Stage Setting
Distinctions and Starting Points

This chapter provides an overview of the main concepts and distinctions about reasons in the literature. It starts with the less contentious distinctions and moves on to the more debated notions and distinctions. Another function of this chapter is to establish the boundaries for the discussion to come. It introduces and sets aside a number of notions and debates that are or have been somewhat important within the general philosophy of reasons but will not fall directly within the scope of working out an informative account of what normative reasons are. Yet, having an idea of these notions and debates will be useful when we turn to (reductive) accounts of normative reasons.

1.1 Normative, Motivating, Explanatory Reasons

The first thing that philosophers typically observe when discussing reasons is that the word 'reason' (as a countable noun) has a multitude of distinct uses in our ordinary language and that our common-sense judgments can vary a lot when considering different cases involving reasons. According to one popular view and following our ordinary language use and common-sense judgments, there are two or perhaps three distinct roles (or kinds) of reasons.[1] We are focusing here on 'reason' as a countable noun only and setting aside 'reason' as a mass noun, which seems to refer most prominently to a faculty or a general disposition, somehow related to rationality. So, according to the distinction in question, 'reason' (a countable noun) can refer to (i) a consideration that speaks somehow in favour of an action or an attitude (e.g. a belief, an emotion). This use is exemplified in constructions having 'a reason to' + a verbal construction in infinitive form – such as 'that the gas tank indicator shows that the car is almost out of gas is a reason for John to stop at the gas station', 'that the weather

---

[1] For paradigmatic examples of this distinction, see Alvarez (2010, 2016) and Engel (2015a).

forecast announces heavy rain for tomorrow is a reason for Mara to postpone the hike', or 'that John's car is parked in front of Mara's house is a reason to think/believe that John is at Mara's place'. Constructions involving verbs in the infinitive are characteristic of expressions that refer to normative aspects, such as oughts, obligations, or permissions (see Hawthorne and Magidor 2018 for more on this) – for example, 'we must go' or 'you are allowed to eat a cake'. Thus, it is not surprising that expressions with the structure 'a reason to' combined with a verb in the infinitive are commonly understood as bringing in a normative aspect too. Thus, the first kind of reasons (or, alternatively, the first role of reasons) is the normative kind/role of reasons. Reasons of this sort are called 'normative reasons' and are the primary object of our investigations in what follows.

In the second place, according to the popular view, 'reason' (a countable noun) can also refer to (ii) a consideration on the basis of which one acts or has an attitude. This use is exemplified in our ordinary talk about *one's reasons for* which one acts or has an attitude, as, for instance, in 'John's reason for parking in front of Mara's place was that he couldn't find a free spot elsewhere,' or 'Mara's reason for thinking that Zoe will come was that she said she will.' Reasons in this sense are the considerations that have actually played (or can potentially play) the role of the foundation of one's actions and attitudes from one's own perspective. It is common to call reasons of this kind 'motivating reasons'. Of course, it is also widely agreed that if we focus on reasons for attitudes and not exclusively on reasons for action, then explaining reasons in this sense by reference to motivation as ordinarily understood might be a bit misleading. For, typically, it is conceded we don't have a motivation for believing this or that, yet our beliefs are based on some considerations – that is, we believe often (if not always) for some reason. Thus, some philosophers prefer to use the label 'operative reasons' to refer to this kind of reasons (see Scanlon 1998: 19). In what follows, we stick with the established use and talk about 'motivating reasons', assuming the relevant technical sense (e.g. the basis for S's F-ing from S's own perspective) and not the ordinary sense of 'motivating'. Yet, note that the technical sense here might not be so alien to a somewhat archaic sense of 'motives'. Consider, for instance, the Scholastic notion of *motiva credibilitatis* (motives of credibility) understood as a basis for rational faith that can be discovered by reason alone (without divine revelation), such as considerations about the origin of the universe as speaking in favour of the existence of God constituting motives of credibility for theism.

Finally, we may also think that 'reason' (a countable noun) can also refer to (iii) considerations that explain or contribute to explaining why an action/attitude occurred. This use seems to be exemplified by our talk of *reasons why*, as, for instance, in 'the reason why I was late was that I was stuck in traffic', 'the reason why she was not selected for the job was that the hiring committee was biased', or 'that you grew up in a religious environment is the reason why you believe in God'. The referents of 'reasons' in this sense are commonly called 'explanatory reasons', since they explain or participate in explaining why one acts in the ways one does or why one has the attitudes one has. Of course, it is also the case that typically considerations that play the role of motivating reasons (considerations on the basis of which one acts/has an attitude from one's own perspective) will also help explain the action or the attitude in question. That John couldn't find a free spot elsewhere explains (partly) why he parked in front of Mara's place. That Zoe told Mara she will come explains at least partly why Mara believes she will come. However, as the aforementioned examples demonstrate, not distinguishing the two, the motivating and the explanatory reasons, would leave a number of cases unexplained. Some considerations seem to be able to play an explanatory role for actions or attitudes without also being considerations on the basis of which one acts or has an attitude from one's own perspective. That I was stuck in traffic can explain why I am late, but it is not a consideration on the basis of which I base my being late. Similarly, one will hardly accept that one's own implicit biases and prejudices are reasons for which one acts in the ways one does, yet they may still be part of an explanation for one's actions and attitudes. Also, as Maria Alvarez (2016) has observed, some cases seem to be best interpreted as cases where one acts for no reason at all, yet there seem to be reasons why one acts in the ways one does. Consider, for instance, the case of one going for a run for no specific reason, or one deciding to grow a beard for no reason, just on a whim. Even if there are no reasons *for* which one does these things, we may still come up with some plausible reasons why one does the things in question.[2] Conversely, one might also think that, at least in some contexts, considerations for which we act don't really contribute to explaining our action. Dramatic cases of implicit bias might instantiate such a possibility. That candidate A has all the skills that are required for the job might be one's motivating reason, the consideration for which one selects candidate A. And yet, arguably, that candidate A has all the necessary skills doesn't

---

[2] See Hawthorne and Magidor (2018) for further examples and discussion.

even partially explain why one selects candidate A, given that candidate B is equally skilful. The choice in this case is explained by one's implicit bias against B. Thus, one might think there are three distinct kinds of or roles for reasons (understood as references to the countable noun 'reason'): normative reasons, motivating reasons, and explanatory reasons.

Note, however, that the case for distinguishing motivating from explanatory reasons gets somewhat complicated if we reject the idea that deliberation (or reasoning) is necessary for acting for reasons (and for having attitudes for reasons). If deliberation and acting (having an attitude) for a reason can come apart, as recently argued by Arpaly and Schroeder (2012), then presumably some of the aforementioned cases might lose their appeal. Think, for instance, about the case of going for a run for no apparent reason or the case of deciding to grow a beard just on a whim. Perhaps these only appear to be cases where one doesn't have a reason for doing what one does, because these are cases where one doesn't undertake prior deliberation that concludes in the relevant action/intention. But once we distinguish prior deliberation from acting for a reason and admit of the possibility of action for a reason without prior deliberation, we may hesitate to conclude that these are genuine cases where one has no reason for doing what one does. At any rate, we don't aim to solve this debate here. The crucial point for what follows is that normative reasons – that is, the object of our investigation here – are commonly distinguished from motivating reasons. The question of whether motivating reasons and explanatory reasons should be further distinguished lies beyond the scope of the present work.

1.2   A (Recent) History of Reasons in Three Acts (and the Reasons–Causes Distinction)

Reasons are central to contemporary philosophy. They play a prominent role in contemporary debates in meta-ethics, epistemology, and philosophy of action, to name but a few. Yet talk of reasons hasn't always been so popular in philosophy. How did the notion of *reasons* become so central? This section proposes one possible reconstruction of the rise to prominence of *reasons* by distinguishing three major episodes that have shaped the understanding of this notion over the twentieth and twenty-first centuries. The story proposed here is a rough one, and as such it certainly misses some important historical details and oversimplifies some complex theoretical debates. Yet the aim here is not to provide an exhaustive genealogy of *reasons* or a complete history of debates in normative

philosophy; rather, it is to draw attention to some aspects of past debates that may help us to better situate the object of our inquiry (i.e. the nature of normative reasons). A central theme in our reconstruction of the rise to prominence of *reasons* will be the supposed difference between reasons and causes.[3]

The debate that inspired the contemporary interest in reasons, the first episode in our reconstruction of contemporary reasonology, is exactly the early twentieth-century debate about the distinction between reasons and causes. According to late Wittgenstein (in particular in his Blue Book: Wittgenstein 1958), we can distinguish between two kinds of explanations: on the one hand, there are explanations by causes, and on the other hand, there are explanations by reasons.[4] One way of further developing these thoughts from Wittgenstein is to conclude that only the explanation by reasons is appropriate in explaining action. This conclusion, drawn explicitly by Wittgenstein's followers (see Melden 1961, and Anscombe 2000 in particular), was thought to have far-reaching theoretical consequences. To see the importance of this distinction, we have to revisit the early twentieth-century discussions concerning the unity and methodology of sciences, and, in particular, of the nature of scientific explanation. According to one trend in the philosophy of science, inspired by Carl Hempel, a genuinely scientific explanation is causal, and crucially it must appeal to empirical laws. A notable and controversial consequence of this positivist view is that, if social sciences, psychology, and the humanities (e.g. history) are genuinely scientific disciplines, then they

---

[3] In my working through this, I am much indebted to Pascal Engel, especially with respect to a better understanding of the Davidsonian approach and its discontents (in particular, I am inspired here by 'la Présentation' in Davidson 1993, translated by Engel 1993: V–XXXI). My general overview here is also partly inspired by Livet (2005) and Ogien (1995: 35–51). See also Wiland (2012), especially chapter 1, for some relevant historical references beyond the twentieth century in connection with the faculty of Reason, which is assumed to be connected to reasoning and hence to judging and doing things for reasons (cf. Wiland 2012: 9). The focus of the present section is on contemporary history only.

[4] Arguably, this distinction between reasons and causes has its predecessor in the Fregean distinction between grounds that can justify our judgments and mere causes of our judgments. Frege distinguishes the two in his discussion of psychologism in logic; see, for instance

> With the psychological conception of logic we lose the distinction between grounds that justify a conviction and the causes that actually produce it. This means that a justification in the proper sense is not possible; what we have in its place is an account of how the conviction was arrived at, from which it is to be inferred that everything has been caused by psychological factors. This puts a superstition on the same footing as a scientific discovery. (Frege 1979: 147)

I discovered this passage from Frege in Benmakhlouf (2018), who suggests that Frege distinguishes here between reasons and causes; see Benmakhlouf 2018: 13–14.

must also conform to this model, and they have to provide causal explanations while appealing to empirical laws (or universal generalisations). This, within the positivist framework, amounted to quite problematic behaviourist assumptions in the human sciences. Thus, taking Wittgenstein's remarks about explanation by causes and explanation by reasons at face value in this context leads to a rejection of the positivist model of scientific explanation as inadequate, or at least as inadequate with respect to human and social sciences. For, if the distinction between causes and reasons is on the right track, positivists' assumption about the necessarily causal nature of explanation is radically misguided when considering a (scientific) explanation of action, which is presumed to be connected to reasons and not causes. In this context of the early twentieth century, then, the nature of reasons as opposed to causes becomes a central philosophical battlefield with respect to debates about human action, scientific methodology, and the unity of the sciences more generally.

It is, then, no surprise that Wittgenstein's followers invested so much energy and effort in working out in detail positive arguments in favour of distinguishing reasons from causes. Roughly, according to the Wittgensteinians, one can give at least three distinct lines of argument in favour of the distinction between causes and reasons. (Wittgenstein himself, in fact, does not enter much into argumentative details concerning the distinction, and even seems to accept the possibility of reasons that are causes without explaining how this might be.) The first kind of argument relies on some epistemological assumptions. In short, it begins with an assumption that causal explanations are discoverable by observation. More specifically, causal explanations can be reached by means of repetitions of experiments that confirm or give a basis for rejecting a hypothesis. Typically it happens when we (i) observe an effect E, following an event C; (ii) we repeat the event; and (iii) we observe the effect again. This then provides the basis for inferring a causal correlation between C and E. This is thought to be a standard procedure in discovering a cause and hence uncovering a causal explanation. However, when it comes to explanations by reasons, the procedure for reaching them is nothing like this. We do not discover our reasons for acting by repeated observations that can confirm the hypothesis that such and such is a reason to act. On the contrary, according to the Wittgensteinians, we have non-observational knowledge of our reasons (see Anscombe 2000: 14). This non-observational knowledge is, according to this line of thought, of the same nature as our knowledge of our own body. This epistemological difference between causes and reasons, then, is thought to uncover a

substantial difference between the two that should give us pause in lumping the two together. The second consideration that has been proposed by Wittgensteinians in favour of distinguishing reasons from causes is of a conceptual nature. It begins with an assumption that there exists an intrinsic link between reasons and actions. The idea is that the very concept of *action* presupposes or subsumes the concept of *reasons*. In other words, we cannot understand action (i.e. grasp the concept of *action*) without grasping the concept of *reasons*. The intrinsic link between the two is supposed to be of the same order as the conceptual link between the concept of a *premise* and the concept of a *conclusion*. Understanding one involves understanding the other. Yet, crucially, there is no such intrinsic link between *cause* and *effect*. Indeed, according to this line of thought, we can quite easily conceive of a cause without conceiving of its effect. The third line of argument in favour of the reasons–causes distinction relies on an observation about the ontology of causes. The observation here is that causal chains can be, in principle, infinite. That is to say, there is nothing in principle that would prevent the possibility of infinite regress in the search for causes of an event (and, if the past is infinite, then such an infinite regress in causal chains even makes sense). However, this is not the case with respect to reasons. Clearly, reasons do not admit even in principle the possibility of infinite regress. Reasons are always reasons for someone, and they do stop at the level of someone. One way to think about it would be to think of reasons as pointing to someone's responsibility for doing something. Once one's responsibility for doing something is established, it doesn't make sense to look further. So, for instance, if you are going to explain my action of going to the grocery store, you might appeal to the fact that I am out of milk or beer, or whatever, but once you have established the relevant considerations (that I am out of milk and need some more, etc.), there is no sense in going further back in explaining *my* action of going to the grocery store by, say, appealing to considerations about why I am living where I do or why I eat what I do.

Whatever the value of these arguments, the point that I would like to observe here is that the modern interest in reasons seems to have first been initiated in the context of the opposition between Wittgenstein and especially his followers to the logical positivist view of methodology and the unity of the sciences, and in particular about the nature of an appropriate explanation of action. The question of whether the explanation of action has to appeal to reasons, as opposed to causes, was of crucial importance to that debate in the early and mid-twentieth century.

In the heyday of Wittgenstein-inspired philosophy of action (i.e. the mid-twentieth century), it seemed that the validity of the distinction between reasons and causes had been established for good. However, things changed rapidly when Donald Davidson came up with his influential – in a sense, neo-Humean – approach in philosophy of action (see Davidson 1963). He challenged established Wittgensteinian views in philosophy of action, and a key element of his attack was the rejection of their reasons–causes dichotomy as over-simplistic. Davidson's novel take on the relation between reasons and causes in the 1960s and the discussion that unfolded after it, in particular in moral psychology, constitute the second major episode in our rough reconstruction of the recent history of reasons.

A central element in Davidson's rejection of the simple reasons–causes dichotomy was to show that, contrary to what his opponents seemed to take for granted, reasons for which a subject acts are actually causes of the subject's action. Reasons for which a subject acts are the bases of one's action and as such are causes of one's action. Another important element in Davidson's view is his adoption of a sort of neo-Humean approach to action, according to which, roughly, action is explained by appeal to the belief–desire pairs. On this view, then, my desire to drink a beer, combined with my belief that there is a beer in my fridge, is my reason for getting up and heading to the fridge. This combination is a reason and also a cause that explains my action. Crucially, the explanation is causal. The desire–belief pair is the relevant cause. In providing this line of thought, Davidson is also rejecting another common assumption in the earlier debates, namely, that causal explanations have to appeal to some general laws and have to be discoverable by repetitions of experimental observations. Such a view of causal correlations is overly restrictive on Davidson's approach, for there are some singular causes. We do not need multiple observations to know that pouring nail polish on my pants will ruin them instantly. A single experience of this kind is largely sufficient to know this. Thus, general laws and multiple observations are not necessary for establishing a causal correlation. Certain reasons, reasons for which we act (e.g. our psychological states), are causes of our actions in this sense according to the Davidsonian approach.[5]

The debate between Davidson and Wittgenstein's followers focused much on motivation in debating the correct account of action. An interesting point to note is that the parallel debates in meta-ethics at that time

---

[5] For a recent critical discussion of the Davidsonian idea that reasons are causes, see Dietz (2016).

were also largely focused on aspects of moral psychology. A central meta-ethical debate at that time concerned the correct account of motivation (see Scanlon 2014: 1–2 for related historical observations about the centrality of morality and in particular the centrality of the question of moral motivation to the mid-twentieth-century debates in meta-ethics). This focus can be explained in part by the growing interest in non-cognitivism in meta-ethics during the 1970s. A central argument for non-cognitivism – the approach according to which, in a nutshell, moral judgments such as 'this action is/is not morally appropriate' are expressions of non-cognitive states (e.g. desires) – relies on the very idea that moral judgments should motivate us to act and that only desires (given certain background beliefs) can motivate us to act in a certain way. Thus, a Davidsonian (and broadly neo-Humean) mentalistic theory of reasons, where reasons are psychological states (e.g. desires with background beliefs), becomes central to meta-ethical debates, for it provides crucial construction blocks for non-cognitivist arguments in meta-ethics (and moral psychology). In short, these debates, initiated in a sense by the Davidsonian rejection of an oversimplified dichotomy between reasons and causes, have the notion of *reasons* as a key element and in that sense have certainly contributed to placing this notion at the very epicentre of contemporary normative fields of philosophy.

The third major episode in the growth in popularity of *reasons* as I see it corresponds to the increasingly widespread realist objections to mentalist, non-cognitivist, or psychologising approaches in meta-ethics, starting in the 1990s and continuing well into the twenty-first century. Indeed, many of these objections are directly based on a divergent understanding of the very nature of reasons. In short, we can observe a certain anti-Humean movement in meta-ethics towards the end of the 1990s and in the 2000s. It is characterised by turning the focus specifically on reasons understood as facts, facts that speak in favour of an action or an attitude. Authors like Derek Parfit, Thomas Scanlon, and John Skorupski are among some of the most important players in this realistic turn. Instead of focusing on reasons for which we act (with their link to motivation), these philosophers have insisted that we focus on reasons to act (or have an attitude). Reasons to act (and have attitudes) are, in a sense, independent of our motivations and psychological states. According to this approach, the fact that there is a fire in the building is a reason for everyone in the building to get out of it immediately. Crucially, it is a reason regardless of the psychological states of people inside the building. A further thought, then, is that reasons understood in this way are considerations that play a central or even

definitional role in determining what one ought to do, what is right or fitting, what is justified or rational, and so on. Those who take this further step often think of reasons as fundamental and prime elements in the normative domain, as things that cannot be further analysed but can be appealed to in order to understand other normative notions and statuses (the next section provides a more detailed explanation of this prominent approach within contemporary philosophy).

We may sometimes hear contemporary philosophers complaining about the omnipresence of reasons in recent debates – the worry being that *reasons* could mean so many different things and that philosophers often use it in ways that don't seem to exhibit any unity. And indeed a healthy scepticism about putting too much emphasis on this notion may be sensible, given the proliferation of distinctions and ways of understanding it. However, to give up on reasons too quickly is also to forget the tremendous importance that this notion has had in past debates, starting at least in the early twentieth century. We have surveyed above what seem to be three major episodes in philosophical debates where the notion of *reasons* has played a crucial role: the debate between logical positivists and Wittgensteinians on explanation of action; the debate between Wittgensteinians and Davidson about the reasons–causes distinction; and the debate between neo-Humeans and realists in meta-ethics. All these debates have relied in one way or another on appeals to reasons and to some extent have been fruitful. It is apparent that the prominence of reasons in contemporary philosophy has not come out of the blue; it is largely due to the role of reasons in these past debates. Insofar as these debates have not all been futile, we can conclude that theorising about reasons is a worthwhile endeavour.

1.3 Objective, Subjective, Possessed, Unpossessed

1.3.1 Introducing the Problem

Contemporary philosophy of reasons is characterised by an increasing number of notions and distinctions. While it is certainly true that some crucial notions are universally accepted (such as the distinction between *normative* and *motivating* reasons), it is also true that recent reasonology debates have become increasingly idiosyncratic. One response to such idiosyncrasy might be to appeal to Ockham's razor by investigating whether and how some notions within the debate can be reduced to others. So, for instance, instead of having both the concept of *motivating*

reasons and the concept of *operative* reasons, philosophers are willing to accept that, in fact, people have used two different names for the same concept.

In such a context, the tendency is to assume the same kind of reduction with respect to *subjective* reasons (as opposed to *objective*) and *possessed* reasons (as opposed to *unpossessed*). According to this assumption, there is no distinction between possessed reasons and subjective reasons at all: thus, there are no reasons that are subjective but not possessed, and equally no reasons that are possessed but not subjective. Rather, just as 'motivating' and 'operative' are two names for the same thing – that is, a certain sort of reason – so 'subjective' and 'possessed' are two names for a sort of normative reason.[6] Here are two [explicit] examples attesting to the popularity of the identification of subjective with possessed reasons:

> . . . But in some sense or other, Freddie [i.e. the guy who likes to dance and knows that there is going to be dancing at the party], unlike Ronnie [i.e. the guy who likes to dance but has no clue about the party], *has* this reason [to go to the party], since he knows about it, and Ronnie does not. This second sense of 'has a reason' is the one I will later distinguish as the *subjective* sense of 'reason'. (Schroeder 2008: 59)

> I reject the Factoring Account, so I deny that subjective reasons are a subset of objective reasons. I also deny that the status of something as a subjective reason is independent of its being possessed. (Smithies 2018: 20 fn 29)[7]

Not everyone within the debate accepts that there are any objective/unpossessed normative reasons. Some think that all normative reasons are *subjective/possessed*. Yet, even in this case, the assumption is often that the distinction makes sense and, in particular, that 'subjective' and 'possessed' are merely two different names for one set of normative reasons, whereas 'unpossessed/objective' (or perhaps simply 'unpossessed')[8] refers to another

---

[6] Hawthorne and Magidor (2018) is one rare place where this identification has been rejected. They rely on this rejection in their objection to Mark Schroeder's arguments against the so-called Factoring Account of reasons. Sylvan (2016) also appears to distance himself from the identity view.

[7] Here is another example: 'One common way of drawing the distinction [between "objective" and "subjective" reasons] is in terms of the reasons there are for some agent S to A (so-called objective reasons) and the reasons S has to A (so-called subjective reasons)' (Fogal and Sylvan 2017: 6, fn 6). Note that Fogal and Sylvan do not endorse such a distinction themselves.

[8] This is to indicate that I do not want to imply that the approach that I criticise in this section has to assume that the sets of subjective and objective reasons are disjoint. My opponents may well hold that subjective reasons constitute a subset of objective reasons and that it makes sense to distinguish between possessed objective reasons (i.e. subjective reasons) and unpossessed objective reasons (i.e. merely objective reasons). Thanks to an anonymous referee for this suggestion.

[possible] set of normative reasons. That is, it makes sense to accept a distinction between two *sorts* of reasons even if, in fact, nothing falls under one term or the other.

The aim of this section is to challenge the identification of *subjective* with *possessed*. I will argue that *subjective* reasons are distinct from *possessed* reasons. More precisely, I will argue that while there are possessed subjective reasons, there are also unpossessed subjective reasons.[9] The distinction between subjective and possessed reasons is not like the [merely apparent] distinction between motivating and operative reasons. In what follows, I will first review the basics of one theoretical framework that appears to imply the distinction between subjective and possessed reasons. Then, I propose intuitive considerations in favour of the distinction, before indicating some of its further theoretical implications.

1.3.2   A Background Theory

To keep things clear, let us focus on the following senses of 'possessed' and 'subjective' reasons:

> **Possessed reasons (PR)**: a reason *r* for S to F is possessed by S, just in case S is in a position to use *r* in deliberation.[10]
>
> **Subjective reasons (SR)**: a reason *r* for S to F is a subjective normative reason for S to F just in case *r*'s being a reason for S to F depends on S's evidence.[11]

---

[9] While I believe that *subjective versus objective* and *possessed versus unpossessed* are orthogonal and not merely different distinctions, in what follows I will focus exclusively on the *subjective* (*possessed or unpossessed*) reasons.

[10] Compare to: '[T]here is a familiar distinction between the reasons there are for a person to act and the reasons she possesses for acting, where a person possesses a reason for acting only if she is in a position to act for that reason . . . A common suggestion is that to possess a reason requires standing in an epistemic relation to the relevant consideration' (Whiting 2018: 3). See also: 'The contrast between possessed and unpossessed reasons we have in mind is fairly intuitive. When a glass contains poison but an agent is unaware of this, there is a reason for the agent to avoid drinking from the glass, but that reason for avoidance is something that the agent is not in a position to use as a consideration when acting' (Hawthorne and Magidor 2018: 2).

[11] See, for instance:

> Roughly our envisaged objective 'ought' ranks actions according to the best outcome, while the subjective 'ought' ranks according to the best expected outcome by the lights of the agent's evidence . . . (We think of evidence as what the subject knows, though much of what we say could be adapted to other frameworks for thinking about evidence.) This basic structure covers both reasons to act and reasons to believe. (Hawthorne and Magidor 2018: 2)

Compare to: 'Second, it is fairly common in the literature we are concerned with (and elsewhere) to distinguish between *objective* and *subjective* "should's". On a simple way of drawing this distinction, what you objectively should do is determined by the facts of your situation, whereas

Three clarifications are in order here. First, note that ordinary English possessive constructions such as 'S has a reason to F' or 'S's reason to F' are extremely context-sensitive and do not always capture the sense of 'possession', which is of interest for normative debates.[12] Consider (a) 'The building is on fire. She has a reason to leave.' The reason attribution here may be appropriate in a sense, and, importantly, the two phrases may be true, even if the subject has no clue whatsoever about the ongoing fire.[13] Crucially, the sense of 'having' reasons in (a) is not the one that we intend to capture by (PR) (nor by [SR]). In short, the context sensitivity of possessive constructions calls for extra caution when relying on linguistic data to theorise about the possession of reasons. Our specification of 'possessed reasons' does not pretend to correspond to all possible uses of possessive constructions involving 'reasons'.

Second, 'being in a position to do something' is a context-sensitive expression. Consider, for instance, (i): 'Carl is in a position to prove Gödel's Incompleteness Theorem.' The utterance (i) may well be true in a context where we focus on, say, comparing typical humans to some cognitively less developed species. There is a sense in which a human, named Carl, and not, for instance, a jellyfish, is in a position to prove the Incompleteness Theorem. This is a very weak sense, a sense that attributes the ability to prove the theorem to Carl merely because he is a human with a sophisticated cognitive capacity (and not a jellyfish). In a different context, (i) will not come out as true. For instance, it will be false in a context where we focus on Carl's ignorance of mathematics and logic. The sense of 'is in a position to', which is relevant for our discussion here, is not a weak one. We can follow Whiting, Hawthorne, and Magidor (see footnote 10) and think of 'is in a position to use r in deliberation' as introducing an epistemic constraint. It requires that the subject's epistemic situation does not prevent the subject from using the relevant consideration in deliberation. Crucially, the mere fact that p follows from the subject's evidence does not guarantee that the subject is in a position to use p in her deliberation in the relevant sense. After all, one may believe or

---

what you subjectively should do is determined by your perspective on your situation' (Way 2018: 14).

[12] Thanks to an anonymous referee for reminding me of the importance of this context sensitivity.

[13] See Hawthorne and Magidor (2018) for similar and more sophisticated examples. An anonymous referee also proposes the following example as an illustration of the context sensitivity in question: 'It turns out that we had good reason to proceed with caution, though of course we could not have known it at the time'. See also Fogal and Sylvan (2017) for further observations about the context sensitivity of the possessive constructions.

know that p and fail to know or even believe a proposition that follows from p.[14]

Third, while 'depends' in (SR) is a bit vague, hopefully it is still clear enough for our purposes here. The crucial point is that 'depends' differs from 'consists of'. More specifically, the point of (SR) is that we do not characterise one's subjective reasons as consisting only of one's evidence. Certain facts that are *determined* by one's evidence will count as subjective reasons on this specification even though they are not themselves part of one's evidence. One might think of a subject's evidence as a set of propositions that the subject knows (e.g. Williamson's E=K thesis).[15] Yet what follows is also compatible with different views about evidence.

One theoretical framework that vindicates the distinction between subjective and possessed reasons is the view that takes seriously the context sensitivity of 'ought' and other modals.[16] According to this linguistically informed approach, 'ought' can have different senses, since it is taken to order actions (and attitudes) according to a standard.[17] There are objective oughts and subjective oughts. The former rank actions (and attitudes) relative to what is best, given all the facts; whereas the latter provides a ranking relative to what is expected to be best in the light of some agent's evidence (cf. Hawthorne and Magidor 2018: 2). The suggestion, then, is to apply a similar line of thought to normative reasons, since the construction 'a reason/reasons to F' is taken to encode the ought modality (cf. Hawthorne and Magidor 2018). A natural conclusion is that 'a reason/reasons to F' is context-sensitive in the same way as 'ought' is. This supports the idea that there are at least two sorts of normative reasons – objective and subjective. Subjective normative reasons are restricted by the subject's evidence, whereas objective normative reasons are not restricted in this way. This approach does not identify subjective normative reasons with possessed normative reasons. For a consideration, r, to be a subjective normative reason for S to F, just is for r to speak in favour of F-ing for S, given the set of S's evidence. In numerous cases, when r is a subjective normative reason for S, S will possess r. Yet, the

---

[14] Thanks to an anonymous referee for drawing my attention to the need to clarify this issue.

[15] Cf. Williamson (2000: 184–237). See also Logins (2017), among others, for a recent defence of E=K.

[16] The view sketched here has been proposed recently by Hawthorne and Magidor (2018).

[17] A more precise formulation would appeal to the rankings of states of affairs. The linguistic theory in the background of this view comes from Kratzer (1981) (and elsewhere). This presentation is a very rough and incomplete one. A number of important elements need to be added here (such as a reference to modal bases or domains of states of affairs that are ranked).

familiarity with such common cases need not lead us to the identification of the two. There might be cases where a consideration r is a normative subjective reason for S to F (it is determined by S's evidence), yet S does not possess r, since S is not in a position to use r in her deliberation.

This was a quick summary of a theoretical view that supports the distinction between subjective and possessed reasons. However, we have not yet seen a convincing case that cannot be well understood unless we adopt the view in question. That is, now that we have filled in a conceptual space, we need some motivation for taking this option seriously. Does it actually have a useful application? The next section aims to explore this question.

1.3.3 A Case

Consider an ordinary participant on the famous *Let's Make a Deal* show (also known as the Monty Hall show).[18] Let us call him Ben. He stands in front of three doors. There is a luxury car behind one of these doors and goats behind the two others. Ben has been given a chance to choose one of the three doors. Let us say Ben chooses door number 1. Now, the show's host, Monty Hall, is obliged to open one of the three doors (that is the rule). Yet he is not permitted to open the door that Ben has chosen. Neither can he open the door with the car behind it. Ben knows the rules of the game. Let us say Monty Hall opens door number 2. Of course, there is a goat behind it. Next, Ben is offered the chance to change his initial choice. That is, Ben can change his choice from door 1 to door 3. Given what Ben knows, the thing to do (as long as he wants to win the car) is to switch to door 3. Given Ben's evidence, it is clearly more probable that the car is behind door 3. In fact, by opening door 2, Monty Hall gave Ben a crucial piece of information. Given the 1/3 probability that Ben's first choice was the lucky one, there was a probability of 2/3 that Monty Hall did not have any choice other than door 2. That leaves a 2/3 chance that the car is behind door 3. Now, it makes sense to think that the consideration (r) 'it is more likely that the car is behind door 3 than door 1' (alternatively, 'the car is more unlikely to be behind door 1') is a normative reason for Ben to switch to door 3. That is, r is a reason for Ben, in a sense.

---

[18] This case is well known (especially within probability theory) as giving rise to the Monty Hall Problem; see vos Savant (1992: 199–209).

It may still be the case that the car is behind door 1. Hence, in another sense (an *objective* sense that is not tied to Ben's evidence), *r* may also be no reason at all to switch. However, as long as we focus on the evidence that Ben possesses, *r* is, intuitively, a normative reason (in a sense) for Ben to switch.

Crucially, Ben has the 'Monty Hall condition' – that is, the condition of being unable to see that it is more probable that the car is behind door 3 than door 1. As a matter of fact, a number of participants on the actual show were unable to see that, given what they knew, it was much more probable that the car was behind door 3 (i.e. the door that was not initially chosen by the participant and was offered as a possible choice for a switch). It may take some time, repeated calculation, reading explanations, and watching tutorials to overcome the 'Monty Hall condition' and finally be able to understand that it is more probable that the car is behind door 3 than door 1. After all, it is common to describe the theoretical choice the participants face on the show (when they have been offered the chance to switch a choice) as a problem or puzzle or even a paradox.

Assuming that Ben has the Monty Hall condition, we do not want to say that he possesses the consideration 'it is more likely that the car is behind door 3 than door 1' as a normative reason. For Ben is not in a position to use it in any kind of deliberation. Crucially, the 'being in a position to', which is relevant here, is an epistemic one, exactly as in (PR) – namely, there is something in Ben's epistemic position that prevents him from using the relevant consideration in his deliberation. After all, Ben does not believe and does not know that it is more likely that the car is behind door 3 than door 1. The distinction between subjective normative reasons and possessed normative reasons seems to be the best way of making sense of the intuitive judgments about this case. The consideration 'it is more likely that the car is behind door 3 than door 1' is a subjective reason for Ben to switch, yet he does not possess it. It is an unpossessed subjective normative reason.

Those who collapse the distinction between the subjective and possessed normative reasons might object to the claim that *r* is a normative reason for Ben to switch. One line of objection to this claim relies on the idea that only facts that determine [evidential] probability and not the probability facts (such as the fact that it is probable that *p*) can be reasons. According to this line of objection, only the relevant pieces of Ben's evidence are reasons for him to switch (for example, that Monty opens door 2, that Monty cannot open door 1, and so on). The fact that, on Ben's evidence, it

is more likely that the car is behind door 3 than door 1 is not a reason, according to this line of thought.[19]

However, this line of objection leads to unacceptable conclusions, since it would generalise to a myriad of ordinary considerations that we typically take to be our reasons to act. Giving up the idea that probability facts, such as that *p* is probable or that *p* is more likely than *q*, can be reasons for a subject to *F* would result in a massive denial of our ordinary intuitions. For example, consider a situation where it is likely, on my evidence, that it will rain in five minutes. Do we really want to say that we are not authorised to hold that the consideration 'it is likely on my evidence that it will rain' is a reason for me to close the windows or to look for my umbrella? To the contrary, the fact that the rain is very probable on my evidence speaks in favour of closing the windows or taking the umbrella. The probability fact here is a normative reason for me to act in certain ways.

The following examples may help to illustrate this point further. Let us say that a mountain expedition has been organised, and the participants are set up at the base camp and planning their ascent to the summit. Given what they know, it is 80 per cent probable that there will be a heavy storm tomorrow. Do we want to say that this probability fact does not speak in favour of not planning the ascent for tomorrow? Given what his doctors know, it is more likely that a patient has the rare, severe, and extremely contagious disease X than a simple flu. That probability fact speaks in favour of the doctors recommending the patient's immediate hospitalisation. It is unlikely, given the available scientifically informed evidence, that a plague epidemic will break out in the city you live in anytime soon. It would seem odd to say that this consideration is not a reason for you to dismiss allegations of a major plague risk in your city. Rejecting the idea that probability facts can be normative reasons leads to an unwarranted scepticism about a large number of ordinary normative reasons.

Hence, I conclude that the Monty Hall example provides at least prima facie support for the possibility of subjective unpossessed reasons.

1.3.4   Theoretical Implications of the Distinction

Taking subjective unpossessed reasons seriously has further theoretical relevance. Here are two places where they might play a significant role.

A major theme in epistemology during the last twenty years or so is our presumed *cognitive homelessness*. Roughly, this view states that our inner

---

[19]   Thanks to an anonymous referee for drawing my attention to this possible objection.

life is not always available to us. According to an influential argument by Williamson (2000: 93–113), there is no non-trivial condition C, such that when C obtains, one can always know that one is in C. In particular, Williamson has argued that it is not the case that when we know that *p*, we always know that we know that *p*. His Anti-Luminosity argument for that conclusion has received sustained interest throughout the early twenty-first century. While it is far from being universally accepted, the Anti-Luminosity argument has, nonetheless, moved contemporary mainstream epistemology towards taking the possibility of our cognitive homelessness more seriously.

The situation within the contemporary meta-normative debates seems to be quite different in this respect. The fact that it is common not to distinguish between subjective and possessed normative reasons seems to reveal the prevailing tendency within the meta-normative field to assume that people's inner lives, and in particular their inner normative lives, are always available to them. However, if what precedes is on the right track, then this common assumption needs to be revisited. The Monty Hall case shows that items within our inner (normative) life (such as subjective reasons) are not always available to us: even people's subjective reasons to act are not something that they are always in a position to know they have.

On a somewhat related note, much debate in contemporary epistemology has centred on the so-called internalism/externalism divide about epistemic justification. Even though the debate does not seem to be over yet, there are substantial lessons that epistemologists have already learned from it. One of these seems particularly relevant for our discussion, namely, that there are two ways of making the general characterisation of internalism more precise. Internalist *accessibilism* is, roughly, the thesis that justification is determined by one's internal states that are accessible (to oneself) upon reflection alone; whereas internalist *mentalism* is the view that justification is determined by one's (internal) mental states, regardless of whether they are accessible or not. Now, it makes an important argumentative difference whether one defends internalist accessibilism or internalist mentalism, and epistemologists have noticed it. A similar contrast seems to apply to normative reasons as well. However, its importance has not yet been fully appreciated. More precisely, the distinction proposed above, between possessed and subjective normative reasons, parallels in a way the accessibilism/mentalism distinction about justification. One may think that practical rationality is determined either by one's subjective reasons or by one's possessed reasons. These are two different theses that

the distinction between subjective–possessed reasons enables us to clearly separate. This matters, since one of these theses seems to be more demanding than the other. At least, certainly, the same arguments will not work for or against both of them. Overall, just as epistemologists have made progress by distinguishing accessibilism from mentalism, meta-normative debates could benefit from taking the possessed/subjective reasons distinction more seriously.[20]

1.3.5 Concluding Remarks about Objective–Subjective Reasons

Philosophy of reasons has reached a point of increased terminological complexity. In such a situation, a natural tendency is to simplify the debate and reduce the number of postulated distinctions. One such tendency is to reduce subjective normative reasons to possessed normative reasons. I have argued here that such a tendency has to be resisted or at least not accepted from the outset. A respectable (yet, of course, debatable) and well-understood theoretical framework licenses such a distinction. More importantly, some cases can be best understood with this distinction in place. Hence, simplifying the debate in this way has theoretical costs. A better strategy for advancing philosophy of reasons might be to pursue deeper theoretical issues, as many do, such as assessing the overall merits of contextualist approaches.

More fundamentally, what the Monty Hall case shows is that our cognitive limitations inevitably lead to the disparity between normative reasons that derive from what we know (or, at any rate, what we have as evidence) and normative reasons that we are in a position to use in deliberation. Not all members of the former category belong to the latter. Sadly, even when reasons are properly connected to a specific person, the person may not always be in a position to take advantage of that connection.

---

[20] Note also that internalist accessibilism, as defended recently by Declan Smithies (cf. Smithies 2012), has to focus on possessed reasons to believe and not merely subjective ones (assuming that what justifies one to believe something are normative reasons to believe). Subjective reasons that are not possessed are not accessible and, hence, cannot determine justification according to internalist accessibilism. Therefore, it is unfortunate that Smithies fails to distinguish between possessed and subjective reasons (see the aforementioned quotation). This makes his view problematically ambiguous between a reading that cannot be true, given the very formulation of accessibilism, and a reading that is internally coherent but very demanding. Thanks to an anonymous referee for drawing my attention to the problem that the subjective–possessed distinction applied to reasons to believe might raise for Smithies's accessibilist account.

1.4   Reasons-First and the Wrong Kind of Reasons

According to a prominent, indeed arguably the majority view until recently, reasons cannot be analysed or defined in other terms; they are rather to be taken as fundamental with respect to other normative properties/notions. This view goes under the name of *reasons-first*, or *reasons fundamentalism*. There are thus two elements in this approach – reasons are taken to be prime, in the sense of not being analysable or substantially explainable in other terms. And reasons are explanatorily fundamental in the sense that we have to appeal to reasons in order to explain all other normative notions. Among prominent defences of the reasons-first programme are Scanlon (1998), Parfit (2011: especially 31–42), Schroeder (2007), and Skorupski (2010); a more recent defence is Rowland (2019); see also Engel (2015b) for an overview and exploration of the general aims and prospects of the reasons-first programme.

   An important thing to note about reasons-first views is that as presented above they need not be *all-things-considered* reasons-first. The general view is that within the normative domain, reasons are prime and fundamental. The general view, as presented here, is not committed *per se* to the claim that there cannot be a plausible reductionist story of reasons by appeal to non-normative properties. Of course, some (perhaps even most) reasons-first theorists would not endorse such a possibility. But some do accept it. Mark Schroeder is one prominent recent proponent of a reductive reasons-first approach in this sense (Schroeder 2007 is a book-length defence of such an option). According to Schroeder, all normative properties can be analysed in terms of reasons, but reasons can be reduced (in the specific, constitutive, non-symmetric sense) to non-normative properties, which on Schroeder's own preferred Humean account amounts partly to appeal to one's desires, promotion, and explanation.[21] As Schroeder puts it:

> But if attractive views about what is distinctive of the normative so often take this *structural* form, then the result that *good*, and *right* and *just* and *reason* and so on are truly normative properties, is one that it is actually incredibly *easy* for a reductive theorist to get right. Since being normative is a matter of a *structural* relation to some basic normative property like that of being a *reason*, the reductive theorist can accept this characterization of the

----

[21] More specifically on Schroeder's account:

> **Reason** For R to be a reason for X to do A is for there to be some p such that X has a desire whose object is p, and the truth of R is part of what explains why X's doing A promotes p. (Schroeder 2007: 59)

normative. Then, she can accept whatever analyses of each *non*-basic normative property in terms of the basic property are accepted by the non-reductive theorists who share this conception of what is distinctive of the normative. And finally, she gives her reductive theory as an analysis of the basic normative property or relation. So it turns out that even the normativity of normative properties is easy for a reductive theory to capture. (Schroeder 2007: 81)

More specifically Schroeder accepts the following fundamentality claim about reasons in the normative domain:

> **Reason Basicness** What it is to be normative, is to be analyzed in terms of reasons. (Schroeder 2007: 81)

In what follows, we will limit our focus only to the question of the possibility of intra-normative analysis/explanation of reasons. That is, we will leave out the discussion about the prospects of a viable reductive account of normative reasons in non-normative terms (though note that Schroeder's account shares some key aspects with 'Explanation' accounts of reasons, ahead, and as such might face some of the same objections that we will present to versions of the Explanation accounts – for example, value-based explanationist accounts). The focus here is on the question of whether reasons are basic/fundamental within the broadly normative domain or whether we can propose a viable view that explains what reasons are by partial appeal to other normative/evaluative properties.

As I see it, there are two main lines of thought that have been presented to support the reasons-first approach. The first line of thought is that there is simply no more informative account available of what normative reasons are. If we try to explain reasons, all we can get at best are circular accounts. Reasons just are, according to this line of defence, unanalysable, since no viable analysis or substantial, informative account is available. I take it that this line of thought is implicit in the now famous passage from Thomas Scanlon that has been taken to be the paradigmatic expression of the reasons-first approach:

> I will take the idea of a reason as primitive. Any attempt to explain what it is to be a reason for something seems to me to lead back to the same idea: a consideration that counts in favor of it. 'Counts in favor how?' one might ask. 'By providing a reason for it' seems to be the only answer. (Scanlon 1998: 17)

Also, this line of thought has more to it than mere appeal to linguistic or common-sense observations. For instance, some proponents of the reasons-first programme have explicitly argued against reductive

accounts/analysis of reasons. John Brunero, for example, has argued at length and on various occasions against some of the most promising existing reductive accounts of reasons (see, for example, Brunero 2009, 2013, 2018; see also Rowland 2019, in particular chapter 11). The idea here is that if it can be shown that all the existing, most promising, informative, reductive accounts of reasons are mistaken, then this alone gives some ground for taking the reasons-first approach seriously. However, as we will see ahead, even if Brunero and others are partially right in criticising the *existing* reductive accounts, there are still other options that have not yet been explored in the literature. The present proposal will actually amount to putting on the table one such overlooked reductive account that avoids the main objections to the existing reductive accounts.

The second line of thought that can be discerned in the reasons-first proponents' texts consists in an appeal to the fruitfulness of the reasons-first programme. In short, according to this line, given that taking reasons as prime and fundamental leads to theoretically useful results. For example, it helps to better understand such and such other thing, or to solve elegantly such and such previously unsolved problem, we should take the reasons-first view seriously. As I see it, there are two slightly different versions of this line of defence. The first appeals to the fruitfulness of theorising with reasons about other normative notions and statuses. The idea here is that given that by reducing all other normative notions to reasons, the reasons-first account provides a simple, theoretically unified, and hence explanatorily very powerful approach. So, for instance, on this view, what one ought to do is what there is most reason for one to do, what one is rational to do is what one's possessed reasons support overall (or, alternatively, what one's subjective reasons support overall), what is good is what anyone has reasons to prefer/desire/have a pro-attitude towards, and so on. Rowland provides a recent version of a proposal along these lines (limited to practical normativity only, however):

> On the account that I have provided we can analyse all moral, evaluative, deontic, and normative notions in terms of normative reasons. So, facts about moral, evaluative, deontic, and normative properties just consist in various sets of normative reasons. Every part of this unified picture of the practically normative has explanatory advantages. And the unified picture itself tells us what unifies the moral, the evaluative, the deontic, and the normative as all practically and normatively important domains. Namely, these domains are unified because facts about morality, the evaluative, the deontic, and the normative consist in facts about normative reasons. (Rowland 2019: 217)

If borne out, such a proposal would provide an explanation of the whole normative domain. It would have an extremely high degree of generality and explanatory power.

The second variant of the fruitfulness argument appeals to the role of reasons in sorting out prominent debates in normative philosophy. The idea here is that various standard opposing views in, say, moral philosophy or meta-ethics achieve better traction when transposed to the domain neutral topic of normative reasons in general. Here is one popular expression of this idea:

> Most meta-ethical debates have been about morality. But I shall first discuss non-moral practical reasons and reason-implying oughts. Our questions here take simpler and clearer forms. These are also the most important questions if, as I believe, normativity is best understood as involving reasons or apparent reasons. Things matter only if we have reasons to care about them. In the conflict between these various theories, reasons provide the decisive battlefield. If Naturalism and Non-Cognitivism fail as accounts of reasons, these theories will also fail, I believe, when applied to morality. (Parfit 2011: 269)

If this is on the right track, then appeal to normative reasons can be fruitful not only in sorting out what other broadly normative/evaluative notions are but also in helping to overcome some of the most persistent disagreements in meta-ethics and, presumably, other normative fields. This would, of course, constitute another consideration that speaks strongly in favour of the reasons-first approach.

Despite its very enticing promises, however, the reasons-first approach has to be set aside in what follows. For one thing, as we've noted already, all the options for developing a viable reductive view of reasons have not been explored yet. Given the methodologically plausible constraint that we should not postulate (normative) entities beyond what is necessary, any view that doesn't introduce an independent entity of normative reasons but manages to reduce/explain reasons by appeal to already known normative notions/properties has the advantage of theoretical simplicity.

Another reason for putting the reasons-first approach on hold is that, as several philosophers have recently observed, its promise of providing an effective reduction of all the other broadly normative/evaluative properties/notions doesn't bear out. Without this element, however, the reasons-first approach loses any bite. If other normative properties/notions cannot be reduced to reasons, then reasons are not fundamental in the sense of being explanatorily indispensable and the ultimate element within the normative domain. Recent discussion about this has focused in particular on the

reasons-first promise to reduce values to reasons. This debate is often referred to as concerning the buck-passing account of values (the idea being that the buck stops at reasons in explaining values; sometimes the term 'buck-passing accounts' is used more broadly to refer to attempts to reduce any normative/evaluative property to reasons; cf. Löschke 2021). Scanlon is a standard source of contemporary reasons-first buck-passing accounts of value:

> Chapter 1 explained and defended my decision to treat the notion of a reason as primitive. In this chapter, I will use the notion of a reason, taken as the most basic and abstract element of normative thought, to provide a general characterization of a slightly more specific normative notion, the idea of value. (Scanlon 1998: 78)

One general way of putting the proposal, a way that abstracts over specifics of concrete proposals from the reasons-first theorists, has been recently helpfully summed up (but not defended!) by McHugh and Way:

> What it is for X to be good is for there to be sufficient reason for anyone to value X. (McHugh and Way 2016: 577)

To value here is understood as having a relevant pro-attitude – for example, desiring, wishing. And good here is understood in the sense of goodness *simpliciter* (not in the sense of good-for, nor in the sense of attributive goodness) (cf. McHugh and Way 2016). Note that this general buck-passing proposal is easily generalised to goodness-for, attributive goodness, as well as to further more specific values – for example, one is admirable when there is sufficient reason to admire one. And indeed it is supposed to be generalised in this way. Moreover, this general aspect of the view might be taken to be a further advantage of the view (see McHugh and Way 2016: 578–579, who make precisely this observation).

The problem with the reasons-centred buck-passing account of value is that to think that values can be reduced to reasons seems to be a mistake given our pre-theoretical judgments about some fairly simple cases. The main objection here appeals to the so-called wrong kind of reasons problem. Formulations of this problem that are already classical appear in Crisp (2000), D'Arms and Jacobson (2000a, 2000b), and Rabinowicz and Rønnow-Rasmussen (2004); see Gertken and Kiesewetter (2017) and Engel (2020a) for recent overviews of the debate. A well-known example that has been used to state the problem (from Rabinowicz and Rønnow-Rasmussen 2004) is, roughly, that of a despicable demon that threatens to punish one severely, say to kill one, unless one admires the demon (and admires the demon for what it is). It is worth noting that, of course, the

strength of the argument would be just the same if transposed to an ordinary situation that preserves the same structure. Two natural observations about such a possible situation when taken together clashes with the reasons-centred buck-passing accounts. The first is that one has sufficient reason to admire the demon in this situation. Indeed, if the threat of terrible pain doesn't give one sufficient reasons, what else could give it? And second, clearly the demon is not admirable; it is despicable. On a buck-passing account, however, if we hold to the view that one has sufficient reason to admire the demon, we have to conclude that the demon is admirable. Thus, something has gone wrong, for we know that the demon is not admirable. If we take these considerations at face value, then it seems that we have to conclude that values cannot be analysed in terms of reasons and hence reasons are not the first and most fundamental elements of the normative domain.

Now, one popular line of response to the wrong kind of reasons problem on the part of reasons-first proponents has been to distinguish between *state-given* reasons and *object-given* reasons (cf. Parfit 2011: appendix A; Piller 2006). Roughly, the idea here is that some reasons arise from or are given by some features, including benefits, of being in a given state, such as from being in a state of admiration of someone; while other reasons are given by the objects of attitudes – for example, by someone's magnanimity in the case of admiration. Crucially, state-given reasons can only be reasons to want to be in a state, or for undertaking actions to try to make oneself have/get into the relevant state (cf. Parfit 2011: 432). They are not reasons to F or reasons to have the relevant attitudes/states (e.g. admiring someone). On the other hand, object-given reasons are reasons to F. Applying this distinction to the case of the despicable demon, the proposal is that the demon's threat to kill one unless one admires the demon can only be a state-given reason. It cannot be an object-given reason, since it is not connected to the demon being magnanimous or generous or otherwise admirable, for the demon is not magnanimous or generous or admirable. The threat can only be a state-given reason in this case. Thus, the threat can only be a reason for one to want to admire the demon or to attempt to (try to) make oneself get into the state of admiration of the demon. If so, then the project of analysing values in terms of reasons might still be maintained. Having state-given reasons in the demon's case – that is, reasons to want or to try to admire the demon – even if sufficient, don't entail that the demon is admirable, since these are not reasons to admire. Admirability reduces only to sufficient reasons to admire, not to reasons to want or to try to get into the state of admiration.

Naturally, the story extends to other values as well: on this picture, roughly, X is good just in case there are sufficient reasons for one to value X, which are different from reasons to want to value X or to try to get into the state of valuing X (see the quotation from McHugh and Way 2016: 577 above).

I think there are two key worries with this line of reply to the wrong kind of reasons problem for reasons-firsters.[22] First, such a proposal looks somewhat ad hoc and arbitrary, given the fact that reasons-first proponents are in no position to provide a substantial account or a definition of reasons. To have a theoretically well-motivated distinction between two kinds of the same thing, we need to know well what is the common element that unifies the two kinds. But the only characterisation that we have from reasons-firsters is that reasons are considerations that speak in favour of a response. The threat in the demon example certainly seems like something that speaks in favour of admiring the demon. Do reasons-firsters have sufficient theoretical grounds that are independent from the need to respond to the wrong kind of reasons problem, for maintaining that normative reasons are of two kinds and that only sufficient object-given reasons to value X really entail that X is valuable? It is not clear how proponents of the reasons-first view could provide us with a satisfactory response to this question. If, say, correctness considerations or justification considerations are worked into the account of object-given reasons by suggesting, perhaps, that only object-given reasons can render an attitude correct/justified (and assuming in addition, for example, that one being admirable and it being fitting/justified to admire one are connected, perhaps, by a biconditional), then reasons would not be first after all. For one would then also appeal to correctness/justification in explaining what object-given reasons are and thus the overall explanation of values would not rely exclusively on normative reasons to value.

Pamela Hieronymi seems to propose a similar, if not the same, line of objection; for instance, she writes: 'As long as a reason is simply a consideration that counts in favor of an attitude, we are left without an obvious way either to draw a useful distinction between these very different sorts of reasons or to say why one of them seems to be the "real" sort of reasons' (Hieronymi 2005: 443). Hieronymi also argues convincingly that it is not clear at all that all state-given reasons are reasons to want or to try to be in a state F, rather than reasons to F, as, for instance, in the case of

---

[22] Thanks to an anonymous reader for Cambridge University Press for making me realize the need to develop these further worries in detail.

imagining that there is no heaven (cf. Hieronymi 2005: 442). Thus, by itself the state-given/object-given reasons distinction cannot help with the problem of the wrong kind of reasons. Reasons-firsters have to appeal to a more substantial and independent account of reasons to distinguish in a non ad hoc way between reasons to which values can be reduced and reasons to which values cannot be reduced. But this is precisely what proponents of the reasons-first approach cannot do, for, according to them, there is nothing more substantial we can say about reasons than that they are considerations that speak in favour of some response. In short, the first line of objection to the object-given/state-given distinction based response to the wrong kind of reasons problem for reasons-firsters is that such a move creates a dilemma for reasons-firsters. On the first horn of the dilemma, they need to provide an independent, theoretically motivated account for why there is a significant distinction between object-given and state-given reasons. Such an account would involve telling us something above and beyond the mere characterisation of reasons as considerations that count in favour. And thus, it would undermine the reasons-first project, according to which we have a robust grasp of reasons and we cannot say anything more substantial about them than that they are considerations that speak in favour. On the second horn, refusing to provide a more substantial, theoretically motivated, and independent account of the distinction between object-given and state-given reasons while relying crucially on this distinction in the treatment of the wrong kind of reasons problem is merely ad hoc and arbitrary and should not be accepted in the present context of debate.

Second, and somewhat connectedly, is that at the end of the day this line of response amounts to an attempt to dissolve rather than solve the wrong kind of reasons problem (cf. Engel 2020a). However, it is not clear that the proponents of the reasons-first approach are in a position to justify such an attempt. The initial assumption that triggered the wrong kind of reasons problem was that, for example, threats appear to constitute genuine and arguably sufficient reasons to admire the despicable demon. After all, it is specifically admiring the demon, and not attempting or wanting to admire the demon, that could save one's life in the imagined case. Insisting that only object-given reasons are reasons to, for example, admire the demon would amount to a scepticism about the wrong kind of reasons; that is, to the denial of the very idea that the wrong kind of reasons are genuine normative reasons and bear normative force. Now, of course, scepticism about the wrong kind of reasons is a possible move within the overall debate about these cases. And philosophers have been willing to

endorse this option while trying to explain why we have the intuition that the wrong kind of reasons considerations speak in favour of a response F without equating these considerations with reasons to F (cf. Way 2012, see also Skorupski 2010, see also Gertken and Kiesewetter 2017 for further references and discussion). However, the worry in our present context is that scepticism about the wrong kind of reasons is not a viable option for reasons-first proponents (cf. McHugh and Way 2016). The worry is that, contrary to reductive accounts of reasons, which propose to explain reasons in other terms and to explain our intuitions away in the wrong kind of reasons cases, reasons-firsters are ill-placed to provide an error theory of why we are, allegedly, massively mistaken in thinking that threats and similar considerations can speak in favour of admiring and other attitudes. Suggesting that we are all massively mistaken in thinking that threats speak in favour of admiring becomes problematic if one's view entails that we are supposed to have a robust grasp of reasons and of speaking in favour. Reasons-firsters presuppose that we have a robust, intuitive grasp of reasons and of speaking in favour of F-ing. But if they endorse scepticism about the wrong kind of reasons, then they also have to admit that our grasp of reasons is not so robust after all, since we are massively mistaken about an important category of considerations that seem to be reasons and speak in favour of F-ings. There is then an unresolved tension in such a proposal: if reasons are first, then we both have and don't have a robust grasp of what normative reasons are. And this sounds very much like a paradoxical conclusion.

Moreover, even if we bracket the issue with the robust grasp of reasons, it is not clear that reasons-firsters could provide an independent theoretical motivation for taking the scepticism route. Again, they cannot say anything more than that reasons are considerations that speak in favour of F-ing. We don't get a theoretical and independently motivated answer from reasons-firsters of why we are massively mistaken in thinking that some considerations appear to speak in favour but are, contrary to appearances, not reasons at all to F. But without a plausible error theory, scepticism about the wrong kind of reasons is unsatisfactory. It would seem, then, that combining the reasons-first approach with scepticism about the wrong kind of reasons is unpromising. The wrong kind of reasons problem is still an important concern for the reasons-first approach.

Note also that appealing to the idea that somehow the wrong kind of reasons problem is an instance of a more general and pervasive issue that is a problem for everyone, not only for reasons-firsters (cf. Schroeder 2012, 2013), doesn't seem to help either. As Kieswetter and Gertken have

persuasively argued, there is not necessarily a problem for alternative accounts (see also Hieronymi 2005, 2013 among others). A fittingness-first approach, for instance, as defended by McHugh and Way (2016), doesn't seem to be vulnerable to such a problem, since they don't endorse the claim that correctness is explained in terms of reasons, but can explain an X being valuable directly as it being fitting to value X. It is not clear then that there is a more general wrong kind of reasons problem that everyone has to face. It would seem to be a genuine problem first and foremost, if not only for those who think that reasons come first in the order of explanation in the normative domain.

Thus, I conclude the wrong kind of reasons problem is still a major problem specifically for reasons-first views, since notable attempts to solve it seem to be unsuccessful. I don't aim here to suggest that the issue has been closed and that the reasons-first approach has been ultimately shown to be mistaken. Indeed, we have not even examined all the possible proposals to deal with the wrong kind of reasons problem. The debate is still on-going. However, the suggestion that I would like to make here is that in the light of the present situation of the debate, it is still worthwhile to explore alternative, non-reasons-first accounts of normative reasons. Again, finding a successful reductive account of reasons would be enough by itself to put the reasons-first approach on hold. We now turn to exploring the prospects for this task.