the idea: 'The right kind of reasons with respect to any activity, A, are all and only those reasons which are shared by necessarily every able person engaging in A, because they are engaged in A, together with all reasons which are derivative from such reasons' Schroeder (2010, p. 39).

might also share certain reasons in virtue of that. So here is a salient hypothesis: The kind of background condition that makes a reason have the 'right kind' property is engagement in an activity.

That idea looks fairly promising. But there is obviously much work left to be done to show that there is some substantive notion of 'activity' which can do all of this explanatory work. Sadly, Schroeder himself tells us little about what he thinks activities are, or how reasons obtain 'in virtue of' engagement in them. Nevertheless, I think this work can be done. And I also think that the resulting account provides more than just an independently plausible solution to the Wrong Kind of Reason problem. As I shall later argue, it provides the kind of solution which illustrates why Pragmatism is a particularly plausible account of the nature of the reasons in virtue of which our obligations obtain.

2. Activities

My suggestion is going to be that activities are complexes of actions. So let me start by clarifying what I mean by 'action'. Following a recent account defended by Yair Levy, I shall say that an event is an action by an agent when that agent has the specific ability to directly intentionally control whether that event occurs.[6] Formally:

ACTION: ϕ is an action by agent S at time t if and only if S has the specific ability at t to intentionally and directly stop and continue the event identical with, or partly constitutive of, S's ϕ-ing.

I believe this view is both independently plausible and well-suited for developing an account of both physical and mental activities. To illustrate this, let me clarify

---

[6] Levy (2016). The clarificatory remarks below paraphrase Levy's.

in further detail what it says. Here, 'specific ability' is meant to combine two conditions. The first condition we can call 'general ability'. Whether I have the general ability to raise my arm does not depend on what I am now in a position to do. Thus, I have the general ability to raise my arm, even when I am asleep and therefore in no position to raise it. The second condition we can call 'opportunity'. Whether I have the opportunity to raise my arm does not depend on whether I have the general ability to do so. Sitting in my chair when my arm is limp after an operation, I may still have the opportunity to raise it, since nothing would prevent me if had the general ability. When I have the specific ability to raise my arm, then, I have the general ability to do so, and the opportunity to exercise it.

Further, 'ability' is gradable. If Jones shoots Simpson, whether or not he thereby kills Simpson may be something Jones has little or no ability to intentionally control. Once the bullet has hit, he may be unable to control whether the doctors subsequently save Simpson's life. The event 'shooting Simpson' can thus be fully under Jones' control while the event 'killing Simpson' is not, but (if Simpson dies as a result of being shot) each event can constitute an action. They will both be actions insofar as Jones had the specific ability to intentionally control one event (the shooting of Simpson) that (partly) constituted another event (the killing of Simpson).

Consequently, which *which type* of action a given token event amounts to may be subject to more or less intentional control.[7] Indeed, *this* aspect of the action may be completely outside of any individual agent's direct intentional control, as when Jones cannot control whether the event he brings about will constitute an action of the type 'killing'. So what matters primarily to the question of agency is whether the

---

[7] A precise account of the relation between tokens and types is more than I can provide here, so I will assume for simplicity that for a token action to be of a certain type is for it to instantiate a given property.

event itself that constitutes the occurrence of the action was something the agent had the specific ability to intentionally control.

For simplicity, I shall assume that 'intentional control' is a primitive, not subject to further analysis. This idea can be motivated on much the same grounds as the idea that knowledge is a primitive: no satisfactory analysis appears to be forthcoming. And the proposal here is similar to that of knowledge-first epistemology; to reverse the order of explanation and let intentional control do the basic explanatory work in our theory.[8] Nevertheless, I shall not pursue the point here, since ACTION is compatible with a causal or counterfactual analysis of 'intentional control' if one wishes to defend it.

A final point is that ACTION is highly inclusive. For instance, it allows that events brought about by habits can be full-fledged actions. Habitually scratching one's head may be an action even if it is not brought about by a desire, insofar as one has the specific ability to directly stop or continue scratching one's head. This inclusiveness is a virtue, because it also allows ACTION to easily accommodate the phenomenon of mental agency. For instance, things like attending, listening, calculating, and daydreaming all come out as instances of mental agency, insofar as they are events that one has the specific ability to directly stop or continue. Thus, it is well-suited to develop a maximally general account of activities which includes physical agency, mental agency, and both. And that seems to be precisely what we need to develop the above solution to the Wrong Kind of Reason problem.

---

[8] This point is defended in detail in Levy (2013). Williamson (2017) provides a further defence of the primacy of both knowledge and intentional action and the analogies between them. The notion that direct intentional control is primitive in this sense also seems to me to be motivating the type of compatibilist position about moral responsibility defended by R. Jay Wallace: '[T]he conditions of responsibility do not involve freedom of the will, but primarily include the possession of certain rational powers: the power to grasp and apply moral reasons, and the power to control one's behaviour by the light of such reasons' Wallace (1996, p. 7).

So the basic idea is that we can use activities to explain the difference between right and wrong kinds of reasons, where activities are complexes of actions, in the specific sense of ACTION. What does it mean to say that they are 'complexes of actions'? Start with an idea that is definitely wrong. Suppose we said that activities are just sets of actions. So, for instance, when playing golf is a type of activity, this just means that it is a set of action types that one can perform. That idea is both simple and metaphysically innocent. Unfortunately, it also has some disastrous implications. Because if the right kind of reasons are those people share in virtue of being engaged in some activity, and if activities are just sets of actions, then every reason is going to be of the right kind relative to what it is a reason for, because for every single action there is a set with that action as its only member. So if activities are going to explain any difference between right and wrong kinds of reasons, there must be *some* difference between activities and sets of actions.

The difference, I believe, is that some complex actions are *prime*. To say that a condition is prime is to say that it is not composite; that is, not analysable as the conjunction of its components.[9] To see what this means in our context, we can give a simplified mathematical analogy. Assume, for the sake of the argument, that we can identify individual actions with cardinal values. Assume also that conjunction can be represented as addition such that a composite action is just the sum of the values of its component actions. Then there is a simple way to show that some complex actions are prime.

Here is my example: Al and Betty are in Prison. They are accused of a crime, and they are facing a Dilemma. If they both cooperate with each other and say nothing,

---

[9] Primeness is well-known from Timothy Williamson's argument that knowledge is not a composite state, but other types of conditions and properties may equally be prime, cf. Williamson (2000, p. 68).

then they each get x. If each rats on the other, then they each get y. If Al cooperates while Betty rats, then Al gets z and Betty gets w, and *vice versa*. Al and Betty are in a bad fix because their payoffs satisfy two conditions. The first is that the values of their individual rewards are ranked such that $w > x > y > z$. The second is that the individual rewards for mutual cooperation are strictly greater than the average of the individual rewards for cooperating when the other defects and *vice versa*; that is $x > (w + z)/2$. And together, those two conditions make it the case that Al and Betty's situation is a terrible Prisoner's Dilemma.[10]

To demonstrate a case of primeness, we start with a case α in which Al cooperates and Betty defects. We construct a case β by reversing the inputs such that Betty cooperates and Al defects. We now construct a case γ which is like α in respect of Al and like β in respect of Betty. In γ, Al and Betty both cooperate, so γ is a case of mutual cooperation. Now suppose we identify Al and Betty's individual actions in α and β with the value of their individual rewards. Then the value of Al's action in α is w and the value of Betty's action in β is also w. So if γ were composite, its total value should be $2w$. But the total value of γ is $2x$, and the first condition of the Prisoner's Dilemma ensures that $2x < 2w$. Now suppose instead we identify the value of Al and Betty's individual actions in α and β with their equal 'share' in the total output. Then the value of Al's action in α is $(w + z)/2$ and the value of Betty's action in β is also $(w + z)/2$. So if γ were composite, its total value should be $(w+z)/2+(w+z)/2 = w+z$. But the total value of γ is $2x$, and the second condition of the Prisoners' Dilemma ensures that $2x > w + z$. So there is apparently no way to construct γ as a composite of α and β. And if the condition in γ is not composite,

---

[10] The first condition is standard. The second condition is seldom discussed, but nevertheless necessary to ensure that defection remains the dominant option in iterated versions of the game, see e.g. Rapoport and Chammah (1965, p. 34).

it is prime. So any instance of mutual cooperation in a Prisoners' Dilemma is an instance of a complex action that is prime.[11]

So here is my suggestion: Activities are complex actions that are prime. They can be composed of all sorts of individual actions, mental and physical; their individual parts may have further parts; those parts may themselves be prime; and all those parts may bear any number of sophisticated relations to each other. Of course, the procedure I have used to bring out this idea will not apply straightforwardly to most other activities, because it is often unclear how to assign cardinal values to actions. But the point here was just to get the idea on the table and illustrate that the notion of prime complex action makes sense. It captures, I believe, the sense in which some activities are not just sets of actions, even though they are nothing over and above the actions that compose them. To use G.E. Moore's phrase, they are like organic wholes.[12] And, similarly to elements of organic wholes, each component's contribution to the constitution of the event that is the activity is dependent on the contribution of the others. To fix terms, let us say that:

ACTIVITY: Actions ϕ_1 and ϕ_2 form an activity Ψ if and only if ϕ_1 and ϕ_2 depend on each other for their relation of parthood to the event that constitutes Ψ-ing.

To be clear, we should not expect ACTIVITY to help us spell out necessary and sufficient conditions for when two actions form an activity. The mutual dependence

---

[11] Chant (2010) similarly argues that the Prisoner's Dilemma is a case of what she calls a 'non-additive' effect of a joint action, but provides no general criterion of non-additive effects.

[12] The use of 'organic whole' here is illustrative, but also slightly misleading, since it is less like Moore's own use of the term (which refers specifically to the intrinsic value of wholes) and more like what he referred to (in 1903) as 'its present unfortunate usage' which he claimed had 'no distinct sense'. See Moore (1993, p. 82).

relation it refers to is merely primeness expressed as a relation between actions.[13] But that type of relation is important for our purposes.

Prime conditions are important because of their explanatory power. In the epistemic case, knowledge is said to have explanatory power in virtue of being prime because knowledge requires the interaction between internal and external conditions; the kind of interaction which explains why knowledge is more apt to sustain action over time than composite conditions such as true belief.[14] In our case, the primeness of activities has explanatory power because the performance of activities depends on the interaction between component actions. For illustration, compare:

(1a) Jones takes this step because he is taking this step, and another step, and a third...

(1b) Jones takes this step because he is exercising.

(2a) Jones evaluates his options because he is evaluating his options and making a decision.

(2b) Jones evaluates his options because he is deliberating

It seems (1b) and (2b) are better explanations than (1a) and (2a), respectively. They are both simpler. Since both subsume their counterparts, they are also more parsimonious. And both allow for more accurate predictions: (1a) and (1b) both predict that Jones will take more steps, but (1b) further predicts *how* he will take them, say, at a certain pace. Similarly, (2a) and (2b) both predict that Jones makes a decision, but (2b) further predicts that he will *base* his decision on the evaluation of his options. Primeness of exercise and deliberation would explain this difference.

---

[13] It is important to note that primeness does not rule out analysis as such; what it rules out is straightforward conjunctive analysis. The claims above are thus consistent with the possibility that a relation like mutual parthood dependence can be further analysed. Bader (forthcoming) develops a formal framework for supervenience relations that applies to both classical and non-classical mereologies, one that allows for dependence relations between complex entities and their parts including order-sensitivity, repetition-sensitivity, compositional structure, and the priority of wholes to parts.

[14] Williamson (2000, Sect. 3.4).

Importantly, this difference also applies at the level of reasons. If exercising is prime, having a reason to engage in exercise implies a reason that merely having a reason to perform a series of steps does not: a reason to take them at a certain pace. If deliberating is prime, having a reason to deliberate implies a reason that having a reason to evaluate options and make a decision does not: a reason to base one's decision on the evaluation of one's options. Of course, we could build these relations into the description of each individual action, e.g. by saying that a reason to exercise implies a reason to take $step_n$ at such-and-such a time after $step_{n-1}$ and before $step_{n+1}$. But surely there would be no reason to do that, unless the individual steps depend on such temporal relations to bring about the exercise, that is, unless the whole they constituted were prime.

The idea that the primeness of activities can explain the existence of reasons raises a question about how far this type of explanation can extend. While it is widely agreed that there are reasons for both attitudes and physical actions, it is also widely agreed that there are important differences between these kinds of reasons. Even if reasons for *some* mental states, like attending and deliberating, can be responded to at will, it seems that others cannot. Our beliefs and emotions are not under our direct intentional control in the way that our physical movements are. So the idea that activities could explain reasons for these kinds of attitudes might seem to rely on an implausibly voluntarist conception of our mental lives.

It is a virtue of the present account that it does not imply this type of voluntarism. As illustrated above, ACTION allows that an individual event over which an agent has direct intentional control may partly constitute a type of event over which she does not. Jones may not be able to control whether the act of shooting Simpson amounts to an act of killing Simpson. And neither Al nor Betty can control what

type of action their individual cooperation partly constitutes an instance of mutual cooperation. This is important, because it further allows that whether a given set of individual actions amount to a particular type of activity may also be outside of any agent's *direct* intentional control. And that offers us a compelling compromise between outright voluntarism and the idea that one has some kind of responsibility for one's attitudes. After all, if one had no control *at all* over one's mental processes, it would be mysterious how one could reasonably be held responsible for the states they result in. That the control is there explains how such responsibility is possible. That it is indirect explains why it is significantly different from responsibility for most actions.

3.   'In Virtue of'

While this account of activities does not settle all questions, I hope to have shown that it is well-motivated and independently plausible. And it gives us enough, I believe, to clarify how a solution to the Wrong Kind of Reason problem can be developed which satisfies the constraints identified in Section 1. The basic idea is that reasons of the right kind for a person to respond in some way correspond to the reasons that this person has in virtue of the activities he is engaged in. This raises two further questions. When does a person count as being engaged in an activity? And in what sense do people have reasons in virtue of such engagement?

To answer the first question, contrast two cases. In Case One, Bert is breaking eggs, he has the pan on the stove, everything is ready for him to make an omelet. Bert tries to make the omelet and, happily, he succeeds. In Case Two, everything is as in Case One, except that the future is already determined: Though Bert tries, he will never succeed in making the omelet. Suppose we know all the facts in both

of these cases. In Case One, it would be appropriate to explain Bert's breaking eggs by saying that he is making an omelet. In the latter case, however, this is less clear. After all, it is not true that Bert's actions will ever amount to making an omelet. We would, then, be invoking an explanans we knew to be false. And this would be misleading since true assertions of the form '*p* because *q*' imply that both *p* and *q* are true. For the same reason, I suggest, it would be misleading for us to describe Bert as being engaged in making an omelet.

This point may seem controversial. But this is because it is easy to slide between two different ways of using the same words to describe different events. Ackrill gives a nice example of this.[15] Suppose someone intends to walk from Oxford to Reading, may yet succeed, and is currently going through Wallingford, some distance short of Reading. In such cases, it is often perfectly appropriate to use 'he is walking from Oxford to Reading' to refer to the event that constitutes his walking the distance between Oxford and Wallingford. For this reason it will often be vague whether, in describing someone as 'walking from Oxford to Reading', one is referring to an event that constitutes part of the process or referring to an event that constitutes the process as a whole.

Once we are clear on this, however, it is also clear that, although both uses can be conversationally appropriate, they do not pick out the same token event. So although we can *say* of someone who is breaking eggs that he is 'making an omelet' when we know that will not finish this process, that is because we do not *in this case* invoke the process of 'making an omelet' as a *whole* as the *explanans* for what he is doing. Such conversational appropriateness is compatible with the central point here: To be engaged in any particular activity, not only must one perform some of its

---

[15] Ackrill (2001, p. 152).

constituent actions, it must remain at least *possible* for the others to be performed. That is:

ENGAGEMENT: If x is engaged in Ψ-ing, and Ψ-ing is constituted by $\phi_1, \phi_2, \ldots, \phi_n$, then x performs at least one of $\phi_1, \phi_2, \ldots, \phi_n$ and it is possible for x to perform each of $\phi_1, \phi_2, \ldots, \phi_n$.

This answers the first question.[16] The second question was how reasons for agents obtain in virtue of their engagement in activities. My answer to this question is not uncontroversial, but it does seem to me a natural way to cash out the intuitive idea. Reasons for an agent to do things that are part of some activity she is engaged in obtain in virtue of her engagement in it *because* doing those things promotes that activity. That is, engagement in an activity *explains* reasons in the sense that it is a background condition for a fact about the promotion relation to be a reason. [17] This idea is Consequentialist in a sense, because it holds that the type of relation in virtue of which reasons obtain is a promotion relation. But it is unlike Consequentialism in denying that what should be promoted are overall states of affairs. Rather, the promotion relation obtains between each activity and the individual actions that promote *it*, much in the way some desire theorists claim that the satisfaction relation that grounds reasons obtains between each desire and the actions that would satisfy *it*. More precisely:

REASON: The fact that [ϕ-ing promotes Ψ-ing] is a reason for x to ϕ if and only if x is engaged in Ψ-ing and [ϕ'ing promotes Ψ-ing].

---

[16] It may seem that there is a further question as to why ENGAGEMENT is a conditional rather than a biconditional, but this follows from the primeness of activities. If activities are not reducible to their component actions, even actually performing each of those component actions would not entail that the activity is performed, hence performing one of them while the others remain possible will not entail engagement in it.

[17] See Maguire (2016) for a recent alternative statement of the view that reasons are facts about the promotion relation. See Schroeder (2007) for a similar about the relation between reasons and background conditions.

I accept that REASON is a bold claim. But it is not as bold as it might seem. It is not a claim about what reasons *are*. REASON is merely a biconditional. And it is not even a biconditional that states necessary and sufficient conditions for *all* reasons. It merely says that a certain *class* of facts are reasons for a particular agent if and only that agent is engaged in a certain activity; namely the class of facts that concerns all and only those actions that promote this activity. This is compatible with the claim that there are other kinds of reasons for which these claims is not true. And it is also compatible with the claim that there is some deeper theory of what it is to be a reason. The claim that engagement in an activity is a normative background condition on reasons does not, in principle, rule out the claim that there are further background conditions. So whether there are other kinds of reasons, and whether there is a unified theory of all reasons, will largely depend on how many of our reasons REASON can account for by itself.

For these reasons, I am not going to defend REASON as a unified theory of what it is to be a reason. My claim here is just that it is enough for us to develop an account of the distinction between reasons of the wrong kind and reasons of the right kind. That, I shall argue, is all we need to develop a general account of normative correctness standards, including the particular correctness standards that obligations are. And that is enough, I believe, to suggest that REASON could be a plausible unified theory of what reasons are. But the normative endgame is for another occasion.

What we have before us, then, are four core claims: ACTION, ACTIVITY, ENGAGEMENT, and REASON. Putting those together, my account of the distinction between reasons of the right kind and reasons of the wrong kind is this: When you are engaged in some kind of activity, certain options are open to you. The options that are open to you are events over which you can exercise intentional control. Of

these, some will promote this activity, some will not. For each of the options that do promote it, you have a reason to perform it. But there is a particular property that only some of these options share: They are all mutually dependent for their contribution to bringing about the kind of event that this activity is, that is, for their promotion of it. This shared feature concerning the promotion relation is, I propose, what makes them reasons of the right kind. Where ϕ denotes some act type and Ψ some activity type:

Right Kind: R is a reason of the right kind for x to ϕ if and only if (i) ϕ is part of some activity Ψ in which x is engaged, (ii) R is necessarily a reason for everyone engaged in Ψ-ing to ϕ, and (iii) the fact that they are engaged in Ψ-ing is sufficient to explain why R is a reason for them to ϕ.

The letter of the proposal here is highly similar to the one suggested by above Schroeder.[18] But the content of the core claims that make it up allow us to explain how it actually works.

Let me give an example with respect to reasons for belief and a type of activity that such reasons might be explained in virtue of.[19] It is often said that belief aims at truth. What kind of process promotes the state of believing truly? If one has no beliefs, the process of belief formation. If one already has beliefs, the process of belief revision also. But simply forming or revising beliefs will not promote believing truly. For those processes to reliably result in true belief, they also need to involve some level of attention to one's evidence and existing attitudes.

So, at least two things go into these kinds of mental processes. Let us refer to the one factor as 'evaluation' and to the other as 'attention'. Reliable doxastic processes

---

[18] See fn.5 page 153 above.

[19] This example is modelled on Hieronymi (2009)'s account of critical belief revision as a type of mental agency.

involving these two kinds of mental events are arguably prime, since it is possible for a mental process to involve both evaluation and attention without resulting in the formation of beliefs in the way that reliably leads to believing truly. This can happen, for instance, if one does these things in the wrong order: Attending now to my evidence and mental states, I discover that I have been forming beliefs without attending to the evidence then. Plausibly, that type of process does not reliably lead to true beliefs.

If these doxastic processes are also composed of mental events over which one can exercise some kind of intentional control, then it follows from ACTION that they are actions. And if the processes they make up are prime, then ACTIVITY implies that they are activities. Now consider the class of agents who have the specific ability to exercise any intentional control over their attending and evaluating, and who are doing at least one. It follows from ENGAGEMENT that they are engaged in these activities. REASON then implies that they have reasons to do whatever actions promote those activities.

There are likely to be many such reasons. And, given variations in the background conditions of the environments in which those processes occur, they are likely to vary greatly. For instance, some environments may make available the option of giving myself a drug that somehow prevents me from forming a belief if it is false, and that would provide me with a reason to do so. But whatever the environmental background conditions, only some of these actions are such that they mutually depend on each other for the performance of the activity to occur; only some are constitutive of it. Giving myself this kind of truth-drug is not one of these. So, which facts make up the *right kind* of reasons for those engaged these doxastic activities? For each constitutive action, the fact that doing *it* will promote this ac-

tivity. Because those are all and only those reasons that everyone engaged in these activities necessarily share, in virtue of their engagement in it.

This might seem to rule out many *bona fide* reasons of the right kind. For instance, asking experts for their opinion will not be constitutive of these doxastic activities as I have just characterised them, so the present account seems to imply that those engaged in them could only have reasons of the wrong kind to do so. And it is far from clear that they are. There is a general strategy, however, that a proponent of RIGHT KIND can appeal to in order to show that such reasons may be ruled in.

These reasons, we might claim, are of the right kind relative to a more extensive activity that constitutively includes such doxastic activities as a part. For instance, we might think there is an activity such as 'enquiring' which includes not just attending to evidence, but gathering it, figuring out which is relevant, and so on. It seems to me plausible that this will be true of many kinds of enquiry. Since everyone engaged in such enquiries would necessarily share reasons to consult the experts, those would be reasons of the right kind, relative to that activity. What if the experts are mischievous? Then it seems everyone engaged in enquiry would have reasons of the right kind *not* to ask the experts. Sure enough, and then this would have to be explained by the fact that enquiry is part of some other activity that would be promoted by not asking mischievous experts, and so on.

Naturally, these particular accounts of belief revision and enquiry are not uncontroversial. According to some, for instance, moral enquiry does not constitutively include asking the experts.[20] How do we settle whether some particular type of enquiry involves asking the experts? Well, on the present proposal, the fact that

---

[20] See e.g. Hills (2009).

an expert says so would be a reason of the right kind for belief only if it is the kind of fact that it would necessarily promote a reliable doxastic process to attend to; evidence, that is. Is the fact that an expert says so evidence for a moral claim? That is a substantive question.

What RIGHT KIND predicts, then, is that substantive claims about the particular forms of agency that different types of activities involve have substantive implications for the kinds of facts that will provide people with reasons of the right kind. And if the correct account of correctness standards is CORRECT, then those claims will have implications for the normative standards that govern various types of activities, such as belief formation, or moral enquiry. That seems to me precisely what we should expect. Our account of what belief revision involves, for instance, surely *should* be constrained by plausible claims about the normative standards of belief revision, and *vice versa*.[21]

So I submit that RIGHT KIND provides us with a robust procedure for ruling in those reasons which are, intuitively, of the right kind. But it also provides us with a robust criterion for ruling out reasons which are, intuitively, of the wrong kind. For suppose an evil demon threatens Jane with death unless she believes that he is benevolent. Since believing this will prevent Jane from being killed, believing it will promote any activity of hers that will be promoted by her being alive. Since it is quite plausible that we can find some such activity, REASON predicts that the fact that believing that the evil demon is benevolent will promote this activity is a reason to do so. But RIGHT KIND predicts that this reason is of the wrong kind. For since the evil demon is not threatening anyone else, this is not a reason that Jane shares with everyone else engaged in doxastic activities.

---

[21] The idea that correctness standards of attitudes somehow derive from the nature of those attitudes thus *follows* from this proposal, it is not a *rival* to it, pace McHugh and Way (2016, p. 600, fn. 54) and Sharadin (2016).

Of course, we can easily imagine that the evil demon does threaten all these people. Then they would all share this reason. But that would still not make it a reason of the right kind, according to Right Kind, since they would not *necessarily* share it. It is perhaps less easy to imagine that the evil demon could necessarily provide everyone with such a reason, because then he would have to necessarily exist. But at least some philosophers can imagine this.[22] So we should not rule out of hand that the evil demon might necessarily exist, and that he necessarily threatens everyone on pain of death to believe that he is benevolent. And then it might seem that the reason they would thereby have for this belief would be of the right kind, according to Right Kind.

Now, it is easy for a proponent of Right Kind to reply here that this seems so only if we forget the careful stipulation that reasons of the right kind obtain 'in virtue of' engagement in activities. For while it is true that everyone would necessarily share the reason provided by the evil demon, they would not share it in virtue of engaging in this activity. After all, it seems they would necessarily share this reason whatever they were doing. So this reason would not, according to Right Kind, be of the right kind for a belief. But this will not get a proponent of Right Kind off the hook. For suppose the necessarily existing evil demon necessarily makes the following *conditional threat*: For all agents, if they engage in any doxastic activity, then he will strike them dead unless they believe that he is benevolent. Then it would not be true that everyone would necessarily have this reason no matter what they were doing. In fact, they would have it only if they engaged in doxastic activities. And that certainly sounds an awful lot like saying they would have that reason in virtue of engaging in those activities. But it would still not be a reason of the right kind to believe that the evil demon is benevolent.

---

[22] McHugh and Way (2016, p. 600, fn. 54); Howard (forthcoming).

This conclusion seems to me correct as far as it goes. But it also seems to me to bring out an implicit assumption of the many counterexamples that are often invoked against proposed solutions to the Wrong Kind of Reason problem. Solutions like the one I am advocating here aims to pick out some feature of the world, *F*, which is meant to *explain* why certain reasons are of the right kind (for admiration, say) and thus why they factor into a higher-order normative property (admirability, say) when other reasons do not. The counterargument to any such solution is that the evil demon could always ensure that reasons to admire him obtain when *F* does. As we have just seen, it seems that a necessarily existing demon could ensure that those reasons obtain *if and only if* the *F* features do. And if that is so, then, by assumption, the *F* features are not going to explain the difference between right and wrong kinds of reasons. So any such solution must be bound to fail.

This is a tempting line of reasoning if one thinks that the 'right kind' property is explained by the *F* features by *supervenience*. For then one will be drawn to think that the right kind of reasons are the reasons that supervene on the set of possible worlds in which the *F* features obtain. And since possible worlds are always going to be big enough to accommodate the existence of an evil demon, there is no principled way of ruling out that he could give one reasons to admire him in all of these worlds. But this is simply begging the question against the proposed solution as I just formulated it. Because it holds that the kind of thing which explains why reasons are of the right kind are *features* of the world, not entire *worlds* or sets of them. And it simply is not a compelling objection to this account that we cannot distinguish between the reasons that obtain in virtue of the *F* features and the reasons that obtain in virtue of the evil demon *at the level of possible worlds*.

What we have been looking for all along are the kinds of relations and entities

in virtue of which the reasons that ground obligations obtain. That idea itself involves a commitment to the idea that entities smaller than possible worlds can do real explanatory work, and this is a substantive view about the nature of normative explanations. The point here is just that there are not any *further* commitments we need to take on to be entitled to reject the central premise of all these counterexamples. Even if it turns out that an evil demon can provide reasons with the exact same modal profile as engagement in certain activities, that does not commit us to the claim that the evil demon or the reasons provided by his threats thereby obtain in virtue of anyone's engagement in those activities.

4.   Reasons, Correctness, and Obligations

This concludes my case for the above solution to the Wrong Kind of Reason problem. It is a solution to the problem because it is does not require appeal to any other normative properties than the property of being a reason. Nor does it require giving up the idea that there is something it is to be a reason. And it gives us a robust criterion for distinguishing systematically between the reasons that are, intuitively, of the right kind from those that are, intuitively, of the wrong kind. For a reason to be of the right kind is for it to be necessarily shared by all those who are engaged in some type of activity, in virtue of their engagement in it. And this allows us to state the account of correctness introduced in the previous chapter:

CORRECT: It is correct for x to ϕ if and only if reasons of the right kind for x to ϕ outweigh reasons of the right kind for x to ϕ.

So, for instance, it is correct for everyone engaged in admiring people to be motivated to emulate those they admire. And it is correct for everyone engaged

in getting true beliefs to attend to their evidence. And this is because being motivated to emulate people is constitutively part of the process of admiring them, just as attending to one's evidence is part of the doxastic process which leads to true beliefs. But there are also correctness standards other than those which govern attitudes. For instance, it is correct for everyone who gives a sincere promise to do what they have promised to. As I argued in chapter 1, this is because everyone giving sincere promises have reasons to hold the kinds of attitudes which modify the weight of their reliance-based reasons to do what they promised. They have those reasons because having those attitudes is part of what it is to give a sincere promise. And those reliance-based reasons are also shared between all those sincerely giving promises because they all hold the kinds of attitudes that constitutively involves a responsiveness to reliance; those attitudes that constitute the second-person standpoint.

If this view is roughly correct, we can say something more general about the kinds of correctness standards that obligations are. As this example with promising illustrates, the reasons that constitute each specific type of obligation are identified in virtue of being shared by all those who are party to a specific type of activity. And what those activities all share are the feature that they involve second-personal attitudes. So obligations are simply correctness standards of second-personal activities; that is, the kinds of activities that constitutively involve doing something with another person on the basis of second-personal attitudes. Accordingly, it is actually quite simple to give a non-normative characterisation of the kinds of reasons that ground obligations: They are precisely the kinds of reasons that everyone shares in virtue of doing something with other people.

This point, however, also raises a more specific challenge. Because this type of

account is going to give us a non-normative criterion for identifying these reasons only if there is a non-normative characterisation of what it is to do something with another person, in the specific recognitional sense that makes it qualify as a second-personal interaction. So, in the final part of this chapter, I illustrate how we can arrive at such an account.

To do this, we first need to draw a distinction between two kinds of agency. Consider again the example of mutual cooperation in a Prisoners' Dilemma. When Al and Betty mutually cooperate, they are both doing something individually. But there is also an argument for saying that they are doing something together. The argument is that mutual cooperation is a prime condition, one we cannot analyse simply as the conjunctive effect of two actions performed individually. But think about how the Prisoners' Dilemma is set up: It is not going to be much of a dilemma if Al and Betty have already agreed what each is going to do. The whole point is that they cannot decide what to do individually on the basis of information about what the other does. So there is a substantive sense of 'joint action' here in which it does not seem to follow from the fact that Al and Betty are acting 'jointly' that they are assuming, believing, or intending anything in particular about what the other does.

There is, of course, also another substantive notion of joint agency which does seem to involve such attitudes. Most of the things that we take ourselves to have reasons to do with other people are 'joint' in this richer sense. For instance, most relationships worth having involve doing things with other people, walking together, eating together, and enjoying aesthetic experiences together, and these are rarely activities we take ourselves to share with people regardless of our attitudes towards them. As I argued in chapters 1 and 2, it is constitutive of these kinds of relationships that people engage in these activities on the basis of certain kinds of attitudes.

And those attitudes, I suggested in chapter 3, are all forms of recognition; the kind of attitude we hold towards other people in taking them to be independent agents whose reasons count as much as our own. And my argument there was that we should think of 'taking up the second-person standpoint' as having or forming a standing intention to engage with other people in that way.

A more precise way of stating the idea that obligations are correctness standards of interpersonal activities, then, is to say that they are correctness standards of interpersonal activities in this richer, second-personal sense. So the challenge here is to show that we can give a non-normative characterisation of what *this* kind of joint agency is. And we can do this, I believe, by showing that the central case of second-personal address is itself amenable to such an analysis. For if the account suggested above is correct, the joint activity of second-personal address will stand to more specific kinds of valuable joint activities as the state of knowledge stands to more specific states like seeing, remembering, and hearing: It will be the most general form of these more specific states. And, as I suggested in chapter 3, second-personal address is precisely the type of joint agency that constitutively involves mutual recognition. So what we need to do is to show that this kind of interaction can be cashed out in non-normative terms.

To do so, one needs to answer two questions. First, *what* is it one recognises, in engaging in reciprocal recognition? And, second, what is it to *recognise* something? To this end, I believe we can helpfully consult what Darwall refers to as *Fichte's Analysis*. This analysis, Darwall claims, shows that a commitment to the recognition of another person's dignity follows from the structure of second-personal delibera- tion itself.[23] Central to this analysis is the idea that second-personal deliberation

---

[23] Darwall (2006, p. 244).

involves 'positing' both one's own agency and that of another. When engaged in such deliberation in response to another person's address,

> we are required to posit the free agency of addresser and addressee alike as part of our reasoning. Although agency is no doubt assumed in the background somehow in any deliberation, it is only from a second-person standpoint that the addresser's free agency (and that of addressees) must be posited, that is, brought into their reasoning as a premise.[24]

If I understand Darwall's claim here correctly, the idea is that second-personal address is a form of agency which constitutively aims at doing something in the awareness that it necessarily involves the agency of both another and oneself. As Darwall further points out, it is part of the normative felicity conditions of such an address that it involves reciprocal recognition, that is, that the recognition of this interdependence of agency is itself mutual. To intend to engage in such an address, then, both agents have to intend not merely that the other exercises his agency, but that he does so *on the basis of an intention to engage in this kind of interdependent agency*. Accordingly, second-personal address involves mutual awareness of interdependence, both at the level of action, *and* the level of attitudes. And that interdependence must, as Darwall puts it, feature *as a premise in practical reasoning* for both parties.

Now, as I introduced the term in chapter 1, to rely on a proposition is simply to use it as a premise in one's reasoning. So in our terminology, Darwall's claim here seems to amount to the claim that both parties must *rely* on each other, both at the level of action, and at the level of attitudes. For practical deliberation to be second-personal, then, is for it to involve a type of intention which posits reciprocal pairs of a particular relation: reliance. And given that awareness of such intentional

---

[24] Darwall (2006, p. 256).

states is a factive mental condition, this is to say that the agents have to act in the *knowledge* that those conditions obtain.

The important point here is that *if* this analysis succeeds in the way Darwall intends it to, then both the form and content of this distinctively second-personal kind of 'recognition' will follow from these purely structural features of second-personal deliberation. For if the structural features of second-personal address really are as the above construal of Fichte's Analysis holds them to be, it turns out that the *attitudes* involved in an instance of second-personal address can be schematized as follows:

(1) A intends that A and B *J*

(2) A relies on B to do B's part in *J*'ing.

(3) A relies on B to intend to do B's part in *J*'ing.

(4) B intends that A and B *J*.

(5) B relies on A to do A's part in *J*'ing.

(6) B relies on A to intend to do A's part in *J*'ing.

(7) (1)-(6) are common knowledge between them.

What follows from Fichte's Analysis, then, is that an interaction is based on a second-personal attitude of recognition *if and only if* it satisfies these conditions. That is the *form* of an intention to engage in such an interaction. This is of crucial importance. Because if it is correct, it means that the form of the intention that characerises second-personal interaction is just the form of what philosophers like to call a 'shared intention'.

Here is an argument for this claim:[25] Suppose Dan and Nancy would like to do some dancing together. To do that intentionally, each of them must intend individually that they dance. That implies that each of them also needs to intend to do their bit in their joint dancing. But this is not quite enough. For instance, suppose Dan wrongly thinks that Nancy is under the spell of an all-powerful, dance loving demon. Dan thinks Nancy will do the locomotion, but that is because he thinks the demon will make her, not because Nancy intends to. Dan does not think Nancy has any kind of intentional control over her body at all. And then Dan cannot intend that they together do something which involves each of them intentionally dancing, because that would at least require thinking that Nancy intentionally dances.

So their intentionally dancing together depends on Dan not thinking that something like this is the case. But it is not just going to be enough for Dan to *believe* that Nancy's moves are under her intentional control either. After all, Dan could believe *that* without intending to dance with her at all, for instance if he was merely intending to dance next to her. And he could do that while thinking that Nancy was really just intentionally ignoring him. So Dan is at least going to have to *rely* on Nancy to do her bit. Still, this is not quite enough. For Dan might also think that Nancy is making all the right moves because she is just practising her own moves. For instance, Dan might think this and intend just to dance alongside her while she is practising, *as if* they were dancing together. Then he would rely on Nancy to do her bit in bringing about a kind of joint performance, but not because she intends to do *that*. So Dan is also going to have to rely on her to intend to make her moves as part of *their* dancing together. Of course, all of this applies to Nancy as well. And

---

[25] I am building here on the account in Alonso (2009) who argues persuasively that it captures the most plausible elements of standard accounts of intentional joint agency.

if both of them are literally going to *see* each other as contributing in these ways to their joint dancing, both are going to have to know that all of this is the case.

If these points are correct, Fichte's Analysis supports the claim that the structure of second-personal deliberation matches the structure of deliberation about intentional joint agency. And what is constitutive of such agency is that it involves mutual reliance. This is good news for something like the Reliance View. For if 'taking up the second-person standpoint' is simply a matter of having an intention with that form, it follows that it is constitutive of taking up the second-person standpoint that one is responsive to the reliance relations it constitutively involves. For to take up the second-person standpoint is to take it to be a fact that doing something another person relies on one to do will promote an activity that one is engaged in with them. According to REASON, that is precisely the kind of fact that is fit to be a reason. And according to the Reliance View, that is precisely the kind of reason that make it the case that people are obligated to each other.

Is this a non-normative analysis of second-personal address and the attitude of recognition it involves? Well, clauses (1)-(7) that specify the form of shared intentions seem to me not to involve any normative terms, or appeal implicitly to any normative properties. If this construal of Fichte's Analysis is along the right lines, then, deliberation based on that kind of intention is second-personal. And if Darwall is correct, it is precisely the structure of that intention which explains why second-personal deliberation involves a recognition of another person's *dignity* and therefore amounts to a form of *respect*.

So if REASON is true, then it will also be true that whenever one engages second-personally with people, one does have a reason to do what others rely on one to do. And, if the Reliance View is true, it will also be true that there is no way to

engage second-personally with other people without being obligated to them. If one understands what it is to engage second-personally with others, it will even be true that one cannot rationally do so without recognising the normative force of such obligations. Together, these claims imply that second-personal interaction is an inherently normative phenomenon. The point here is just that second-personal interaction does not involve these normative features as a matter of its *analysis*. The analysis itself can be given in wholly non-normative terms.

Moreover, this analysis has a number of virtues. It explains why it should seem so plausible that recognition is a mode of valuing or caring about persons that is different from merely valuing or caring about objects. Unlike objects, people are agents, and recognition differs from valuing mere objects precisely in constitutively positing the independent agency of other persons as a premise in one's own agency. It is in this way that recognition involves seeing someone 'as a person', rather than just as an object useful for the promotion of other things one values.[26]

The idea that taking up the second-person standpoint is a matter of holding a certain type of standing intention also gives us a plausible account of the relation between reasons and motivation. Recognising another person does not merely involve registering facts about how he relies on one, or how doing what he relies on

---

[26] This point also seems to me to allow for the existence of obligations concerning our treatment of individuals who do not have the requisite cognitive abilities to qualify as 'persons' in the technical sense introduced in chapter 3, as in the case of some animals, very young children, and those with severe mental impairments. Insofar as these individuals are nevertheless *agents*, recognition of this fact would require positing their agency in one's reasoning and this would help explain why the normative standards concerning our treatment of them should take the character of obligations. Indeed, it is consistent with the RELATIONSHIPS thesis from chapter 2 that we can share relationships with such individuals in virtue of our attitudes towards them. Conversely, the fact that they lack the capacity for similar reasoning would help explain why both these kinds of relationships and the obligations they involve are different in kind from those I have referred to as 'second-personal'. Such an account would go together with the fact that these kinds of individuals are not typically regarded as subject *to* obligations concerning others, even if they are subjects *of* obligations that apply *to us*. Regrettably, I must set aside further discussion of this question here.

one promotes one's activities. It involves the particular taking-it-to-be-a-reason attitude towards such facts. But on the widely accepted view that motivation is built into the structure of intention itself, it is no wonder that taking up the second-person standpoint should motivate us to do what we are obligated to do. And if the Reliance View is correct, it is not mysterious either why those taking up the second-person standpoint would be motivated *by the reasons that obligate them* to comply with their obligations.

Finally, conceiving of the second-person standpoint in this way allows us to pin down more precisely the nature of the reasons that we might have for taking it up. We can do this, I believe, by looking at how the content of the intention itself explains why there are reasons for people to form an intention with this particular form. As I have stated it above, the generic form of the content of this type of intention is simply that two people *J*. For this analysis of intentional joint agency to avoid circularity, however, *J* cannot itself be specified as an intentional joint action. How, then, should we specify *J*?

Well, suppose that *J* was simply a set of two individual actions.[27] For instance, suppose Dan and Nancy intended that Dan dance in the morning and Nancy dance in the evening. If they did as they intended, Dan and Nancy would together bring it about that [Dan dances in the morning and Nancy dances in the evening]. In this case, Dan and Nancy *could* form a shared intention with this content. But it is not clear why they would each have a reason to do this *rather* than forming an

---

[27] The tendency to think of the objects of shared intentions in this deflationary sense is endemic in the literature. For instance, Christopher Kutz tells us that the objects of shared intentions are 'collective ends', by which he means 'the object of a description that is constituted by or is a causal product of different individuals' acts' in Kutz (2000, p. 10). And Michael Bratman holds that an even more minimal notion will do: 'We have, for example, a concept of our painting the house that involves only the idea that, roughly, we are each intentionally painting that house in ways that avoid collisions. We then use such neutral concepts in the contents of the intentions involved in our construction of initial cases of shared intention' in Bratman (2009, p. 47).

individual intention to do their own bit. To do *his* bit, Dan need not rely Nancy to do anything and so he need not, *inter alia*, rely on her do anything on the basis of a particular intention. And recall the point from chapter 1, that one relies on something to happen only if one does not intend to bring it about. So the point of relying on Nancy cannot be to bring it about that Nancy does her bit. In this case, it seems Dan has as much reason to form an individual intention to do his bit and simply cross his fingers that Nancy will do hers.

It is tempting to reply here that if Dan did not rely on her in these respects, the object of his intention could not be a joint action. But recall that we cannot appeal to intentions in specifying the content of *J* itself when we are giving an analysis of shared intentions. This would be circular. So if the objects of shared intentions were simply composite actions, it would be difficult to explain why shared intentions would even *seem* to require attitudes of mutual reliance. This is not difficult, however, if the objects of shared intentions are joint actions in the sense I introduced above. For if the object of one's intention is to do one's bit in a *prime* action, then whether or not one succeeds in doing one's own bit *does* depend on what the other does. This is the sense in which ACTIVITY specifies that individual actions are mutually dependent. Thus, we can explain why there are reasons for people to form intentions of this particular kind by pointing to do the contents of those intentions themselves. They would have those reasons if they had reasons to be engaged in joint, prime activities.

This, I believe, makes it considerably more plausible that we can find some reasons to take up the second-person standpoint without characterising those reasons in terms of second-personal or other normative notions. Of course, if our aim is to show that everyone is subject to the normative requirements of obligations, we will

also need to show that everyone has *most* reason to be engaged in these kinds of joint activities. This would be an important result, not least because it would make it much more plausible that our reasons to take up the second-person standpoint are going to turn out to be of the right kind. After all, the idea that we all have reasons to intend to do what we have most reason to do has at least *some* plausibility to it. And if we all shared reasons to hold this kind of intention, in virtue of our engagement in some activity, then those reasons would, according to the view I have defended here, be of the right kind to hold it. So there might be non-second-personal reasons of the right kind to take up the second-person standpoint after all. It might be correct for all of us to do so.

Chapter 6

Interpersonal Agency

1.   End of the Line

Now that we are reaching the end, let us remind ourselves how we got here. We started out in chapter 1 with a programmatic statement of Pragmatism. It holds that facts about obligations are really just facts about reasons. They are a particular type of correctness standard, fixed by how certain sets of reasons weigh against each other. In identifying the nature of those reasons, Pragmatism is motivated by the particular 'bipolar' features that many obligations have. They obligate one person *to* another because they involve some kind of normative relation between two people. This, I suggested, is the relation of reliance.

In chapter 2, I defended in further detail this Reliance View of obligations, arguing that it gave us a plausible account of both promissory and special obligations that reflects the normative significance of relationships and the way this is reflected in the attitude of trust and blame. In chapter 3, I then argued that this account could be extended to give us a plausible account of all obligations. And I argued for that by defending an account of a maximally extensive moral relationship, one which involves default attitudes of trust and reliance towards everyone. If one holds the attitudes constitutive of that relationship, one will take as reason-giving the facts

that, according to the Reliance View, make it the case that one is obligated to other people.

In chapter 4, we saw that Pragmatism's commitment to the normative fundamentality of reasons raised some concerns about its ability to give a satisfactory account of reasons to hold such 'second-personal' attitudes. And I argued that a failure to address this Wrong Kind of Reason problem forced a proponent of Pragmatism to accept some rather uncomfortable conclusions. So, in chapter 5, I addressed that problem by defending a solution to it. According to that solution, reasons to hold some attitude are of the kind that make it correct to hold that attitude just in case they are reasons shared by everyone engaged in some activity. And that showed us how we might, on Pragmatism's premisses, defend the claim that it is correct for everyone to hold those second-personal attitudes that constitute the moral relationship. For this would be correct, I argued, just in case it would be correct for them to have an intention which constitutively involves the relation that, according to the Reliance View, grounds obligations: reliance.

We arrive now at the final chapter, the purpose of which is to fill in those details that will take us the final step towards Pragmatism as I stated it in chapter 1. The preliminary steps are these: I have argued that shared intentions constitutively involve reliance, because they are intentions to do something for which the individual agency of two people is mutually dependent. So to show that it would be correct for everyone to have an intention with that content, we need to show that it would be correct for everyone to be engaged in that kind of activity. And to show *that*, on the account of correctness in chapter 5, we need to show that they are already engaged in some other type of activity in virtue of which they have most reason to do that. So that is the aim of this chapter.

Before doing so, however, we will need to look at some objections that purport to show that these claims can never be made to work because the account of correctness here is fundamentally mistaken. In Section 2, I address two objections that each purport to show that reasons, even reasons of the right kind, could never make the attitudes we have been discussing 'fitting' or 'correct'. In Section 3, I address another objection, that even if reasons could ground correctness standards, those reasons could not come from facts about agency; normativity has to come from somewhere else. If either of these objections are correct, we cannot possibly take the final step, so that is why we need to address them first.

Hoping that I have done enough to make you reasonably sympathetic towards Pragmatism, I hope you will be relieved to find that these conclusions are not correct. The final step is at least possible. It is also, however, a slightly complicated step, and so I break it up into a series of smaller ones. I first show, in Section 4, how to defend the claim that anyone would have reasons to engage in joint activities, whatever they were doing. Together with the account of correctness from chapter 5, this allows us to argue that the existence of these reasons ensures that the second-person standpoint becomes normatively inescapable once taken up. This result is enough to argue that Pragmatism has the resources to defend an account that is *at least* as plausible as those discussed in chapter 4.

The reason I think Pragmatism is an attractive view is that I think it also has the resources to do a little bit more. So, in Section 5, I propose an account of the weight of reasons which, together with the claims of chapter 5, implies that anyone has most reason to engage in joint activities, and therefore take up the second-persons standpoint, whatever they are doing, even if they do not already hold any second-personal attitudes. Finally, in Section 6, I address a few remaining concerns. I trust

this account will not be uncontroversial, and more will need to be said about it than I can do here. But I shall argue that it is independently plausible, and that it is enough to show that the account I have been defending can make good on the claim that our obligations are ultimately grounded in facts about the basic interdependence of our activities, or 'interpersonal agency' if you like. And that makes it deserving of its full name, Deontic Pragmatism.

2.   Strictness and Gradability

One objection to the account of correctness defended in the previous chapter starts from some observations about the way that reasons work. Suppose you can walk into a shallow pond only by getting your fancy new clothes all dirty. That is a reason for you not to walk into the pond. Suppose this is the only reason you presently know of. Can you infer from this that it would be correct for you not to walk into the pond? It is plausible that you cannot because you may be ignorant of other reasons. If, for instance, a child were drowning in that pond, then that would be a reason for you to walk into the pond and help. That reason would be weightier, and so it would plausibly be the case that it would be correct for you to walk into the pond in this case. The reason you cannot know this just by thinking about one of your reasons is that no reason can determine this individually. One reason may be enough to make it the case that it is correct to do something, but that will usually be due to the absence of other reasons. Individual reasons contribute to determining such facts, but they do not settle them.

When several individual reasons count in favour of the same action, their weights typically combine. When they count in favour of different, incompatible actions, they weigh against each other, or compete. When all these reasons are in, we can

settle what it would be correct for you to do by determining how all these reasons combine and compete against each other. But according to Barry Maguire, that is not how facts settle which kind of attitude it is correct for people to have, at least not *affective* attitudes like the ones we have been discussing: trust, blame, valuing, respect, and so on.[1] Maguire's first argument for this claim is that facts provide normative support for affective attitudes individually and without competing. And he argues that they should compete, if they were reasons.

Maguire gives the following example: Suppose your good friend undeservedly got a job that you were a much better candidate for. It is fitting to be pleased that a friend had good fortune, it is fitting to be frustrated that one did not land a much desired job, and it is fitting to be indignant that someone got a job they did not deserve. But it is not as if it would be fitting *only* to feel pleased, or frustrated, or indignant in this case. The fitting response is just to have all those attitudes to obtain in one complex, ambiguous whole. Does that not sound right? I think it does. But if these facts provide normative support for a complex whole, then they do not provide normative support for incompatible attitudes; they do not compete. And if we can know that each of these facts individually make each part of this complex fitting, then they are not contributory. So they cannot be reasons, because reasons compete, and they are contributory. Does that not sound right? Not to me.

I think there are two reasons why this is argument is misleading. First of all, what is needed here is a principled argument to the effect that facts providing normative support for affective attitudes *cannot* compete and that they *cannot* be contributory. For if any facts providing normative support for affective attitudes were *ever* to compete or contribute, they would, by the lights of the argument here, be

---

[1] Maguire (forthcoming).

reasons. And then it would not be true that reasons cannot make attitudes fitting, or correct. Are there any such cases?

Well, recall Mary from our discussion of the Wrong Kind of Reason problem. Mary is steadfast, considerate, and wise, and she is neither cruel, dishonest, or the like. It is clearly correct to admire Mary in this case. But suppose Mary *were* dishonest, indeed *very* dishonest. If that affects your judgment of whether Mary is admirable, then it should seem like a fact that competes with the other ones in settling whether or not Mary is admirable. And insofar as each of them appear to be relevant with respect to determining whether her dishonesty decisively counts against admiring her, they should seem like facts which contribute to making Mary admirable. The fact that in some cases such facts do not compete or contribute simply does not show that a case like Mary's is not possible. At best, Maguire's case suggests that there is such a principled argument to be had which will show that our intuitions about Mary are misleading.

I believe, on the contrary, that we are mislead by an assumption which leads to the conclusion Maguire draws from his example. For consider: How do we get from the claim that the facts in his case do not compete to the claim that they cannot be reasons? The explanation cannot be that one could not respond to these facts with such a complex of attitudes. If the facts make a complex of attitudes correct, that complex of attitudes is possible to hold by assumption. If those attitudes are jointly compossible, why would the facts supporting them have to compete to be reasons? Surely not simply because the attitudes are in tension with each other. The corresponding claim seems to be implausible about reasons for actions. You might have one to reason to sing Love Me Do at 9am sharp and another to play Rachmaninov's second piano concert at 9am sharp. That is hard, the tunes do not

sit well with each other, but it does not follow that these reasons could not together make it correct for you to do both at 9am sharp.

What else might the explanation be? It *would* be an explanation if there were only one correctness standards for attitudes for these reasons to fit into. But that is precisely what I have been arguing is *not* the case. Different attitudes have different correctness standards, because reasons are always of the 'right kind' relative to different activities. Indeed, I spent most of Section 5.1 arguing that this sensitivity to variations in background conditions was an inherent feature of the 'right kind' property that any plausible account of the distinction between right and wrong kinds of reasons should account for. To draw the conclusion Maguire does, we have to assume away that possibility. But if we grant that possibility, it is no wonder that we can find cases where individual facts seem to make affective attitudes correct all on their own. If some reason is the only one that is of the right kind relative to the correctness standard for indignation, say, that reason will make indignation correct, no matter whether other reasons make some other attitude correct, say, feeling pleased.

Maguire's second argument that reasons do not make affective attitudes correct relies on the claim that these attitudes are gradable. One can be more or less sad, for instance. To appreciate the force of this argument, consider the following analogy: Suppose one act is *more* wrong than another. On a simple buck-passing account of the correctness standard of wrongness, if an act can have this property to a higher degree than some other act, it seems this must be because the total weight of reasons not to do this act is greater than the total weight of reasons not to do the other. As we can say, there is *more* reason not to do it. This line of reasoning seems intelligible. Applied to normative support for attitudes, however, it seems to locate

the gradability in the wrong place. To see this, let 'n-support' be a neutral term for whatever type of normative support facts provide for attitudes. Now compare:

(a.1) That A's death is *more* regrettable (than B's) n-supports feeling *more* sad about A's death (than B's).

(a.2) That A's death is *more* regrettable (than B's) *more* n-supports feeling sad about A's death (than B's).

Maguire holds that such pairs of claims are problematic for those who believe affective attitudes are supported by reasons. Whereas (a.1) seems perfectly natural, (a.2) appears infelicitous. And the problem here is not merely to do with the surface structure of English. For the buck-passing view seems to predict that the *property* of it being correct to feel more sad reduces to the *property* of there being more n-support for some generic degree of sadness. So, while this view should be able to explain how something like (a1) could be true, it seems that it can only explain how something like (a2) could be. So, it seems not to be able to explain how it could be correct for people to feel more or less sad, and that is clearly a major lacuna for any view on the nature of normative support for *affective* attitudes.

This, I believe, is forceful objection. But I also think the account I have defended allows us to explain why it merely looks like an objection, even though it is not. As I shall now argue, claims like (a2) have genuine, substantive content. Moreover, I shall argue, it is the content of claims like (a2) which explains why claims like (a1) are true, when they are. To do this, I will draw on the account of blame I have already defended in earlier chapters. This should be no problem since 'blame' is an affective response, and the property of being blameworthy, I have argued, is normative in the sense that it is a property people can have only if they have other properties that provide reasons for people to blame them. Now consider,

(b.1) That A is *more* blameworthy (than B) n-supports

  fill-in-the-blank more (than B).

The buck-passing view of the normativity of reactive attitudes I have defended so far predicts that what we should fill in for *fill-in-the-blank* in (b.1) is *blame*. So (b.1) should become:

(b.2) That A is *more* blameworthy (than B) n-supports

  blaming A *more* (than B).

Note that the support relation here might be very weak. It does not follow from the fact that someone is blameworthy that the normative support for blaming them is strong, or that it amounts to a requirement to blame them. It merely follows that there must be *some* normative support for blaming them. On the view I have defended since chapter 1, the blaming response that there is such support for is a complex attitude, one which involves judgments that certain reasons have been disregarded, negative emotions, withdrawal of trust, and revision of intentions. The key point here is that three of these four types of responses are *not* gradable. Thus:

(b.3) That A is *more* blameworthy (than B) n-supports

  # judging *more* that A is blameworthy (than B).

    judging that A has disregarded reasons *more* (than B).

    judging that A has disregarded *more* reasons (than B).

  # withdrawing trust *more* from A (than B).

    withdrawing *more* trust from A (than B).

\# revising intentions *more* with respect to A (than B).

revising *more* intentions with respect to A (than B).

\# feeling negatively *more* towards A (than B).

feeling *more* negatively towards A (than B).

The prefix '#' indicates an infelicitous sentence. The statements thus prefixed here are infelicitous because the attitudes they are about are non-gradable. One cannot judge that *p* more or less, one either judges that *p* or one does not. Similarly, one cannot withdraw trust that a person will do something more or less, one either withdraws trust or one does not. And one cannot revise an intention more or less, one either revises it or one does not.

There may be such things as partial intentions, for instance. But to have a partial intention is not to intend something to a lesser degree, it is to hold a different type of all-out attitude than an intention.[2] And while it may be vague whether one intends that *p* or only partially intends that *p*, vagueness is not gradability. It may be vague whether something is true, but truth is not a gradable property. Where these kinds of responses admit of degree is in their content. One may judge that someone disregarded more reasons because the number of reasons he disregarded was greater. Or one may judge that he disregarded reasons more because he was in a better position to respond to them. One may withdraw more trust in the sense that one trusts in fewer domains of interactions. And one may revise more intentions, either in the sense that one revises a greater number of intentions, or that one revises more fundamental or encompassing intentions.

---

[2] See Holton (2014) for an argument that extends this claim to other attitudes, such as belief.

Now, recall that like judgments and intentions, trust takes propositional objects. And, I have argued, to be blameworthy is to be such that it would be reasonable for people to revise such attitudes. So we might take the fact that someone is *more* blameworthy as saying that one has n-support for forming or revising attitudes of certain kinds with respect to a greater number of propositions. In turn, this means that one will have n-support for forming or revising a greater number of individuated attitudes. To revise the judgment that p and q is also to revise the judgment that p and the judgment that q. Let β denote this class of propositional attitudes that blame involves a revision of. Then we get:

(b.4) That A is *more* blameworthy (than B) n-supports

revising *more* β attitudes with respect to A (than B).

Of course, it is arguably not true that to blame someone more is *merely* to revise more propositional attitudes. For instance, suppose I now blame you for something and revise all my intentions to interact with you. It is arguably not true that if I had formed a few more intentions to interact with you a moment earlier, then I would now be blaming you more. What does seem to be true is that the degree of blame corresponds to the number of *dominant* intentions; that is, intentions that are not merely sub-intentions of a more comprehensive intention. And it is not difficult to see which kinds of dominant intentions are relevant to blame on the view I have defended. They are intentions to share substantive relationships with people that involve second-personal attitudes.

Consider the case of Bert and Ernie, for instance. They are friends, but Ernie has not been nice to Bert recently. Bert does not like that, naturally, and he blames Ernie for it. Now whether Bert revises his intention to remain friends with Ernie seems

to make a difference to how intensely he blames him. Of course, if Bert and Ernie were not already friends, he would not have been able to revise that intention. But he might instead have revised a standing intention to be friendly with Ernie. And whether he did, it seems to me, would still make a difference to how intensely we should say he would be blaming Ernie. But to revise such intentions, he will also have to revise all of his standing sub-intentions to do things with Ernie that involve being friendly. And if he was so angry with Ernie that he intended not to interact with him at all, he would have to revise even more such sub-intentions.

Now suppose it would be *correct* for Bert to blame Ernie more. Then there would be n-support, for instance, for revising a more dominant intention. Then there would be n-support for forming or revising a corresponding set of sub-intentions. And then there would have to be derivative n-support for forming or revising each of those attitudes individually. The more dominant the intention, the greater the corresponding set of derivative n-support relations for revising individual sub-intentions must be. Since that will affect the extent to which Bert relies on Ernie, and since reliance is a condition on trust, similar remarks apply to withdrawing trust. This means that there will be more individual relations of the kind 'n-support' for the kinds of attitude revisions involved in blame, 'β attitudes' that is. So we get:

(b.5) That A is *more* blameworthy (than B) *more* n-supports

revising β attitudes with respect to A (than B).

Now it seems to me that there are two sources of resistance to this move. The first is that the notion of support involved might seem spurious. It might be argued that when we say that some fact supports an attitude more, what we mean is not that it provides a greater number of support relations for a greater number of attitudes.

Rather, it provides a single source of support for more of something. But that idea is clearly not forced upon us by the notion of support itself.

Imagine a group of builders trying to prevent the impending collapse of a bridge. 'More support!' one of them shouts to the engineer. This engineer could intelligibly understand the builder to say that he wants a greater number of supporting beams for the bridge. It is a substantive question whether this *is* what the builder means, but it is a meaningful interpretation of what he is saying. That is the kind of interpretation offered here. Nor can the interpretation be forced upon us because the thing supported could not be made up of things that are each individually supported by a separate support beam. The bridge is made up of individual stones, each of which might be supported by a support beam that must be carefully coordinated with those supporting other stones. That would also be similar to the interpretation of blame I have offered.

The second source of resistance is to the idea that we can pass the gradability to normative support rather than the response supported by it. But note that, so far, we have been passing the gradability to the support for *non-gradable* attitudes involved in blame. The 'β attitudes' here are non-gradable propositional attitudes: judgments, trust, and intentions. If these attitudes are not gradable, it cannot be an objection to this account that it locates the gradability somewhere else. Of course, I have also been arguing that blame also has an emotive component, a 'negative feel' and that *is* plausibly gradable. So the objection Maguire has in mind must be that *this* is what cannot be accounted for on the view that reasons provide the normative support for affective attitudes like blame.

I believe it can be accounted for on this view. To do so, we can appeal to the notion of *resonance* discussed in chapter 2. As Kolodny expresses this idea, one's

reactive attitudes towards people should resonate with how they treat the things one cares about. If someone willfully destroys your much cherished vase, for instance, your reactive attitudes ought to reflect this. It is something to feel resentful about. What I have argued is that the reason *willfully* destroying your vase (as opposed to accidentally destroying it) makes this kind of reaction appropriate is that the action is thereby based on attitudes that impair a relationship you have reasons to have with them. In the case of blame, then, we can appeal to the idea that it is the *degree of impairment of such a relationship* that your negative emotions should resonate with.

The *degree* to which a person's attitudes impair a relationship is arguably a function of several factors. There is the extent to which their reasons to have such attitudes were weighty, the extent to which they were in a position to respond to those reasons, and the extent to which their actual attitudes match the attitudes they have reasons to have. Precisely how this works is not important here. Insofar as it is correct for your emotions to resonate with the state your relationships, it is equally plausible that the *intensity* of the negative emotion you have a reason to feel is a function of the degree to which such a relationship is impaired.

This is important. For it means that the explanation for the intensity of the emotions you have reasons to feel *matches* the explanation for the extent to which you have reasons to revise other types of attitudes, like trust and intentions. And that further explains why there is an important kind of internal *coherence* in the complex of attitudes you have reasons to respond with when other people are blameworthy. If that is right, the fact that you have more n-support for revising attitudes like trust and intentions implies that you have n-support for feeling blame more intensely. In other words, your reasons to revise attitudes can fully explain why, when some-

one is more blameworthy, there are reasons for you to blame them more. Rather than constituting an objection to the idea that reasons can make affective attitudes correct, the phenomenon of gradability illustrates its explanatory power.

3.   The Shmagency Argument

Pragmatism holds that the reasons which make attitudes correct, not least affective attitudes, are those that are shared by everyone engaged in some activity, in virtue of their engagement in it. And reasons, I have been assuming, are normatively fundamental. This view is therefore part of a broader class of views according to which normativity is somehow grounded in agency. How might we defend the claim that there are reasons and normative standards that apply to everyone in virtue of their engagement in some activity? According to so-called Constitutivist views of normativity, this follows from the fact that agency itself has a constitutive aim.

Christine Korsgaard, for instance, holds that acting is a matter of constituting oneself, and substantive norms of agency apply to all agents, she argues, simply in virtue of the fact that self-constitution can be more or less succesful. Building a house poorly, for instance, does not imply that one is not building a house at all. And if one is building a house, then substantive standards of house building apply; that the roof should not leak, that the walls should not cave in, etc. Similarly, bad actions are 'bad' in the sense that they fall short of the standards of self-constituting agency. The difference, Korsgaard claims, is that action, unlike house building, is something we are condemned to. The constitutive norms of agency are thus inescapable for anyone who is an agent.[3]

---

[3] Korsgaard (2009, pp. 1, 24, 32). In Velleman (1989), David Velleman defends a similar view, according to which agency constitutively involves an aim of being intelligible to oneself, one that in turn grounds normative standards. Velleman's view, however, is complex and has evolved consid-

This type of Constitutivist argument seems compelling partly because it makes it easy to understand how genuine normative standards might apply to everyone. The problem is that it is questionable whether agency is really inescapable in the sense that the argument requires. If one wishes to stop acting, there are certainly things one can do to make that happen. One can commit suicide or go to sleep, for instance. Granted, these are things one must necessarily *do*, but that shows little. Matthew Silverstein gives a nice example to illustrate the point:[4] If you are in your office, the only way you can leave your office is by doing something *in your office*. But it does not follow from this that you could not possibly leave your office, you can just walk out the door. Your office and its norms are not inescapable. Why should agency and its norms be?

David Enoch argues that the mere intelligibility of this question undermines the Constitutivist's position. For if agency is not inescapable, it is an intelligible question whether one has reasons to engage in agency in the first place. That is a normative question. So even if the Constitutivist is right that agency involves constitutive reason-giving aims, there is a further normative question that the Constitutivist must answer. And the answer to *that* question, Enoch thinks, cannot be that those reasons are inescapable because agency is, for then the question the Constitutivist must answer would not be intelligible in the first place. And this, Enoch holds, will be the case no matter what the correct account of agency is.[5]

Enoch provides an example to illustrate the point. Suppose you and Jim are sitting in front of a chess board when he suddenly castles out of check. When you

---

erably over the years (see Velleman (1996); Velleman (2000); Velleman (2009)), and a sympathetic statement of it would require more care than I can give it here. Some have also taken Connie Rosati to defend a Constitutivist view, but Rosati has since questioned whether this is a helpful characterisation of her view, see Rosati (2003); Rosati (2016).

[4] Silverstein (2015).
[5] Enoch (2006).

tell Jim that he cannot do that, his response is 'So what?' Suppose you tell him that, if he does that, he is not really playing chess anymore, because one can only play chess if one abides by the rules of chess. Jim does not care. 'Fine' he says, 'I'm not interested in playing chess, I'll call it 'schmess' instead'. It seems intelligible that people could play a game like this; similar to chess, but subject to different rules. So, to argue that the rules of chess *rather* than those of schmess apply to Jim, it seems you have to assume that Jim has a reason to play chess rather than schmess.

Similarly, we might imagine someone who fails to abide by the constitutive norms of agency, whatever they are. This person might respond to a similar type of critique by saying 'Fine, I'm not interested in agency, I'll be a shmagent, engaged in shmagency, the kind of thing that involves shmactions'. If the constitutive norms of agency are to apply to this person, there must be a reason for him to engage in actions rather than such shmactions. But *that* reason, it seems, has to come from somewhere else. Call this the Shmagency Argument.

One response to this argument is to point out that normative explanations always have to end somewhere. A reductive explanation of what it is to be a reason, for instance, is bound to end up saying that for r to be a reason for x to ϕ is for some kind of relation R to be obtain between r, x, and ϕ. For such an account to be coherent, there cannot always be a further question whether there is any *reason* for x to stand in this relation to ϕ-ing. Standing in the relation *is*, on any such view, just what it is for x to have a reason. So the question really only arises if one assumes from the beginning that such accounts are not coherent.[6]

Note, however, that it will not do for the Constitutivist here to assume that *some* such account is correct. It is no help to him if the correct reductive account of

---

[6] This is the reply to the Shmagency Argument given by Silverstein (2015). Schroeder (2005) defends a similar account of normative explanations.

reasons is desire-based, for instance. To defend *Constitutivism* from the Shmagency Argument, he has to assume that a Constitutivist account of reasons is correct. Now this is not ruled out on the account I have defended. For according to the thesis I called Reason, something is a reason for an agent to ϕ if and only if ϕ-ing promotes some activity in which he is engaged. So to avail ourselves of this response, we might simply claim that this is what it *is* for something to be a reason. But you will have noted that Reason is just a biconditional. It does not actually say what it is for something to be a reason, and it is, in principle, compatible with a more fundamental explanation of why that biconditional is true.

For this reason, it seems to me a weak response to the Shmagency Argument to assume claims about what it *is* for something to be a reason and then point that, if this assumption is true, the argument has no force. If reasons are normatively fundamental, this is simply to assume a reductive theory about all of normativity, and such a theory does not come cheap. It seems to me a better response if we can defend ourselves against the Shmagency Argument without assuming this much. And I believe we can. For as both Velleman and Silverstein point out, Enoch's central question is actually ambiguous.[7] Consider the formulation in this passage:

> Classify my bodily movements and indeed me as you like. Perhaps I cannot be classified as an agent without aiming to constitute myself. But why should I be an agent? Perhaps I can't act without aiming at self-constitution, but why should I act?[8]

This way of phrasing the question obscures whether it is an agent or a shmagent asking the question. And that, I hope you can see, makes *all* the difference. For suppose an agent wants to know if he has a reason to act. If Reason is true, then

---

[7] Velleman (2009, p. 143); Silverstein (2015, p. 1133f).
[8] Enoch (2006, p. 179).

he has such a reason if and only if acting would promote some activity of his. Presumably, this agent is enquiring about something, so he is engaged in *some* activity. If REASON is true, he does have a reason to do something, namely whatever actions will promote it. Conversely, since a shmaction is not an action, it cannot, according to REASON, be the kind of thing that is favoured by a reason. So there are no reasons for him to pursue shmactions. If an *agent* asks if he has a reason to perform actions rather than shmactions, the answer provided by REASON is thus unequivocally 'Yes'. Now suppose instead that a *shmagent* wants to know if he has a reason to perform actions rather than shmactions. If REASON is true, then he has this reason if and only if acting would promote some activity of his. But insofar as the shmagent is not engaged in any activity, he does not have any reasons at all. So REASON is enough to establish that the answer to the shmagent's question is an unequivocal 'No'.

These seem to me the right answers, but there is concern that they leave this defense of REASON vulnerable to the argument I previously pressed against Darwall's account of our reasons to take up the second-person standpoint. I argued then that someone might have the capacity to hold second-personal attitudes, and be blameworthy for failing to exercise it. And this, I claimed, implied that such a person would have reasons to exercise those capacities, reasons that it was difficult to account for on a view like Darwall's. The same problem might be said to arise for a proponent of REASON. Perhaps someone could have the capacity to be engaged in activities, and be blameworthy for failing to exercise it. There would then have to be reasons for this person to exercise this capacity, but those reasons cannot be there according to REASON.

While I agree that the situation here seems to be similar, it also seems to me that we need to buy into a particular assumption to push the analogy this far. Con-

sider an individual, Paul, who has the capacity to perform actions, but only actually performs shmactions.[9] To have the capacity to perform actions, Paul would need to have the capacity for intentional control over his bodily movements and mental states. To have a reason to exercise that capacity, however, it would have to be brought about somehow that Paul engages in some activity, according to REASON.

But this seems to be something *Paul* can bring about by an exercise of his capacity for intentional control – as if it were itself a type of action – only if Paul had some kind of higher-order control over his exercise of intentional control itself. And whether one exercises the intentional control required for an event to be an action cannot be something over which one could have further intentional control. This would make *any* account of action circular or subject to a vicious regress, even one which held intentional control to be a primitive. So to push the Shmagency Argument far enough to ask the question above, it seems one has to assume that no coherent account of action is possible. And that is a lot to assume for an argument that purports to show that normativity could not come from agency, no matter what the correct account of agency is.

For these reasons, it seems to me that we can answer the Shmagency Argument with fairly minimal assumptions. To defend the present account of correctness against it, we do not need to assume that agency is the source of normativity. And all we need to assume about the nature of agency is that some coherent account of it is possible. Accordingly, we can set this objection aside and see whether this account of correctness can do the work that would be required of it to *earn* the claim that activities are source of normativity. That work involves making good on intuitively plausible normative claims, such as the claim that there are reasons for everyone to comply with their obligations. Let us now return to that task.

---

[9] Set aside the terminological question of whether Paul would qualify as an agent or a shmagent.

4.   Locking In

To defend the claim that everyone has most reason to engage in joint activities, we need to show that they have at least *some* reason to do this. How can we do this? According to Reason, a person could have such a reason only in virtue of his engagement in some activity. So one way to show this would be to show that everyone is engaged in some activity that gives them such a reason. But the problem with merely showing that everyone *actually* has some particular reason in virtue of what they are doing is that it does not tell us whether they would have that reason if things were a little bit different. Consider the following example.

Suppose that Stephanie buys herself a lottery ticket each day. Though she does not know it, today is Stephanie's lucky day: she has got the winning ticket. That is a really good reason for Stephanie not to sell her lottery ticket. But had things been only slightly different, Stephanie would not have bought the winning ticket, and she would have had no reason not to sell it. This reason is what we might call *modally unstable*. Just as some reasons have this property, other reasons have the opposite type of property; they are *modally stable*. Suppose, for instance, that John has a reason to believe that the sun has not been extinguished: it is right there on the sky, shining at him. That this is reason for John to believe that the sun has not been extinguished is not down to luck. Even if things had been *very* different from what they are, the sun would still have been there on the sky, shining at John. And this would have been reason for him to believe that the sun has not been extinguished.

Modal stability is a matter of degree. The degree to which Stephanie's reason not to sell her lottery ticket is modally stable will depend on how many winning tickets there are in the draw. If two tickets were winners, her reason not to sell her ticket would be more modally stable than if merely one ticket was a winner, because

at least one more thing would have to be different for her not to have that reason. It would have to be true that Stephanie did not have the ticket she actually has, but it would *also* have to be true that she did not have the *other* winning ticket. We can thus arrange the modal stability of certain conditions along a spectrum. If some condition is impossible, the fact that it does not obtain is maximally modally stable. If some condition is necessary, the fact that it does obtain is maximally modally stable. And then there are all the non-necessary possible conditions in between that may be more or less modally stable.

Now if the account I have been defending is going to show that obligations really *bind* everyone, then it will have to show that reasons for people to engage in joint activities are more like John's reason and less like Stephanie's. These reasons are going to have to be fairly modally stable. Ideally, their status should be *maximally* modally stable, being reasons for everyone *no matter what*. And, of course, it also needs to show that they are reasons of the right kind. It needs to do this in order to explain why it would be *incorrect* for us not to take up the second-person standpoint, and in what sense this would be a mistake. And to do that, on the view I defended in the previous chapter, it needs to show that these reasons are necessarily shared by everyone in virtue of engaging in some activity, no matter what activity they are engaged in.

This could be the case if there were some activity A such that, no matter what activity people are engaged in, they also necessarily engaged in A. Such a view is available if, for instance, one holds a Constitutivist account of agency. But, as we have seen, such accounts are also hard to defend. There is, however, a different way of accounting for the maximal modal stability of reasons, one that relies on weaker commitments. Consider, for instance, how one might explain the modal

stability of an epistemic reason to believe something only if it is true.[10] Suppose you are a first-year mathematics student, having just heard of Fermat's Last Theorem. The theorem seems so simple, it cannot be that hard to show that it is true, so you immerse yourself in the project. Since this involves constructing a proof, knowing how to construct a proof will promote this activity. So not having false beliefs about proof construction will promote that. It will also help not to have false beliefs about where to find books that explain how proofs work. And there are going to be many propositions such that not having false beliefs about them will help finding those books. And for each of those propositions, there are going to be many propositions that are relevant to whether one ends up having false beliefs about those. Given fairly weak assumptions, it will turn out that, for any proposition, you have a reason to believe it only if it is true.

All of these reasons can be explained in virtue of facts about what promotes the activity you are engaged in. But the particular choice of activity we employed in this explanation was completely arbitrary. We could easily have come to the same conclusion by assuming you were engaged in some other activity. In general, it seems that for any activity in virtue of which those reasons might obtain, even if you were not engaged in *that* activity, those reasons would still obtain. So the mere possibility that you might fail to be engaged in either of those activities would fail to show that you might fail to have those reasons. And since this type of argument relies on nothing about the nature of the activities in question, it does not hinge on controversial claims about what anyone is necessarily doing, if they are doing anything at all. This type of argument can provide a strong case for the modal stability of such reasons.

---

[10] The example here is adapted from Schroeder (2007, p. 114).

I hold that a similar type of argument supports the claim that reasons to engage in joint activities are maximally modally stable for everyone. If that is true, then the fact that there are reasons of the right kind to intend to do so will also be maximally modally stable. And to have an intention with that structure, I argued above, is to have the kind of intention which makes it the case that one interacts with people on the basis of a second-personal attitude. This, it seems to me, would be an argument for the existence of non-second-personal reasons to take up the second-person standpoint, and I am going to illustrate shortly how I think it might be made to work. Before doing so, however, it will be helpful to recall Darwall's dictum from chapter 4, according to which all of this cannot be true, because *any* non-second-personal reason to take up the second-person standpoint is of the wrong kind to do so. If that claim is true, this argument falls at the first hurdle. So let me explain why I think we do not need to accept it.

Let us start by granting that, for those within the second-person standpoint, many kinds of considerations are of the wrong kind for various responses. The fact that it would satisfy one's thirst for vengeance is a reason of the wrong kind to punish someone. The fact that it would be expedient is a reason of the wrong kind to blame an innocent person. And the fact that it will promote one's activities is a reason of the wrong kind to hold an attitude of respect towards someone. Does it follow that all such reasons are of the wrong kind to *take up* the second-person standpoint. Well, if it followed from the fact that a reason is of the wrong kind for a certain response within a certain context that it is also a reason of the wrong kind to bring about that context, then this claim would follow.

As I noted in the previous chapter, however, this claim is not generally true. Recall the case in which the evil demon will kill Ronnie unless he does something

he has promised to do. From the fact that this reason is of the wrong kind to *keep* a promise, it does not follow that it is a reason of the wrong kind to bring about a context in which a promise can be kept. It does not follow that it is a reason of the wrong kind to *give* a promise. In other words, a reason may be of the right kind to bring about a context in which it is a reason of the wrong kind. This does not mean that it is no longer a reason. Even once he has given the promise, the fact that the evil demon will kill him is a reason for Ronnie to keep it. Doing so will promote some activity of his. But it is a reason of the wrong kind to keep the promise.

I hold that the same point applies to our reasons to take up the second-person standpoint. But how might we explain this? Well, suppose taking up the second-person standpoint will promote something you have a reason to do, such as engaging in a joint activity with other people. By the argument above, you will have that reason because engaging in this joint activity will promote some other activity you are engaged in. Now suppose you take up the second-person standpoint by forming the relevant kind of shared intention to engage in a joint activity.

Having formed this intention, the fact that another person relies on you to do your bit is *also* a reason for you to do so. It is also a reason because doing this thing will promote the joint activity that you are engaged in. Of course, it is not the *only* reason to do so. There is, for instance, still the reason that it will promote your own individual activity. But it *is* the only kind of fact that is a reason for everyone in virtue of the fact that they have an intention with that form. And that, I have argued, is what it is for it to be a reason of the right kind, relative to the kind of activity that involves that intention. So that is how your reason to take up the second-person standpoint can be a reason to bring about a context in which it is a reason of the wrong kind.

The upshot of this is that there is now a genuinely normative standard relative to which all of your reasons can be assessed. In particular, it is a standard relative to which any reason you might have for revising your second-personal attitudes can be assessed. For instance, take any extant reason for revising your second-personal attitudes that simply concerns the promotion of your own activities regardless of what others rely on you to. This might be the fact that you could make more money if you just did not care about anyone else. That will be a reason of the wrong kind to revise your second-personal attitudes. Any such reason will be. Moreover, it is unlikely that any second-personal reasons are going to favour revising your second-personal attitudes, since, by the argument of chapter 3, anyone who holds them relies on you to hold them too. And add to this the fact that second-personal attitudes generally make your reliance-based reasons weightier. If that is all correct, it seems the second-person standpoint forms a kind of self-supporting normative structure. It is, so to speak, a normative Alcatraz. Once you are in, you never (reasonably) get out.

5.   Weight

The idea that there should be this type of self-supporting normative structure goes a long way, I believe, towards accounting for the normative inescapability of obligations that those of us who reason within the second-person standpoint are faced with. But the argument I have given for this conclusion implies that this structure does not get off the ground all on its own. In one respect, this is a virtue, since it allows it to avoid explanatory circularity. But it also raises a further question about how exactly this structure does get off the ground. On the view I have defended, reasons to take up the second-person standpoint are reasons to intend to do things

with other people; to engage in joint activities with them. These reasons are facts about the kinds of responses that promote activities. So these are the kinds of facts that are supposed to do the heavy lifting here.

So far, however, this account of reasons is somewhat austere. It merely tells us that the fact that ϕ-ing would promote a second-personal activity of Φ is a reason to ϕ, just as the fact that ψ-ing would promote some other type of activity Ψ is a reason to ψ. What Pragmatism needs is some way of weighing the facts about *promotion* of different activities against each other. To be clear, Pragmatism does not need any particular account of how this works. All it needs is *some* way of explaining that people have more reason than not to engage in interpersonal activities, whatever activity they are engaged in. So let me illustrate an explanation along these lines that is both well motivated on the view I have defended so far and independently defensible.

There are many competing views on what it is for a reason be more weighty than another one. Arguably, the simplest and most intuitive is *proportionalism*. Essentially, proportionalism is the view that the weight of a reason is proportional to some non-normative factor. Suppose, for instance, that ϕ-ing will promote Φ-ing while ψ-ing will promote Ψ-ing, and that the degree of promotion in each case is the same. If the reason to ϕ is nevertheless weightier than the reason to ψ this is because, the proportional strategy holds, some feature of Φ-ing is greater than the corresponding feature of Ψ-ing.

Here is a well-known way of employing the proportional strategy. Suppose Bobby desires two things: to eat bacon and to observe a vegetarian diet. According to a simply desire-based theory, the fact that eating bacon would promote the satisfaction of desire of Bobby's is a reason to do that. And the fact that observing

a vegetarian diet would promote the satisfaction of another desire of Bobby's is a reason to do that. But, since Bobby cannot both eat bacon and observe a vegetarian diet, these reasons conflict. And since eating bacon and observing a vegetarian diet are identical in the degree to which they promote the satisfaction of some desire (both of them satisfying it completely), it is not a difference in the degree to which they *promote* something that is going to explain why one reason is weightier than the other.

Naturally, this does not force a proponent of the desire-based view to conclude that Bobby could not possibly have a weightier reason to do one thing rather than the other. As some desire theorists hold, Bobby may have a weightier reason to observe a vegetarian diet if his desire to do this is *stronger* than his desire to eat bacon. As the desire theorist might also say, the weight of a reason to do what promotes the satisfaction some desire is *proportional* to the strength of the desire: more weighty if the desire is stronger, less weighty if the desire is weaker. Similarly, hedonists may argue that even if doing A will provide you with enjoyment for precisely the duration of time that doing B will, you may have a weightier reason to do A because the enjoyment it provides is *more intense*. On this type of view, the weight of one's reason to do something is proportional to the degree of enjoyment it would promote.

On both these types of views, the weight of one's reasons is proportional to a gradable property of a given psychological state, like strength or intensity. Activities are, of course, not psychological states. But a view that grounds normative properties in activities might nevertheless employ a version of the proportional strategy by appealing to the fact that activities have other kinds of gradable properties. It may not be obvious, however, why this type of account of the weight of reasons

should be appealing. This is particularly so given that proportional strategies are often considered problematic. Indeed, the best objection to the idea that the weight of a reason should be proportional to the promotion of some non-normative factor is that it tends to lead to counterintuitive implications. On the hedonist view, for instance, proportionalism seems to imply that a reason to choose the life of an oyster is weightier than the reason to choose the life of Haydn, as long as the amount of enjoyment in the oyster life is of greater proportion than the amount of enjoyment in Haydn's life.[11]

However, we should recall that the normative fundamentality of reasons restricts the available options here. We cannot, for instance, give an account of weight in terms of promotion of value.[12] Of those left, the proportional strategy is arguably the simplest available account.[13] And if the best objection to such an account is its counterintuitive implications, the best way to defend it is to show that it does not have counterintuitive normative implications. But if such a strategy can support the claim that we have most reason to take up the second-person standpoint, the arguments above suggest that it will at least be well on its way to avoiding *morally* counterintuitive implications. So I think the proportional strategy is certainly worth exploring here.

Thomas Hurka defends a view that is broadly of this kind. Hurka argues that for something to be good for a human being is for it to involve the exercise of their essentially human capacities. As Hurka observes, it is a constraint on such a theory that it provides acceptable predictions about which kinds of things are good

---

[11] Crisp (1997, p. 24).

[12] This rules what is otherwise an admirably lucid account defended in Maguire (2016).

[13] The reader can compare competing, and extraordinarily complex, accounts of the weight of reasons that avoid appeal to other normative properties in Schroeder (2007, Chap. 7); Scanlon (2014, Lect. 5).

for people. And this may appear to be a problem for such a theory, since nothing excludes the possibility that a human being could exercise his essentially human capacities in ways that are, intuitively, bad for him. Yet what sets human beings apart from other animals, Hurka also observes, is their capacity for more complex forms of thoughts and activities. And this intuition goes together with the fact that the kinds of activities that we, intuitively, consider it good for people to engage in are comparatively complex.[14]

For instance, most of us believe that it is, other things being equal, better for people to acquire abstract, general knowledge from which they can acquire more concrete, particular knowledge by reasoning than it is simply to acquire the concrete, particular knowledge on its own: the former is a greater intellectual achievement than the latter. As Hurka also observes, a plausible explanation for this is that the former requires a greater exercise of an essentially human capacity like complex theoretical reasoning than the latter. What the theory should hold, then, is that engaging in activity A is *better* for one than engaging in activity B just in case activity A involves the exercise of one's essentially human capacities *more*, in particular by being more complex.

This idea raises some intricate questions. One is how precisely we are supposed to measure complexity. Hurka's suggestion is that two things factor into this: dominance and extent. An end is more extended when it encompasses a greater number of subordinate ends within it. It is more dominant when it encompasses ends of a greater variety. For instance, politics is more complex than repetitive factory work, because it requires the pursuit and integration of ends in a great variety of fields, whereas factory work involves doing things of the same kind over and over again. In general, the best ends are the most extended, dominant ones.

---

[14] Hurka (1993, p. 115f.).

Another question is why more complexity should always be better. Doing something (e.g. factory work) in a roundabout way can make it more complex, but intuition militates against the idea that inefficiency is better. To this we may reply that intuition supports the claim that inefficiency is unvaluable just when and because it prevents the pursuit of other ends. And although inefficiency may add complexity, there is intuitively greater value in achieving a complex goal by more elegant means, such as when a proof establishes the truth of additional lemmas along the way. Thus, taking the value of complexity to be basic does not require a commitment to the idea that *needless* complexity should be counted as valuable, as long as any apparent disvalue of complexity can be explained by reference to *other* complex ends.[15] And which types of ends should serve as the baseline relative to which this is all ultimately assessed on this view? Complex ends, of course.

Arguably, there are further questions one would have to answer to deliver a full-fledged theory of weight. I cannot answer all of these questions here. My aim is merely to show that, if something like this view is acceptable, the kind of argument that Pragmatism needs can be made to work. And the reason this idea is worth exploring in our context is that one has to make a particular assumption for it to work, one that I have already defended. This is the assumption that activities are not identical to the sum of their parts. For if they were, it would be mysterious why the integration of these different parts into a coherent whole should add any kind of valuable complexity to the pursuit of them. Thus, it seems to me that an appeal to something like primeness is precisely what one needs for this type of approach to work.

So while Pragmatism is clearly not committed to this particular account of the weight of reasons, it does provide an independent rationale for developing it. I sug-

---

[15] Hurka (1993, p. 138).

gest the following basic idea: The fact that [ϕ-ing will promote Φ-ing] is a *weightier reason* to ϕ than the fact that [ψ-ing will promote Ψ-ing] is a reason to ψ just in case activity Φ is *more complex* activity Ψ. So, on this view, the fact that something would promote a joint activity is weightier reason than the fact that something else would promote an individual activity *just in case* a joint activity is more complex than an individual activity. Is this so?

Well, I have argued that joint activities are joint precisely in virtue of being prime in the specific sense that they necessarily involve the interdependence of two people's actions. This does not imply that any joint activity is necessarily more complex than any individual activity. Building a space craft by oneself is arguably more complex than playing checkers with someone else. But it does suggest that for any activity that *is* part of such a joint activity, one will have a weightier reason to engage in it than if it is *not* part of such a joint activity. This will be so because that individual activity will come to be integrated with another activity that is also qualitatively different in being the activity of another person; that is, it will involve engagement in a unified agential whole that is both more extended and more dominant.

Of course, sometimes one may have most reason to seek solitude rather than company. But this is no more problematic than the claim that one may have most reason to avoid roundabout complexity. All this requires is that we can explain in terms of *other* joint activities why one would have reasons to seek solitude. So when Albert has most reason to sit by himself in the library to read, for instance, the present view is merely committed to the claim that this is true in virtue of some joint activity of which Albert's sitting by himself in the library is a part. In general, it holds that whatever activity one is engaged in, there will be a weightier reason

to integrate it within another person's activities, because this will promote a more complex activity.

In the abstract, this line of thought may be hard to follow, but its implications are plausible enough. Many of us find the idea compelling that even those activities which do not strictly *require* interaction with other people nevertheless become more worthwhile when they are. Even such things as taking a walk fall in this category. The suggestion above merely extends this idea. By taking up an intention to integrate other people's agency with one's own, by bringing their agency into one's reasoning as a premise, one adds complexity to one's own agency. So the weight of one's reasons to engage in activities on that basis is comparatively greater.

If a conception of the weight of reasons along these lines can be made out, then the weight of our reasons to hold second-personal attitudes will turn out to be *maximally* modally stable. Because for any activity one might be engaged in, there is a weightier reason to engage in it on the basis of second-personal attitudes than not. So no matter what activities one happens to be engaged in, one will have most reason to hold such attitudes. In fact, for any activities one might be engaged in, even if one were not engaged in those particular activities, one would have most reason to hold such attitudes. The correctness standards of second-personal activities and the relationships they give rise to apply inescapably because one always has most reason to impose them on oneself; by intending to engage in joint activities, that is.

Indeed, I believe this type of argument adds further credence to the idea that the type of intention we have reason to form is one that involves the integration of everyone's agency. If Albert and Annie are around, for instance, they will both have more reason than not to intend to engage in an activity that constitutively involves both of them, by the above argument. And if Arnie is also around, Albert,

Annie, and Arnie will all have more reason than not to engage in an activity that constitutively involves the three of them, and so on. By iteration, that argument implies that everyone who is around will have more reason than not to engage in some type of activity which integrates the agency of everyone. Everyone will share those reasons in virtue of what they are doing, regardless of what they are doing. So, for everyone, it is correct to have that aim. And for everyone to engage with other people on the basis of that aim, I have argued, would be for everyone to share the moral relationship.

6.   Closing Remarks

How much closer to does this view take us towards answering the questions we began with? Well, Deontic Pragmatism is a theory about what obligations are. It holds that they are correctness standards that obtain in virtue of reasons of a particular kind, facts about how people rely on each other. On this view, it is the fact that reliance is itself a bipolar relation that explains why obligations involve the bipolar normative features that they do. The bipolar normative significance of that relation is further reflected in the attitudes of trust and blame. Indeed, Deontic Pragmatism holds that this bipolar normativity permeates all obligations at some level or other, and that our disposition to trust other people by default and hold attitudes like indignation at their behaviour simply reflect the more abstract forms that such reliance relations can take.

But Deontic Pragmatism also offers an explanation for why it is correct to take facts about other people's reliance to be reason-giving in the first place. Its explanation is that everyone has most reason to intend to engage with other people generally in joint activities, and that having an intention with that form simply is to take

such facts to be reason-giving. Because everyone shares those reasons simply in virtue of being engaged in any activity at all, those reasons also make it correct for everyone to hold such attitudes, and incorrect not to. And the reasons that make this the case, I have just argued, are facts about how engaging in joint agency with other people will promote one's own activities. These are facts, not just about one's own agency, but about its dependence on the agency of others. Because the interdependence of agency is the relation presupposed by the attitude of reliance, because reliance is constitutive of trust, and because trust is constitutive of moral agency, Deontic Pragmatism thereby locates the normative source of our obligations by drawing out the normative relations presupposed in our moral psychology.

My defence of Deontic Pragmatism is thus indirectly a defence of the idea that moral psychology itself is a guide to the normative source of obligations. We can defend that idea, I have claimed, by analysing obligations as correctness standards that obtain in virtue of a specific kind of reason. And, I have argued, for any genuine normative correctness standard, we can distinguish the reasons that ground it from those that do not in wholly non-normative terms. This defence of Deontic Pragmatism is thereby also a defence of the reasons-first programme in philosophy more generally.

Together, I believe these conclusions allow for a robust response to the dilemma we discussed at the outset. As you will recall, it is no good for a moral theory to tell us that we should do what is right because it is right. Deontic Pragmatism avoids this by saying, in effect, that we should do what is right because it promotes our activities. Because that reply does not depend on the appeal to any other normative notions, and because the relevant notion of activity is thoroughly non-

normative, Deontic Pragmatism avoids both direct circularity and indirect triviality. So it seems to me to avoid this horn of the dilemma.

The other horn of the dilemma is that it is no good for a moral theory to tell us that we should do what is right for reasons that seem to have nothing with the reasons that would motivate a morally decent person. Deontic Pragmatism avoids this by saying, in effect, that the kinds of facts that morally decent people take to be reason-giving are virtually identical to the facts that ultimately explain why they should be moral in the first place. To take up the second-person standpoint is to take facts about reliance to be reason-giving. And the kinds of facts that ultimately explain why that is the correct type of attitude for everyone to hold is that the relation one presupposes in relying on another person *is* fundamentally reason-giving; the dependence of one person's agency on another. This narrows the gap between these two kinds of reasons considerably.

This view may yet seem unsatisfactory for two different reasons. The first is that it seems to force an equivocation into the statement of what obligations really are. When I originally presented Deontic Pragmatism's account of obligations, I said that obligations obtained in virtue of how reliance-based reasons weigh against each other. In giving the account of correctness standards, I later said that obligations were correctness standards of second-personal activities. These two formulations are not in tension with each other, I believe, once we take on board the analysis of second-personal interaction as responsiveness to reliance. But there is a different point at which a problematic equivocation may seem to come in.

I criticised both Darwall's and Scanlon's positions for making it, in different ways, unclear how it might be a mistake for people to fail to take up the second-person standpoint. Such a mistake, I argued, could be blameworthy, and this is

something a moral theory should be able to explain. But I also argued in chapter 1 that blame presupposed the violation of an obligation, and this seems to suggest that we ought to be able to account for an obligation to take up the second-person standpoint. However, the reasons Deontic Pragmatism ultimately posits as reasons to do so are not facts about reliance. So if there is such an obligation, it cannot be true that obligations obtain in virtue of facts about reliance. On the other hand, if we instead characterise these reasons directly as facts about the dependence of agency, rather than the attitude of reliance, then the Reliance View defended in chapter 2 and chapter 3 seems to be false. So the view here may seem to rely on an equivocation about the relation between reasons and obligations at the most fundamental level.

One response to this concern is to accept that there is no final statement of what obligations ultimately are. This is on some level unsatisfactory, but the normative facts would in any case be the same. Facts about reliance really are reasons to respond in various ways for those who have the attitudes they have reasons to have, and facts about dependence and promotion of agency really are reasons for everyone to hold those kinds of attitudes. Since correctness standards derive directly from these reasons, nothing of normative consequence depends on what we say obligations are.

Another response is to revise the view that blame presupposes the violation of an obligation. In principle, the account of blameworthiness defended in chapter 3 is compatible with this response, since it requires merely that blameworthy behaviour impairs relationships that there are reasons to have. It does not require that those reasons amount to obligations. This reply seems to me more satisfactory. But there is another reason to think that this sort of reply is unsatisfactory. Even if we restrict

ourselves to reliance-based reasons for obligations, the account still implies that whether or not those facts actually make it the case that one is obligated depends on whether or not one actually holds the attitudes there are reasons for one to hold. Thus, even if the account can explain why it would be blameworthy not to take up those attitudes, it still seems to rely on the idea that such attitudes can cause a shift in one's normative situation, and this may seem implausible.

This concern can be mitigated, I believe, by pointing out that the notion of such a shift in one's normative situation is built into the phenomena that an account of obligations should explain. In the case of promising, whether the reliance of another person makes it correct to do what he relies on one to do depends on the attitudes of both the promisor and the promisee. So while the situation at the level of taking up the second-person standpoint is not identical, I do not see why something similar could not be the case in this context.[16] Indeed, the idea that there should be such continuity in the way that different kinds of obligations get off the ground seems to me independently appealing.

While these responses seem to me to go some way towards relieving the concerns raised above, I believe there is yet another position available that provides a more unified answer to both of these concerns. The first step of this reply is to grant that there simply is an an ineliminable difference between these reliance- and dependence-based reasons. So while these are both categories of genuine reasons, the claims defended here will never allow us to close the gap between the foundational (dependence-based) reasons to have moral attitudes and the higher level

---

[16] John McDowell argues that something like this *is* the case in the context of reasons to be virtuous, on Aristotle's view: 'To embrace a specific conception of eudaimonia is to see the relevant reasons for acting, on occasions when they coexist with considerations that on their own would be reasons for acting otherwise, as, not overriding, but silencing those other considerations – as bringing it about that, in the circumstances, they are not reasons at all' McDowell (1980, p. 370).

(reliance-based) reasons that those who hold moral attitudes take to obligate them. Pragmatism, as I have stated it here, leaves explanatory residue.

The second step is to point out that, if Pragmatism is correct, there will nevertheless be conditions under which these different kinds of reasons overlap. These conditions, I conjecture, will be conditions of moral knowledge. For when people know what they are doing, their reliance on other people will involve knowledge of their actual dependence on them. For someone who knows that other people knowingly rely on them, reasons grounded in reliance and dependence relations, respectively, will overlap. That would explain why Pragmatism, on its own, could never completely close the gap between these different levels of moral explanation; to do that, we need to do some moral epistemology.

The account defended in chapter 5 suggests that the appearance of a gap between the facts about reliance and dependence may thus be closed, insofar as the conditions of second-personal interaction are satisfied, because those conditions involve common knowledge. It is not that this claim cannot be contested. The point is rather that it illustrates where Pragmatism can acquire the resources to address outstanding concerns; namely, moral epistemology. And this, I believe, is both an interesting and encouraging result for a proponent of Pragmatism. Because it is not hard to see how Pragmatism can supply an epistemology that allows people to get that kind of knowledge without looking very far. And, incidentally, Pragmatism might thereby make the hard problem of the normativity of obligations easier if Anscombe was also right about another thing: knowledge of what one is doing is knowledge without observation.[17] For if Pragmatism is right, moral knowledge just *is* a kind of knowledge of what one is doing. That, I believe, is a possibility we would do well to explore.

---

[17] Anscombe (1963, pp. 13–5).

Conclusion

I have now spent the better part of six chapters discussing a view you have probably never heard of before. I have argued that it is defensible, but provided little by way of explicit positive motivation. So why should you believe Deontic Pragmatism rather than so many other available moral theories? As I have stated it, Deontic Pragmatism is an *explanatory* view, one that aims to explain what obligations are and why they are normative. And if anything characterises a good explanation, it is generality, simplicity, and economy. They should explain as much as possible, as simply as possible, with as few means as possible. Deontic Pragmatism is an attractive view, I believe, because it offers a general explanatory framework that exhibits these characteristics to a high degree.

I started out in chapter 1 by outlining the various features of obligations that a plausible account of obligations should explain. These included the way obligations feature in deliberation, blame, and trust. With respect to each of these phenomena, I argued, obligations have some salient 'bipolar' features that a plausible account should also explain. Deontic Pragmatism does that in a simple way by saying that obligations are determined by what sets of reasons favour, and then by saying that those reasons are facts about the bipolar relation of reliance. This is what I referred to as the more specific Reliance View of obligations.

In chapters 2 and 3 I argued that this type of explanation can also generalise to a wide variety of obligations. It explains the normative features of promissory obligations, obligations of special relationships, imperfect obligations, and the kinds of

obligations that each of us has to each other person, simply because they are a person. And it does this completely generally by pointing to the significance of people's attitudes for the kinds of relationships that they have reasons to have with us, in a way that also explains how it might be reasonable for us to trust and blame them, according to how they respond to their own obligations.

So what I have done in the first three chapters is to show how Deontic Pragmatism provides a very simple and very general explanation of why obligations are normative, in a way that makes sense of the psychology of moral agents and explains why its various components fit together in the way that they seem to do. As I have also argued, the only *normative* property this explanation needs to appeal to is the property of being a reason, and that also makes it very economical. Together, these features make it an attractive explanation of the nature and normativity of obligations.

My case for this explanation was *conditional* on the plausibility of the claim that everyone has reasons to hold attitudes of recognition towards each other. That is a good thing, I believe, because it means that it is acceptable on a variety of views about why people would have such reasons. But, as you will recall from chapter 4, it also raised a general problem about distinguishing the reasons that factor into the kind of correctness standards that obligations are from other kinds of reasons. Indeed, I argued that this was a completely general problem for *any* reasons based approach to obligations.

We discussed different ways of responding to that problem, and I believe we can now appreciate that each of them is less attractive than the response Deontic Pragmatism offers. Deontic Pragmatism is more economical than views that require appeal to other normative properties. It is simpler than views that resort to

splitting explanations of the normativity of obligations into two levels. And it is more general, because it explains more. It explains what the relevant reasons are, how people could act for them, and how they could be blameworthy for failing to act for them. None of the other views explained *all* of this.

As I argued in chapter 5, the way Deontic Pragmatism does this is by saying that reasons are facts about how individual actions promote individual activities. The reasons that ground correctness standards are those that people necessarily share in virtue of being engaged in that activity, and so that is how Deontic Pragmatism can explain the distinction between the reasons that factor into some particular correctness standard, such as an obligation, from those that do not, without assuming other normative properties, and without assuming that there are fundamentally different kinds of reasons.

However, I also argued that activities are really nothing over and above the actions that compose them. And this is important because it offers us a very simple way of connecting different levels of normative explanation: Higher-order normative properties (correctness) obtain in virtue of lower-order normative properties (reasons) *because* higher-order properties of agency (activities) obtain in virtue of lower-order properties of agency (actions). And if we can analyse axiological properties (value) in terms of correctness standards for attitudes, what we get is a unified explanation for how the normative supervenes on the non-normative. And we can do all of this with just two basic explanatory components: reasons and actions. That gives Deontic Pragmatism a strategy for normative explanation that is both simple, general, and *very* economical.

I am careful to say *strategy* here, because I have not actually *claimed* that all of normativity is explained in this way. As I stated the fundamental relation between

reasons, actions, and activities in chapter 5, it is compatible with the claim that there are further background conditions on something's being a reason, other than engagement in an activity. Nevertheless, I take it that an explanation which avoids appeal to further background conditions is more economical than one which does not, and is thus more attractive on that score.

My argument in chapter 6 was that we do *not* need to appeal to any other background conditions on reasons (normative or non-normative) to explain what we want to. Given the argument of the first three chapters, what we want to explain here is how everyone has reasons to engage with each other on the basis of certain attitudes, and I argued that we could explain that by developing an account of the weight of reasons *in terms* of properties of activities, namely their complexity. I am the first to recognise that this view needs to be developed in further detail, but it suffices to show that it is *possible* to develop this type of explanation without further commitments.

Deontic Pragmatism thereby offers a simple, general, and economical explanation of our reasons in general and, in particular, the reasons that make it the case that we are all obligated to other people. And it does this by drawing on an independently plausible and previously underexplored account of the normative content of our moral psychology. These seem to me excellent reasons to consider it an attractive view.

References

Ackrill, J. L. (2001). 'Aristotle's Distinction Between Energeia and Kinesis.' In *Essays on Plato and Aristotle*. Oxford: Oxford University Press, 142–62.

Alonso, Facundo M. (2009). 'Shared Intention, Reliance, and Interpersonal Obligations.' *Ethics* 119 (3): 444–75.

_____ (2014). 'What is Reliance?.' *Canadian Journal of Philosophy* 44 (2): 163–83.

_____ (2016). 'Reasons for Reliance.' *Ethics* 126 (2): 311–38.

Andersen, Svend and Kees van Kooten Niekerk, eds. (2007). *Concern For the Other: Perspectives on the Philosophy of K.E. Løgstrup*. Notre Dame, Ind.: University of Notre Dame Press.

Anscombe, G. E. M. (1958). 'Modern Moral Philosophy.' *Philosophy* 33 (124): 1–19.

_____ (1963). *Intention*. Cambridge, Mass.: Harvard University Press.

Aristotle (2005). *Nicomachean Ethics*. Roger Crisp (ed.). Cambridge: Cambridge University Press.

Arpaly, Nomy (2002). *Unprincipled Virtue: An Inquiry Into Moral Agency*. Oxford University Press.

Bader, Ralf (forthcoming). 'Multiple-Domain Supervenience for Non-Classical Mereologies.' In Miguel Hoeltje, Benjamin Schnieder, and Alex Steinberg (eds.). *Ontological Dependence and Supervenience*. Munich: Philosophia, 347–68.

_____ (2016). 'Conditions, Modifiers, and Holism.' In Lord and Maguire 2016, 27–55.

Baier, Annette (1986). 'Trust and Antitrust.' *Ethics* 96 (2): 231–60.

Baker, Judith (1987). 'Trust and Rationality.' *Pacific Philosophical Quarterly* 68: 1–13.

Becker, Lawrence C. (1996). 'Trust as Noncognitive Security About Motives.' *Ethics* 107 (1): 43–61.

Bratman, Michael (2009). 'Shared Agency.' In Chrysostomos Mantzavinos (ed.). *Philosophy of the Social Sciences: Philosophical Theory and Scientific Practice*. Cambridge University Press, 41–59.

Brunero, John (2010). 'Consequentialism and the Wrong Kind of Reasons: A Reply to Lang.' *Utilitas* 22 (3): 351–59.

Chant, Sara Rachel (2010). 'Two Composition Questions in Action.' In Allan Hazlett (ed.). *New Waves in Metaphysics*. Basingstoke: Palgrave Macmillan, 27–53.

Chappell, Richard Yetter (2012). 'Fittingness: The Sole Normative Primitive.' *Philosophical Quarterly* 62 (249): 684–704.

Ciocchetti, Christopher (2003). 'Wrongdoing and Relationships.' *Social Theory and Practice* 29 (1): 65–86.

Coates, D. Justin and Neal A. Tognazzini, eds. (2013). *Blame: Its Nature and Norms.* Oxford: Oxford University Press.

Cogley, Zac (2012). 'Trust and the Trickster Problem.' *Analytic Philosophy* 53 (1): 30–47.

Crisp, Roger (1997). *Routledge Philosophy Guidebook to Mill on Utilitarianism.* London: Routledge.

_____ (2006). *Reasons and the Good.* Oxford: Clarendon Press.

D'Arms, Justin and Daniel Jacobson (2000a). 'Sentiment and Value.' *Ethics* 110 (4): 722–48.

_____ (2000b). 'The Moralistic Fallacy: On the Appropriateness of Emotions.' *Philosophy and Phenomenological Research* 61 (1): 65–90.

Dancy, Jonathan (2004). 'Enticing Reasons.' In Pettit, Scheffler, M. Smith, and Wallace 2004, 91–118.

Danielsson, Sven and Jonas Olson (2007). 'Brentano and the Buck-Passers.' *Mind* 116 (463): 511–22.

Darwall, Stephen (2002). *Welfare and Rational Care.* Princeton, N.J.: Princeton University Press.

_____ (2006). *The Second-Person Standpoint: Morality, Respect, and Accountability.* Cambridge, Mass.: Harvard University Press.

_____ (2007). 'Reply to Korsgaard, Wallace, and Watson.' *Ethics* 118 (1): 52–69.

_____ (2010). '"But It Would Be Wrong".' *Social Philosophy and Policy* 27 (2): 135–57.

_____ (2011). 'Demystifying Promises.' In Sheinman 2011, 255–76.

_____ (2012). 'Bipolar Obligation.' *Oxford Studies in Metaethics* 7: 333–58.

_____ (2016). 'Taking Account of Character and Being an Accountable Person.' *Oxford Studies in Normative Ethics* 6: 12–36.

_____ (2017a). 'Morality, Blame, and Internal Reasons.' In Peter Singer (ed.). *Does Anything Really Matter?: Essays on Parfit on Objectivity.* Oxford: Oxford University Press, 259–79.

_____ (2017b). 'Trust as a Second-Personal Attitude (of the Heart).' In Faulkner and Simpson 2017, 35–50.

Dietz, Alexander (2016). 'What We Together Ought to Do.' *Ethics* 126 (4): 955–82.

Domenicucci, Jacopo and Richard Holton (2017). 'Trust as Two-Place.' In Faulkner and Simpson 2017, 149–60.

Enoch, David (2006). 'Agency, Shmagency: Why Normativity Won't Come From What is Constitutive of Action.' *Philosophical Review* 115 (2): 169–98.

Evans, Gareth (1982). *The Varieties of Reference.* Oxford: Oxford University Press.

Faulkner, Paul (2011). *Knowledge on Trust.* Oxford: Oxford University Press.

_____ (2015). 'The Attitude of Trust is Basic.' *Analysis* 75 (3): 424–29.

Faulkner, Paul and Thomas W. Simpson, eds. (2017). *The Philosophy of Trust.* Oxford: Oxford University Press.

Feltham, Brian and John Cottingham, eds. (2010). *Partiality and Impartiality: Morality, Special Relationships, and the Wider World.* Oxford: Oxford University Press.

Fichte, Johann Gottlieb (2000). *Foundations of Natural Right: According to the Principles of the Wissenschaftslehre.* Cambridge: Cambridge University Press.

Fink, Hans and Robert Stern, eds. (2017). *What Is Ethically Demanded?* Notre Dame, Ind.: University of Notre Dame Press.

Foot, Philippa (1972). 'Morality as a System of Hypothetical Imperatives.' *Philosophical Review* 81 (3): 305–16.

Friedrich, Daniel and Nicholas Southwood (2011). 'Promises and Trust.' In Sheinman 2011, 277–94.

Frost-Arnold, Karen (2014). 'The Cognitive Attitude of Rational Trust.' *Synthese* 191 (9): 1957–74.

Gibbard, Allan (1990). *Wise Choices, Apt Feelings: A Theory of Normative Judgment.* Cambridge, Mass.: Harvard University Press.

Greenspan, Patricia (2010). 'Making Room for Options: Moral Reasons, Imperfect Duties, and Choice.' *Social Philosophy and Policy* 27 (2): 181–205.

Haase, Matthias (2014). 'Am I You?.' *Philosophical Explorations* 17 (3): 358–71.

Hardin, Russell (2002). *Trust and trustworthiness.* Russell Sage Foundation.

Hare, R. M. (1963). *Freedom and Reason.* Oxford: Clarendon Press.

Hawley, Katherine (2013). 'Trust, Distrust and Commitment.' *Noûs* 48 (1): 1–20.

_____ (2014). 'Partiality and Prejudice in Trusting.' *Synthese* 191 (9): 2029–45.

Hawthorne, John and David Manley (2012). *The Reference Book.* Oxford: Oxford University Press.

Hawthorne, John and Jason Stanley (2008). 'Knowledge and Action.' *Journal of Philosophy* 105 (10): 571–90.

Heal, Jane (2014). 'Second Person Thought.' *Philosophical Explorations* 17 (3): 317–31.

Helm, Bennett W. (2014). 'Trust as a Reactive Attitude.' *Oxford Studies in Agency and Responsibility* 2: 187–215.

Heuer, Ulrike (2010). 'Wrongness and Reasons.' *Ethical Theory and Moral Practice* 13 (2): 137–52.

Hieronymi, Pamela (2005). 'The Wrong Kind of Reason.' *Journal of Philosophy* 102 (9): 437–57.

_____ (2008). 'The Reasons of Trust.' *Australasian Journal of Philosophy* 86 (2): 213–36.

_____ (2009). 'Two Kinds of Agency.' In Lucy O'Brien and Matthew Soteriou (eds.). *Mental Action.* Oxford: Oxford University Press, 138–62.

Hills, Alison (2009). 'Moral Testimony and Moral Epistemology.' *Ethics* 120 (1): 94–127.

Hinchman, Edward S. (2017). 'On the Risks of Resting Assured: An Assurance Theory of Trust.' In Faulkner and Simpson 2017, 51–69.

Holton, Richard (1994). 'Deciding to Trust, Coming to Believe.' *Australasian Journal of Philosophy* 72 (1): 63–76.

_____ (2014). 'Intention as a Model for Belief.' In Manuel Vargas and Gideon Yaffe (eds.). *Rational and Social Agency: Essays on the Philosophy of Michael Bratman.* Oxford: Oxford University Press.

Horton, Joe (2017). 'Aggregation, Complaints, and Risk.' *Philosophy and Public Affairs* 45 (1): 54–81.

Horty, John F. (2012). *Reasons as Defaults.* Oxford: Oxford University Press.

Howard, Christopher (forthcoming). 'The Fundamentality of Fit.' *Oxford Studies in Normative Ethics.*

Hurka, Thomas (1993). *Perfectionism.* Oxford: Oxford University Press.

Jones, Karen (1996). 'Trust as an Affective Attitude.' *Ethics* 107 (1): 4–25.

_____ (2012). 'Trustworthiness.' *Ethics* 123 (1): 61–85.

_____ (2017). '"But I Was Counting on You ".' In Faulkner and Simpson 2017, 90–108.

Kant, Immanuel (2015). *Critique of Practical Reason.* Mary Gregor and Andrews Reath (eds.). Cambridge: Cambridge University Press.

Kavka, Gregory S. (1983). 'The Toxin Puzzle.' *Analysis* 43 (1): 33–36.

Keller, Simon (2013). *Partiality.* Princeton, N.J.: Princeton University Press.

Kelly, Thomas (2002). 'The Rationality of Belief and Other Propositional Attitudes.' *Philosophical Studies* 110 (2): 163–96.

Keren, Arnon (2014). 'Trust and Belief: A Preemptive Reasons Account.' *Synthese* 191 (12): 2593–615.

Kolodny, Niko (2003). 'Love as Valuing a Relationship.' *Philosophical Review* 112 (2): 135–189.

_____ (2010a). 'Which Relationships Justify Partiality? General Considerations and Problem Cases.' In Feltham and Cottingham 2010, 169–93.

_____ (2010b). 'Which Relationships Justify Partiality? The Case of Parents and Children.' *Philosophy and Public Affairs* 38 (1): 37–75.

_____ (2011). 'Aims as Reasons.' In Wallace, Kumar, and Freeman 2011, 43–78.

Korsgaard, Christine M. (1993). 'The Reasons We Can Share: An Attack on the Distinction Between Agent-Relative and Agent-Neutral Values.' *Social Philosophy and Policy* 10 (1): 24–51.

_____ (2007). 'Autonomy and the Second Person Within: A Commentary on Stephen Darwall's the Second-Person Standpoint.' *Ethics* 118 (1): 8–23.

_____ (2009). *Self-Constitution: Agency, Identity, and Integrity.* New York: Oxford University Press.

Kripke, Saul A. (1980). *Naming and Necessity.* Cambridge, Mass.: Harvard University Press.

Kutz, Christopher (2000). 'Acting Together.' *Philosophy and Phenomenological Research* 61 (1): 1–31.

Lahno, Bernd (2001). 'On the Emotional Character of Trust.' *Ethical Theory and Moral Practice* 4 (2): 171–89.

Lang, Gerald (2008). 'The Right Kind of Solution to the Wrong Kind of Reason Problem.' *Utilitas* 20 (4): 472–89.

Levy, Yair (2013). 'Intentional Action First.' *Australasian Journal of Philosophy* 91 (4): 705–18.

_____ (2016). 'Action Unified.' *Philosophical Quarterly* 66 (262): 65–83.

Liberto, Hallie (2016). 'Promissory Obligation: Against a Unified Account.' *Oxford Studies in Normative Ethics* 6: 102–22.

Locke, John (1975). *An Essay Concerning Human Understanding.* Peter H. Nidditch (ed.). Oxford: Oxford University Press.

Løgstrup, Knud Ejler (1997). *The Ethical Demand.* Hans Fink and Alisdair MacIntyre (eds.). Notre Dame, Ind.: University of Notre Dame Press.

Longworth, Guy (2014). 'You and Me.' *Philosophical Explorations* 17 (3): 289–303.

_____ (2017). 'Faith in Kant.' In Faulkner and Simpson 2017, 251–71.

Lord, Errol and Barry Maguire, eds. (2016). *Weighing Reasons.* New York: Oxford University Press.

Lott, Micah (2016). 'Morality, Accountability and the Wrong Kind of Reasons.' *Utilitas* 28 (1): 28–40.

MacCormick, Neil (1972). 'Voluntary Obligations and Normative Powers.' *Aristotelian Society Supplementary Volume* 46 (1): 59–102.

Macnamara, Coleen (2013). 'Taking Demands Out of Blame.' In Coates and Tognazzini 2013, 141–61.

Maguire, Barry (forthcoming). 'There Are No Reasons for Affective Attitudes.' *Mind.*

_____ (2016). 'The Value-Based Theory of Reasons.' *Ergo: An Open Access Journal of Philosophy* 3 (9): 233–62.

_____ (2017). 'Love in the Time of Consequentialism.' *Noûs* 51 (4): 686–712.

Markovits, Julia (2010). 'Acting for the Right Reasons.' *Philosophical Review* 119 (2): 201–42.

Mason, Elinor (2005). 'We Make No Promises.' *Philosophical Studies* 123 (1-2): 33–46.

May, Simon Căbulea (2015). 'Directed Duties.' *Philosophy Compass* 10 (8): 523–532.

McDowell, John (1980). 'The Role of Eudaemonia in Aristotle's Ethics.' In Amélie Oksenberg Rorty (ed.). *Essays on Aristotle's Ethics.* London: University of California Press, 359–76.

_____ (1998). *Mind, Value, and Reality.* Cambridge, Mass.: Harvard University Press.

McGeer, Victoria (2008). 'Trust, Hope and Empowerment.' *Australasian Journal of Philosophy* 86 (2): 237–54.

McHugh, Conor and Jonathan Way (2016). 'Fittingness First.' *Ethics* 126 (3): 575–606.

McKenna, Michael S. (1998). 'The Limits of Evil and the Role of Moral Address: A Defense of Strawsonian Compatibilism.' *The Journal of Ethics* 2 (2): 123–142.

McNamara, Paul (1996). 'Must I Do What I Ought (or Will the Least I Can Do Do)?.' In Mark Brown and Jose Carmo (eds.). *Deontic Logic, Agency and Normative Systems*. London: Springer-Verlag, 154–73.

Mill, John Stuart (1998). *Utilitarianism*. Roger Crisp (ed.). Oxford: Oxford University Press.

Moore, G. E. (1993). *Principia Ethica*. Revised Edition. Cambridge: Cambridge University Press.

Nagel, Thomas (1969). 'Sexual Perversion.' *Journal of Philosophy* 66 (1): 5–17.

_____ (1970). *The Possibility of Altruism*. 86. Oxford: Clarendon Press.

Nickel, Philip J. (2007). 'Trust and Obligation-Ascription.' *Ethical Theory and Moral Practice* 10 (3): 309–19.

O'Neill, Colin (2017). 'Betraying Trust.' In Faulkner and Simpson 2017, 70–89.

Olson, Jonas (2004). 'Buck-Passing and the Wrong Kind of Reasons.' *Philosophical Quarterly* 54 (215): 295–300.

_____ (2009). 'The Wrong Kind of Solution to the Wrong Kind of Reason Problem.' *Utilitas* 21 (2): 225–32.

Owens, David (2006). 'A Simple Theory of Promising.' *Philosophical Review* 115 (1): 51–77.

_____ (2012). *Shaping the Normative Landscape*. Oxford: Oxford University Press.

Parfit, Derek (1984). *Reasons and Persons*. Oxford: Oxford University Press.

_____ (2011a). *On What Matters*. Vol. 1. Oxford: Oxford University Press.

_____ (2011b). *On What Matters*. Vol. 2. Oxford: Oxford University Press.

Pettit, Philip (1995). 'The Cunning of Trust.' *Philosophy and Public Affairs* 24 (3): 202–25.

Pettit, Philip and Victoria McGeer (2017). 'The Empowering Theory of Trust.' In Faulkner and Simpson 2017, 14–34.

Pettit, Philip, Samuel Scheffler, Michael Smith, and R. Jay Wallace, eds. (2004). *Reason and Value: Themes From the Moral Philosophy of Joseph Raz*. Oxford: Oxford University Press.

Piller, Christian (2001). 'Normative Practical Reasoning.' *Aristotelian Society Supplementary Volume* 75 (1): 175–216.

_____ (2006). 'Content-Related and Attitude-Related Reasons for Preferences.' *Royal Institute of Philosophy Supplement* 81 (59): 155–82.

Prichard, H. A. (1949). 'Does Moral Philosophy Rest on a Mistake?.' In *Moral Obligation: Essays and Lectures*, 1–17.

Rabinowicz, Wlodek and Toni Rønnow-Rasmussen (2004). 'The Strike of the Demon: On Fitting Pro Attitudes and Value.' *Ethics* 114 (3): 391–423.

_____ (2006). 'Buck-Passing and the Right Kind of Reasons.' *Philosophical Quarterly* 56 (222): 114–20.

Radzik, Linda (2009). *Making Amends: Atonement in Morality, Law, and Politics*. Oxford: Oxford University Press.

Rapoport, Anatol and Albert M. Chammah (1965). *Prisoner's Dilemma: A Study in Conflict and Cooperation*. Ann Arbor, Mich.: University of Michigan Press.

Rawls, John (1971). *A Theory of Justice*. Cambridge, Mass.: Harvard University Press.

Raz, Joseph (1984). 'Legal Rights.' *Oxford Journal of Legal Studies* 4 (1): 1–21.

_____ (1986). *The Morality of Freedom*. Oxford: Oxford University Press.

_____ (1999). *Practical Reason and Norms*. Oxford: Oxford University Press.

_____ (2000). 'The Truth in Particularism.' In Brad Hooker and Margaret Olivia Little (eds.). *Moral Particularism*. Oxford: Oxford University Press, 48–78.

Reibetanz, Sophia (1998). 'Contractualism and Aggregation.' *Ethics* 108 (2): 296–311.

Reid, Thomas (2010). *Essays on the Active Powers of Man*. Edinburg: Edinburgh University Press.

Rödl, Sebastian (2014). 'Intentional Transaction.' *Philosophical Explorations* 17 (3): 304–16.

Rosati, Connie S. (2003). 'Agency and the Open Question Argument.' *Ethics* 113 (3): 490–527.

_____ (2016). 'Agents and "Shmagents".' *Oxford Studies in Metaethics* 11: 182–213.

Rosen, Gideon (2014). 'Culpability and Duress: A Case Study.' *Aristotelian Society Supplementary Volume* 88 (1): 69–90.

Rowland, Richard (2013). 'Wrong Kind of Reasons and Consequences.' *Utilitas* 25 (3): 405–16.

_____ (2017). 'Reasons or Fittingness First?.' *Ethics* 128 (1): 212–229.

Salje, Léa (2017). 'Thinking About You.' *Mind* 126 (503): 817–40.

Samuelsson, Lars (2013). 'The Right Version of the Right Kind of Solution to the Wrong Kind of Reason Problem'.' *Utilitas* 25 (3): 383–404.

Scanlon, T. M. (1998). *What We Owe to Each Other*. Cambridge, Mass.: Harvard University Press.

_____ (2007). 'Wrongness and Reasons: A Re-Examination.' *Oxford Studies in Metaethics* 2: 5–20.

_____ (2008). *Moral Dimensions: Permissibility, Meaning, Blame*. Harvard University Press.

_____ (2011). 'How I Am Not a Kantian.' In Parfit 2011, 116–39.

_____ (2012). 'Interpreting Blame.' In Coates and Tognazzini 2013, 84–99.

_____ (2013). 'Giving Desert its Due.' *Philosophical Explorations* 16 (2): 1–16.

_____ (2014). *Being Realistic About Reasons*. Oxford: Oxford University Press.

Schafer, Karl (2016). 'The Modesty of the Moral Point of View.' In Lord and Maguire 2016, 241–56.

Scheffler, Samuel (2001). 'Relationships and Responsibilities.' In *Boundaries and Allegiances*. Oxford: Oxford University Press, 97–110.

_____ (2004). 'Projects, Relationships, and Reasons.' In Pettit, Scheffler, M. Smith, and Wallace 2004, 247–69.

_____ (2010). 'Morality and Reasonable Partiality.' In Feltham and Cottingham 2010, 98–130.

Schroeder, Mark (2005). 'Cudworth and Normative Explanations.' *Journal of Ethics and Social Philosophy* 1 (3): 1–28.

_____ (2007). *Slaves of the Passions*. Oxford: Oxford University Press.

_____ (2010). 'Value and the Right Kind of Reason.' *Oxford Studies in Metaethics* 5: 25–55.

_____ (2012). 'The Ubiquity of State-Given Reasons.' *Ethics* 122 (3): 457–88.

Shafer-Landau, Russ (2003). *Moral Realism: A Defence*. Oxford: Oxford University Press.

Shah, Nishi (2006). 'A New Argument for Evidentialism.' *Philosophical Quarterly* 56 (225): 481–98.

Sharadin, Nathaniel (2016). 'Reasons Wrong and Right.' *Pacific Philosophical Quarterly* 97 (2): 371–399.

Sheinman, Hanoch, ed. (2011). *Promises and Agreements*. Oxford: Oxford University Press.

Sher, George (2013). 'Wrongdoing and Relationships: The Problem of the Stranger.' In Coates and Tognazzini 2013, 49–65.

Shiffrin, Seana Valentine (2008). 'Promising, Intimate Relationships, and Conventionalism.' *Philosophical Review* 117 (4): 481–524.

Silverstein, Matthew (2015). 'The Shmagency Question.' *Philosophical Studies* 172 (5): 1127–42.

Simpson, Thomas W. (2012). 'What Is Trust?.' *Pacific Philosophical Quarterly* 93 (4): 550–69.

Singer, Peter (1972). 'Is Act-Utilitarianism Self-Defeating?.' *Philosophical Review* 81 (1): 94–104.

Skorupski, John (2010). *The Domain of Reasons*. Oxford: Oxford University Press.

Smith, Matthew Noah (2010). 'Reliance.' *Noûs* 44 (1): 135–57.

Sommers, Christina Hoff (1986). 'Filial Morality.' *Journal of Philosophy* 83 (8): 439–456.

Strawson, Peter F. (2003). 'Freedom and Resentment.' In Gary Watson (ed.). *Free Will*. Oxford: Oxford University Press, 72–93.

Thompson, Michael (2004). 'What is It to Wrong Someone? A Puzzle About Justice.' In Pettit, Scheffler, M. Smith, and Wallace 2004, 333–84.

Thomson, Judith Jarvis (1990). *The Realm of Rights*. Cambridge, Mass.: Harvard University Press.

Velleman, David (1989). *Practical Reflection*. Princeton, N.J.: Princeton University Press.

_____ (1996). 'The Possibility of Practical Reason.' *Ethics* 106 (4): 694–726.

_____ (2000). 'On the Aim of Belief.' In *The Possibility of Practical Reason*. Oxford: Oxford University Press, 244–81.

_____ (2009). *How We Get Along*. Cambridge: Cambridge University Press.

Wallace, R. Jay (forthcoming). *The Moral Nexus*. Princeton, N.J.: Princeton University Press.

_____ (1996). *Responsibility and the Moral Sentiments*. Harvard University Press.

_____ (2011). 'Dispassionate Opprobrium.' In Wallace, Kumar, and Freeman 2011, 348–72.

_____ (2012). 'Duties of Love.' *Aristotelian Society Supplementary Volume* 86 (1): 175–98.

_____ (2013). 'The Deontic Structure of Morality.' In David Bakhurst, Margaret Olivia Little, and Brad Hooker (eds.). *Thinking About Reasons: Themes From the Philosophy of Jonathan Dancy*. Oxford: Oxford University Press, 137–68.

Wallace, R. Jay, Rahul Kumar, and Samuel Richard Freeman, eds. (2011). *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*. Oxford: Oxford University Press.

Watson, Gary (2004). 'Responsibility and the Limits of Evil: Variations on a Strawsonian Theme.' In *Agency and Answerability: Selected Essays*. Oxford: Oxford University Press, 219–59.

Williams, Bernard (1981). 'Persons, Character, and Morality.' In *Moral Luck: Philosophical Papers, 1973-1980*. Cambridge: Cambridge University Press, 1–19.

Williamson, Timothy (2000). *Knowledge and its Limits*. Oxford: Oxford University Press.

_____ (2017). 'Acting on Knowledge.' In J. Adam Carter, Emma C. Gordon, and Benjamin W. Jarvis (eds.). *Knowledge-First: Approaches in Epistemology and Mind*. Oxford: Oxford University Press, 163–81.

Wolf, Susan (2011). 'Blame, Italian Style.' In Wallace, Kumar, and Freeman 2011, 332–47.

Author Index