Fittingness and Other Normative Categories

Doctor of Philosophy

Ronan Ó Maonaile
Department of Philosophy
September, 2025

Abstract

This thesis is about the relations between different normative categories. Normative categories are things like reasons, oughts, values, and fittingness. The guiding question is this: can we explain the nature of one normative category in terms of another? Specifically, I am interested in explaining the nature of fittingness, a normative category which has received increased attention in recent years. The thesis develops several related claims: (1) fittingness is irreducible to other normative categories; we cannot explain the nature of fit in terms of value or the balance of normative reasons, (2) there is no one normative category which fully explains all the others; all X-First views are seriously flawed, and (3) the project of explaining the nature of normative categories and the relations between them is best advanced by articulating the particular features that unite and divide them.

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Ronan Ó Maonaile

Acknowledgements

There are a number of people without whom this piece of writing would not exist. I owe a debt of gratitude to each of them. First and foremost, I want to thank my parents, Tomás and Siobhán Uí Mhaonaile. I cannot begin to name all of the ways they have supported me. I want to thank my supervisors, Luke Elson and Charlotte Newey. Luke has been a mentor to me from the very first day I started at Reading. He has helped me to think critically about my writing and steered me through moments of self-doubt. I am grateful to Charlotte for offering her support when the need arose. I also want to thank Philip Stratton-Lake. Philip and his work drew me to Reading, and in the short time I had the privilege of studying under his guidance, I learned a few things about ethics, and a great deal about philosophical argument. I sincerely thank the faculty of the philosophy department at the University of Reading, all of the graduate students who participated in weekly research seminars, and numerous philosophers and friends who have provided invaluable feedback on my work over the past four years. For written comments and stimulating discussions that have directly influenced the development of the ideas presented here I want to thank Brad Hooker, David Oderberg, Selim Berker, Alex Gregory, James Hart, Steven Deon Wu, Joshua Jarvis-Campbell, Luke Elson, Philip Stratton-Lake, and Charlotte Newey. Finally, I want to acknowledge the anonymous reviewers for various academic journals whose objections to different versions of the chapters that make up this thesis greatly contributed to their present form.

Table of Contents

Chapter 1. Introduction

1.1. The Project

In normative and practical ethics, we make claims about what we ought and ought not to do, which sorts of outcomes would be good, and which bad. We try to figure out what we have reason to do and what we have reason to refrain from doing, and we determine who is blameworthy for what, and when and why someone is morally admirable. We do similar kinds of things in relation to non-ethical matters. For instance, epistemically, we want to know what we have reason to believe, what we ought and ought not believe, and what it would be fitting to believe. Aesthetically, we are interested in the value of artworks and other aesthetic experiences, when aesthetic pleasure, amusement, or the attitudes associated with tragedy are appropriate responses, and what, if any, are our aesthetic obligations. And of course prudentially, the same thing again: do I have reason to pursue postgraduate studies? Should I do it? Would it be a good thing?

In metaethics, or metanormative philosophy, we do not primarily try to determine which things are good or what we have most reason to do or believe, although some metanormative theories may have practical implications. Instead, we ask questions about these questions, like what exactly is goodness? And how do we explain what it is to have a reason for something? Here we are interested in the 'fundamental nature' of goodness or reasons. There are several different approaches to these kinds of meta-questions. One involves the metaphysical status of normative categories[1]. There are various ways to go here. You might think that goodness is a natural property just like roundness or being a table. Alternatively, perhaps it is non-natural: a property fully existing in the world but not observable via the empirical methods of natural science. You might even deny that goodness really exists: either our normative language attempts to describe the world but fails systematically because there are no properties picked out by terms like 'good', or normative language is not in the business of describing the world at all; it merely expresses certain pro or con attitudes. The present thesis is not concerned with which of these options

[1] Throughout this thesis I use 'normative categories' as a collective term for normative things in the world, such as properties like goodness and relations like reasons and fittingness. As I go on to say in the body of the text, whether or not these normative things actually exist will not be my concern. For those who are not comfortable with the realist connotations of talk about normative categories, what I say can be taken to be about normative *concepts.* Indeed, at certain points in this thesis, especially chapter 6, I will explicitly address myself to normative concepts.

is correct. I take it for granted that what I say about normative categories is compatible with both realism and anti-realism about their metaphysical status.

A different sort of second-order question concerns whether the various normative domains – such as the moral, epistemological, aesthetic, and prudential – are unified. When we talk about values and obligations in the moral domain, are we referring to the same things as when we talk about epistemic or aesthetic values and obligations? Is there a moral kind of fittingness which is distinct from the fit between belief and true propositions, or between amusement and funny things? This thesis is not concerned with these sorts of questions either. I take it for granted, unless otherwise stated, that normative categories and the terms that pick them out are univocal across normative domains.

There is another approach to the question of the nature of normative categories, and this is the one we will take. We can ask, quite straightforwardly, 'what is goodness?', 'what is a requirement?', and 'what is fittingness?'. These are questions about the *constitutive nature* of normative categories. Here we are looking for definitions, analyses or reductive explanations. This thesis is a collection of essays addressing this sort of explanatory question. The kind of explanations I will be interested in are those which purport to explain one normative category in terms of another. A primary motivation behind such explanations is the striking coinstantiation of different normative categories. For instance, it would seem that whenever x is good, there is also some reason to admire x or to approve of it. This has led some people to think that the nature of goodness can be explained in terms of reasons for pro-attitudes, as in Scanlon's (1998) 'buck-passing account' of value, which says that for x to be good is for there to be sufficient reason to respond to x with a pro-attitude. Or perhaps the exact opposite is true, such that reasons are to be explained in terms of value, as proposed by Barry Maguire (2016). This thesis examines the prospects of explaining the nature of one normative category in terms of another, with a special focus on the normative category of fittingness.

Fittingness is generally understood to be a normative relation in which a response – typically an attitude – stands to its object when the object merits or is worthy of that response (Howard, 2019, p. 216). For example, it is fitting to blame a person when they have done something in virtue of which they merit or are worthy of blame, i.e. when they are blameworthy. When we ask for a bit more detail about the nature of the fittingness relation, one might try to explain it in terms of other normative categories, especially categories we are more familiar with. For example, it may be suggested that an attitude is fitting in response to x just in case there is sufficient *reason* to have that attitude toward x, or fittingness might be a kind of *requirement* or *permission*, or maybe the fittingness of a response is to be explained in terms of the *value* of having that response. Each of these views attempts to explain the fittingness of responses in terms of a different normative category. Some writers go the other way and argue that fittingness is unanalysable, and it is in fact this category which explains the nature of the other ones (McHugh and Way, 2016 and 2022). This brings a very strong claim about the nature of normative categories into the picture: perhaps there is one normative category, where this might be reasons, value, ought, or fittingness, which is *First* in the sense that it is prior to all the others; a master category which fully explains each of the categories we are interested in. Fittingness-First is the view that we can explain what it is to be a reason, a requirement, and a good, all in terms of fit. Call views of this sort X-First views. As we will see in the next chapter, Fittingness-First, Reasons-First, Value-First, and Ought-First have all been defended in the literature.

I am sceptical of the whole enterprise of explaining the nature of normative categories in terms of one another. This thesis attempts to develop and defend this scepticism by (1) arguing against X-First views in general; (2) showing how a particular normative category, fittingness, is irreducible to any other category; and (3) demonstrating along the way an alternative means of explaining the nature of normative categories, one which focuses on articulating the features that unite and divide them. In the next two sections of this introduction, I first say something about my motivations for writing about these issues in the first place and why I am drawn to the answers I give. I then provide an overview of the chapters to come, including the argumentative methods they employ.

1.2. Motivation

Every philosophical programme is vulnerable to a peculiar kind of scepticism not normally directed at the physical scientist, the economist, or the software engineer to the same extent. It takes the form of a raised eyebrow and a condescending question: what's the point of this? Moral philosophy is no exception, and metanormative questions about the constitutive nature of normative categories are sure to meet with more than a little suspicion of this kind. I am going to mention two very general motivations for the present project: one which speaks to the philosophical importance of questions about normative categories, and one which addresses the implications of my particular approach in this thesis, an approach I think is appealing for reasons beyond the purely philosophical. In his introduction to the recently published *The Future of Normativity* (2025), which he also edited, Simon Kirchin discusses what we might go as far as to call the 'normativisation' of philosophy:

> "It seems that nearly every field in modern analytic philosophy is riddled with normativity: logic and philosophy of language, philosophy of mind, metaphysics and epistemology, ethics, politics, law, aesthetics, and so on, with figures from the history of philosophy also viewed through the normative lens. Modern writers are writing not just about the normative aspects of art or mind or logic; a whole body of work has developed that focuses on the idea of normativity itself… It seems that one nowadays simply cannot begin to understand or comprehend many areas or problems that philosophers are concerned with without understanding these areas' normative natures" (Kirchin, 2025, p. 1).

Assuming Kirchin is right, and the tendrils of normativity have indeed spread into most areas of philosophy, why we should be interested in explaining the nature of normative categories is plain to see. If normativity pervades such a diverse range of human thought and experience, and normativity essentially involves normative categories such as values, reasons, and fittingness, then determining exactly what these normative categories consist in is a vital project. Fittingness is a good example of a normative category that has found its way into various streams of philosophical discourse. Some cases which will be alluded to at various points in this thesis include aesthetics, where the fittingness of attitudes elicited by artworks may affect the work's aesthetic value; the philosophy of emotion, where the plausibility of the thought that emotions have fittingness conditions may act as a constraint on theories of emotion; and the philosophy of science, where a research project may be deemed 'pursuitworthy' (Fleisher, 2022), or fitting to pursue, even though the current state of evidence counts against it. Answering the question 'what is fittingness?' should therefore be of interest to those working across a broad spectrum of philosophical projects. A central goal of this thesis is to contribute to the growing body of literature on the nature of fit.

As noted in the previous section, the approach to the explanation of normative categories I am going to interrogate involves explaining the nature of one category in terms of another. I am going to argue that fittingness cannot be analysed in this way, and in so arguing I employ an alternative means of elucidating normative categories, one which focuses on describing their constitutive features. That is, the sort of explanation of the relations between normative categories I will demonstrate differs from the reductive method insofar as I will say what fittingness is and how it relates to normative reasons, for instance, by highlighting features that unite and divide them. In the case of the relation between reasons and fit, chapter 4 carries out this method by arguing that fittingness is determined by weighted considerations, normative reasons are weighted considerations, and yet fittingness is not determined by normative reasons because the weighing of fit-making considerations only issues in two distinct normatively significant verdicts – fitting or unfitting – while the weighing of reasons issues in three: impermissible, permissible, or required. This way of explaining the nature of normative categories is in contrast to the reductive approach, which might claim, for instance, that fittingness is to be explained in terms of reasons on the basis of more abstract observations, like the fact that there is a correlation between reason and fit that calls for explanation, or by appealing to considerations of parsimony. How this method works in each of my chapters will be spelled out in the next section.

The anti-X-First and fittingness fundamentalist agenda I am advocating also has implications for how we think about normativity in a wider, more general sense. What my view entails is that there are at least three distinct normative categories: fittingness, value, and reasons.[2] This kind of pluralism about the normative domain is very appealing insofar as it opens up space for ordinary intuitions. Consider the following, somewhat simplified example. Suppose that the atomic bombings of Hiroshima and Nagasaki by the United States of America in 1945 were overall the best thing to do. It was a good thing to end the

[2] Although I do not devote much time to arguing that deontic categories such as oughts, requirements and permissions can be explained in terms of reasons, I do think this is a very plausible view, and I will make use of it in the coming chapters.

war, and although the murder of innocent civilians was beyond terrible, the greater loss of life had the war continued would have been much worse. Even if I accept all of these evaluative claims, I might still think that dropping the bomb was *wrong*. It would be difficult to explain this thought if what is right can be straightforwardly explained in terms of what is good, even if various more sophisticated versions of consequentialism attempt to provide such an explanation. This gives us some initial reason to be pluralists about the right and the good.

Now let's suppose that it's true that dropping the bomb was good overall, and it's also true that it was the right thing to do; perhaps the different kinds of reasons in favour of doing so outweighed the reasons against. Even if I accept each of these claims, I might still think that something is *off* about the treatment of Hiroshima and Nagasaki. It was both right and good, insofar as it ended the war and saved more innocent lives than would otherwise have been lost, and yet it may still be unfitting to approve of the bombing, or perhaps the act of killing civilians in war is itself unfitting, even if it is both right and good. This sort of thought would be difficult to sustain if fittingness were straightforwardly explainable in terms of the right or the good. The anti-reductivist position I am advocating allows us to capture the full range of normative experience and judgement by equipping us with a number of distinct, independent normative categories which more reductivist views risk assimilating.

1.3. Overview of Chapters and Methodology

Chapter 2 sets the anti-reductivist tone of the thesis by advancing a pluralist position within the X-First debate. The chapter argues that there is no X First. I begin by saying exactly what X-First views involve. This is done by distinguishing the central claim from other claims it is easily confused with. I also describe the precise nature of the constitutive explanation of normative categories X-First views are supposed to provide. In doing so, I highlight what defenders of X-First views are committed to, which lays the foundations for my argument against all X-First views. The two most important commitments are these: (1) if a normative category X is to provide a constitutive explanation of another normative category Y, then X must be able to fully explain all the constitutive elements of Y, and (2) this means that X must be able to account for the phenomenology of Y.

Before discussing a serious flaw in all X-First views, the chapter considers two popular motivations that have been cited in their favour - demystification and qualitative parsimony - and argues that neither gives us decisive reason to adopt an X-First view over No-X-First. I then give my argument against all X-First views. This is the argument from phenomenology, which says that there is no one normative category X which can account for the diverse phenomenological features of all other normative categories. This is a representative example of the main argumentative strategy employed throughout this thesis: reflecting on our normative concepts, we often find that there is some feature of a normative category Y – where this category might be goodness, or fittingness, or reason – which is not shared by another normative category X. If this feature is partly constitutive of what it is to be Y, and X can tell us nothing about this feature, then we have good reason to doubt that X can provide an adequate constitutive explanation of Y.

We can illustrate this with a non-normative example: what explains the occurrence of earthquakes? Suppose I suggest that it's the wind; earthquakes are what happen when strong gusts of wind shake the earth. Among the many reasons for rejecting this explanation is the fact that there is a feature of earthquakes which the wind lacks. Earthquakes release massive amounts of energy. Far more energy than even the strongest winds can produce. Unable to account for their power, any explanation of earthquakes in terms of the wind is bound to fail. In the coming chapters, I will make an analogous move in order to reject explanations of a normative category Y in terms of another normative category X. Furthermore, simply describing some of the features of different normative categories goes a long way toward providing constitutive explanations of them. This gives us an alternative to the reductive strategy of explanation, one in which normative categories are seen as part of an interrelated web based on the features that unite and divide them. In the rest of this section, I will call this sort of argument the 'normative features strategy'.

Chapter 3 brings the normative category of fittingness into focus and illustrates some of its basic features by asking what sorts of responses can be fitting or unfitting. Fittingness is generally understood as a relation between an attitude and its object: admiration is fitting whenever the object it is directed at is admirable, and belief is fitting in response to true propositions. There are two other sorts of responses which may be evaluable for fit. Bodily episodes are the first. Oded Na'aman (2022) has recently argued that the bodily episodes which often accompany affective attitudes, such as the increased heart rate and sweaty palms that form a part of your fear, are also evaluable for fit. Action is the second. In the early to mid-20th century discussions of fit by people like C.D. Broad (1930) and W.D. Ross (1939), actions played a central role; it was thought that a particular action could be more or less fitting in response to an aspect of one's situation, or the situation as a whole. Actions also make an appearance in some contemporary discussions too, including Philip Stratton-Lake (2022) and Selim Berker (2022). The chapter asks whether bodily episodes and actions are in fact fit-evaluable.

Following Na'aman's lead, I lay out a list of eight conditions that a response plausibly must meet in order to be evaluable for fit. It is assumed, again following Na'aman, that fit-evaluable responses are also rationally evaluable, and the first five conditions discussed are widely accepted features of rationally evaluable items, such as (1) rationally evaluable responses are responses to which we normally apply norms of rationality and justification and (2) rationally evaluable responses are responses for which we normally seek and offer reasons-for-which explanations. To these I add three more conditions specific to fit-evaluability: (3) If x is fit-evaluable, x has an object, (4) If x is fit-evaluable, it must be possible for x's object to merit or be worthy of x, and (5) If x is fit-evaluable, x must have constitutive standards of correctness. Bodily episodes and actions are then measured against these criteria, and it is argued that while the former are not fit-evaluable, the latter are. This exemplifies the normative features strategy insofar as we have answered the question 'what is evaluable for fit?' by attending to certain features that characterise fit-evaluable entities.

Chapter 4 addresses a potential reductive explanation of fittingness in detail, namely the view that for a response to be fitting is just for it to be sufficiently supported by the right kind of normative reasons. This chapter, along with chapters 1 and 5, contains most of my reasons for the claim that fittingness is an independent normative category, irreducible to any other. Here I offer a new argument for a claim I label 'The Independence of Fit', which says that fit-making considerations are not normative reasons. Instead, the aspects of objects which determine whether or not a certain response is fitting should be thought of as providing a distinctive kind of normative support. I begin by canvassing other arguments for this claim that can be found in the literature. These are the wrong kind of reason problem, arguments from guidance, and arguments from the weight of reasons. According to arguments from the wrong kind of reason problem, fittingness cannot be explained in terms of normative reasons for attitudes because there can be very good reason to admire x even though admiring x is not fitting, as when a demon who is quite the opposite of admirable threatens to kill you unless you admire him. Solving the wrong kind of reason problem involves articulating the nature of 'right-kind' reasons in a non-arbitrary way. Although there is little consensus about how to do this, it is reasonable to assume that the right kind of reasons can be distinguished from the wrong kind, and so a defender of the view that fit is to be explained in terms of normative reasons can simply adjust the position to one about fit as sufficient right-kind reasons.

The next two types of arguments try to show that even right-kind reasons do not determine fit. Arguments from guidance say that we can only have normative reasons for responses we can reason *to*. However, we cannot reason to emotions, and since emotions can be fitting, fittingness cannot be determined by normative reasons. Arguments from the weight of reasons say that normative reasons can be weighed against each other in determining overall deontic facts. Reasons are gradable, and they combine and compete. Fit-making considerations are not gradable, and they neither combine nor compete. Contra Conor Schultz (2025) and Barry Maguire (2018) respectively, I argue that neither of these arguments is fully convincing.

I then offer my own argument in support of The Independence of Fit. Employing the normative features strategy once again, I argue that fit-making considerations and normative reasons are different in an important way: while the weighing of reasons issues in three distinct, normatively significant verdicts - impermissible, permissible, and required - if fit-making considerations can indeed be weighed, the weighing process only issues in two distinct, normatively significant verdicts - fitting or unfitting. Fit-making features thus lack a feature essential to reasons, namely a tripartite structure of normative verdicts.

Chapter 5 considers the relation between fit and two other normative categories, ought and good. This chapter pursues the normative features strategy by taking a popular view about the all-things-considered ought and asking whether the feature it ascribes to the deontic category is shared by fittingness and value. This feature is an epistemic constraint. As a view about ought, *epistemic perspectivism* says that an agent A ought to Φ if and only if A ought to Φ relative to A's epistemic perspective, where an agent's epistemic can include her beliefs, evidence, or knowledge about her situation. I do not attempt to fully defend epistemic perspectivism about ought against its alternative, *objectivism*, the view that an agent's epistemic position is irrelevant to what she ought to do.

Instead, I show how the very same arguments which support epistemic perspectivism about ought also support epistemic perspectivism about fit. So, if you think that an agent's epistemic position partly determines what she ought to do, then you should also think that her epistemic position partly determines the attitude it is fitting to have. I also show that these arguments do not support epistemic perspectivism about good simpliciter to the same extent. There are two upshots. First, epistemic perspectivism about fittingness is made to look a lot more plausible than is generally recognised in the literature, where objectivism about fit is the dominant view (McHugh and Way, 2022, pp. 78-9). And second, since the evaluative category of good simpliciter is not epistemically constrained, while fittingness plausibly is, we have reason to doubt that constitutive explanations of fittingness in terms of goodness will be successful.

Chapter 6 is a further exploration of one of the themes of chapter 5, namely the view that fittingness is agent-relative, or at least less objective than it might appear. According to the widely accepted fittingness biconditionals, x is V if and only if a corresponding attitude A is fighting in response to x. For instance, x is admirable if and only if admiration is fitting in response to x, or x is shameful if and only if it is fitting to be ashamed of x. Chapter 5 argued that whether admiration or shame are fitting in response to x plausibly depends on a particular agent's epistemic position with respect to x, or at least we should consider this a plausible view if we think what one ought to do depends in part on her epistemic position.

Epistemic perspectivism about fittingness involves the rejection of objectivism, which is the view that whether some attitude is fitting depends solely on the objective features of the attitude's object, irrespective of any agent's epistemic perspective on it. If being shameful and being admirable are purely objective properties, then the truth of biconditionals like 'x is admirable if and only if it is fitting to admire x' would suggest an objectivist view about 'fitting to admire', rather than an epistemic perspectivist one. This chapter aims to show that properties like being shameful and being admirable are less than fully objective by arguing that the concepts we use to represent them, which I call Affective Concepts, or ACs, are essentially contestable, which is to say they admit of a number of alternative conceptions, none of which is determinately the best or correct conception. Drawing on the work of W.B. Gallie and his successors, I lay out some conditions on essential contestability and show how ACs meet these conditions. The upshot is that ACs do not refer to stable, objective properties, but are open and multivocal. Here the normative features strategy has been employed to articulate a feature of affective concepts - they are essentially contestable - in the service of providing further support to the subjectivist conclusions of the previous chapter.

Chapter 7 seeks to drive home the point that we should not think of the relations between different normative categories as a kind of hierarchical ordering, with one category X in a position to explain what is to be Y, but rather as a web of independent categories which are nonetheless interrelated. The chapter makes this point by focusing on the relations between the deontic and evaluative categories. Here the normative features strategy is flipped on its head. Responding to Selim Berker's (2022), I show that the deontic family of categories, which includes things like requirements and permissions, shares certain logical and relational features with the evaluative family of categories, which includes things like goodness and the 'better than' relation. This is to contradict the traditional view, which Berker defends, that says the deontic and evaluative families are entirely distinct, independent sets of normative categories.

Following Berker, I discuss five potential features of the deontic and evaluative categories. These are duality, opposition, gradability, neutrality, and alternatives dependence. I argue in a piecemeal way, showing how each of these features can be found among both the deontic and evaluative categories. Crucially, however, I do not think this lends support to views which attempt to explain ought in terms of the good, vice versa. Instead, it should be taken as evidence that the two families are interrelated. The deontic and evaluative categories share certain logical and relational features. In revealing these similarities, I also highlight some differences that remain. The final part of the chapter proposes a way of understanding these similarities and differences: normative categories are related via the roles they play in reasoning. **Chapter 8** provides a very brief conclusion to the thesis.

Chapter 2. No-X-First: A Plea for Normative Pluralism

2.1. Introduction

In this chapter, I offer a general argument against all X-First views of normative categories. There is no normative category X which can fully explain all other normative categories because there is no X which can account for the diverse phenomenological profiles of goodness, oughts, reasons, and fittingness.

We refer to normative categories all the time. Normative categories are the things in the world picked out by our normative terms, including properties and relations. For example, we talk about what is good and what is bad, as in 'You donated your kidney to save a relative? That was a very good thing to do', and 'Most Adam Sandler comedies are pretty bad'. 'Good' and 'bad' have comparative and superlative forms as well; 'better'/'worse' and 'the best'/'the worst'. These are the evaluative categories.

We also talk about obligations, permissions, and oughts, as in 'You ought to keep your promise', or 'It is permissible to break a promise in order to prevent a greater harm'. These are explicitly action-guiding normative categories; they tell us what to do. We refer to these categories when faced with dilemmas and those ubiquitous moments in life when a choice needs to be made: I have promised to pick my friend up at the airport. When the time comes to leave, I am playing a video game, in the middle of a really fun and important mission. I would rather continue playing, but I feel as though I *ought to* keep my promise, and so I get in the car and drive to the airport. However, driving along the motorway, I come across a traffic collision and a person screaming for help. There is no one else around. I stop to offer my assistance, knowing that I am breaking my promise to my friend but also realising that it is *permissible* to do so, in light of the circumstances. Relatedly, we talk about what we have reason to do. Normative reasons *count in favour of* responses (Scanlon, 1998, p. 18). Although this is not universally accepted as an analysis, it is a standard way of elucidating normative reasons, and throughout this thesis, I will assume without argument that normative reasons are considerations that count in favour of (or against) a response. On this view, for the fact that I have to get to work to be a reason to leave the house is for this fact to count in favour of leaving the house. Conversely, if having the flu is a reason not to leave the house, this means it counts against doing so. Plausibly, normative reasons play a role in determining what I ought to do: if my reasons to leave the house outweigh my reasons to stay at home, I ought to leave the house.

Finally, we have the ‘aptic’ categories[3]. These include fittingness, being appropriate, and being merited. These terms pick out relations between responses and their objects. This is the relation of *worthiness* (Howard, 2019, p. 216). Attitudes are the paradigmatic case of the sorts of responses that can be fitting, as they clearly have objects. Fear, belief, and admiration are all directed at things in the world: I fear a dangerous animal, believe a proposition, and admire public figures. Our attitudes do not always get things right however. When I admire a stone, for instance, it would seem my emotions have got things wrong; my response is unfitting. On the other hand, it is plausibly fitting to admire Nelson Mandela. This is to say that Mandela is worthy of admiration, while an ordinary stone is not. From this normative relation of fit we can derive ‘thick’ fittingness categories, as in ‘Mandela is *admirable*’. Thick-fittingness categories are analysable in terms of the thin fittingness relation: x is admirable if and only if it is fitting to admire x.[4]

How are we to explain the nature of these normative categories?[5] As we saw in the introduction, a prominent set of answers involves the explanation of one normative category in terms of another. So perhaps we can explain value in terms of reasons, or reasons in terms of fittingness, or any explanatory combination of the four categories introduced above. X-First is the view that there is *one* normative category in terms of which *all* other normative categories can be explained. The most prominent options discussed in recent literature are Reasons-First, advocated by Scanlon (1998), Derek Parfit (2011), Rach Cosker-Rowland (2019) and Mark Schroeder (2021a and b), Fittingness-First defended by Conor McHugh and Jonathan Way (2016, 2022) as well as Christopher Howard (2019), Value-First proposed by Cosker-Rowland (2022)[6] and Ought-First, a view suggested by the work of John Broome (2013 and 2015) when he endorses an explanation of reasons in

[3] This label is due to Selim Berker (2024, p. 1053).
[4] As we move through this thesis, various aspects of the generic formulations of the four normative categories just given will be questioned and discussed in greater detail.
[5] There are of course other potential normative categories, such as virtue and vice, and rationality and irrationality. However, since these do not feature as candidates for X in any prominent X-First views that I am aware of, I will focus in this chapter on the four kinds of categories mentioned above.
[6] Several authors suggest value-based reductions of certain normative categories without fully committing to a Value-First programme. For instance, Barry Maguire (2016) defends a value-based theory of reasons but does not explicitly endorse value-based theories of all other normative categories. In the specific sense that I am going to elaborate in section 2, this would disqualify him from being an X-First theorist.

terms of ought. This chapter makes a case for the rejection of all X-First views. Call this position No-X-First. I argue that there is *no* master normative category in terms of which all others can be explained.

Section 2.2 isolates the version of X-First I am interested in by distinguishing it from adjacent views. Two key commitments of X-First views are established. These provide the foundations for my objection to all X-First views in section 2.4. Section 2.3 presents two considerations that are sometimes thought to motivate X-First views. I argue that these putative motivations do not in fact favour X-First views over No-X-First. Section 2.4 then offers a reason for favouring No-X-First over X-First: there is a high degree of phenomenological variation among the four normative categories. This variation in how the different normative categories *feel* poses a general problem for all X-First views. Section 2.5 briefly concludes.

2.2. X-First Views

Let's begin with four closely related theses:

> **X-Priority**: there is at least one normative category X such that some other normative category Y can be explained in terms of X.
>
> **X-Fundamentalism**: there is at least one normative category X such that X cannot be explained in terms of any other normative category.
>
> **X-First**: there is a single normatively fundamental normative category X such that all other normative categories can be explained in terms of X.
>
> **The Mark of the Normative**: there is a single normative category x such that normativity itself can be explained in terms of x.

Disentangling X-First from the other three theses will help to clarify what exactly the view says. First, X-First views are independent of The Mark of the Normative. It will be noticed that I have so far made no attempt to define *normativity*. This is because X-First views can remain neutral on this issue. We might think, along with Joseph Raz (1999, p. 67), that what it is to be normative is to be related in some way to reasons, i.e. reasons are the mark of the normative. X-First views do not go this far. Reasons-First, for example, is the claim that the deontic, evaluative and aptic categories can all be explained in terms of reasons, where 'being explained in terms of' can mean a number of different things, like being reducible to reasons, analysable by reasons, or grounded in reasons. I will say more about the kind of explanation I will be working with shortly. To distinguish X-First from the Mark of the Normative thesis, notice that Reasons-First is compatible with the further claim that the nature of normativity itself is best understood as an area of thought or activity which essentially involves reasons, but it is also compatible with the claim that normativity is ultimately explainable in terms of natural features of the world, or that it is unexplainable, or even that it does not exist. X-First is an ordering of the set of normative categories, if they exist and if they are normative, where X is explanatorily prior to all the others, whether or not the set as a whole is to be explained in terms of X, or in some other way.

There is one feature of normativity itself that is often stressed in discussions of X-First views: they concern ‘robust’ or ‘authoritative’ normativity. This is best understood in contrast to ‘formal’ normativity. The norms of etiquette are merely formal in the sense that it does not really matter which hand I use to wield my fork. The cutlery norm is no more than a social convention. However, “it seems to matter in and of itself that you keep your promises, look after your health, and follow the evidence” (McHugh and Way, 2022, p. 9), and this is because the norms that apply in these cases are authoritative in some way, unlike the merely formal norms of etiquette. Authoritative norms are not *optional* in the sense that non-authoritative norms such as those associated with good etiquette do seem to be optional. Whether you sign off an email in the appropriate way is optional. Whether you refrain from theft is not. This can also be glossed in terms of a kind of *binding*: we are bound by the norms of morality, which is the paradigm of authoritative normativity, but we are not similarly bound by merely formal norms. I will assume that the kind of norms discussed in this chapter in the context of the X-First debate are authoritative, rather than merely formal, although it should be noted that this is itself up for debate: for instance, Cosker-Rowland (2022b) provides an interesting discussion of when the fittingness norm is authoritative, and when it may not be.

Secondly, although X-First entails X-Fundamentalism, the latter does not entail the former. This is not always made clear in the literature. At one point, Mark Schroeder seems to conflate the two views when he says in his paper entitled ‘The Fundamental Reason for Reasons Fundamentalism’:

> "As I will be using the term, 'reasons fundamentalism' denotes the thesis, not that reasons are absolutely fundamental, but that they are *most* fundamental, among the class of normative properties and relations. It is the thesis that every normative property or relation must be explained in terms of reasons, or as it is sometimes put, in a phrase that would make my title less cute, that reasons come *first*" (Schroeder, 2021a, p. 3110).

Whether cuteness is a good reason to run the two views together I will leave up to the reader. In any event, here is a reason to prise them apart: identifying fundamentalism with firstness unduly biases the debate against the view that there is *no* category which grounds all the others. If X-Fundamentalism and X-First amount to the same thing, then the thesis that X cannot be defined in terms of any other category, and the thesis that all other categories can be defined in terms of X, would amount to the same thing. But this blocks the possibility that there is no X in terms of which all other categories can be defined. X-Fundamentalism just says that X cannot be explained in terms of some other normative category. It does not say that all other normative categories can be explained in terms of X. So, X-First is not entailed by X-Fundamentalism, although if X is first, then it is also Fundamental.

Finally, while X-First does not go as far as The Mark of the Normative, it goes further than X-Priority. X-Priority is a family of views about the relative priority of one normative category over another. An example is a fitting-attitude theory of value, a standard version of which says that for x to be good is for x to be a fitting target of a pro-attitude. According to this view, the evaluative category 'good' is grounded in the aptic category 'fitting pro attitude'; goodness is explained in terms of fitting pro-attitudes. Although this view implies that fittingness is first relative to value, we do not yet have a Fittingness-First view. For this we would need an accompanying story about how reasons and oughts are reducible to fittingness as well. If Fittingness-Priority is true relative to each of the three other normative categories, then Fittingness-First is true.

Now, in characterising X-First so far, I have been talking quite loosely about some normative category X 'defining', 'explaining' or 'grounding' other normative categories. The literature reflects this fuzziness. Different writers use different terminology. For instance, Cosker-Rowland (2019, p. 193) talks about 'reductive analyses', McHugh and Way (2022, p. 13) offer a 'constitutive account' of reasons as fittingness, while Schroeder (2021a, p. 3108) and Selim Berker (2018, p. 730) use the term 'grounding'. To make things precise, I will use the notion of grounding, and say what I mean by it. There is good precedent for this. In an overview of the fitting-attitude analysis of value, one component of Fittingness-First views, Howard (2023) points out that common to most contemporary fitting-attitude theorists is the claim that facts about the fittingness of certain human responses *ground* evaluative facts.

Additionally, Daniel Wodak, who explicitly discusses the general X-First programme in a similar vein to the present chapter also talks in terms of grounding (Wodak, 2020, p. 50). So, X-First says that all normative categories are grounded in X. But what exactly is grounding? Following Berker (2017, p. 730) and Shamik Dasgupta (2020, p. 75), it simply means *because*. However, it is a particular kind of 'because'. Here is Dasgupta to illustrate what I have in mind:

> "Why is there a table here? One answer: Because someone put it there yesterday. Another answer: Because there are pieces of wood arranged table-wise. These answers are not in competition. The first has to do with the causal history that led to the table being here; the second explains what it is about the current situation that makes it the case that there is a table here. The former is called a causal explanation, the latter a *constitutive* explanation"
> (Dasgupta, 2020, p. 75)

The grounding involved in X-First views is what Dasgupta calls 'constitutive explanation'. When a defender of Fittingness-First says that reasons, oughts and values are grounded in fit she is saying that facts about fit *make it the case that* facts about the other normative categories obtain, or, to use Berker's (2017, p. 731) preferred locution, it is *in virtue of* facts about fit that facts about the other categories obtain, just as it is in virtue of wood being arranged table-wise that a table obtains. So, 'grounding' is a kind of 'constitutive explanation', and I will use these terms interchangeably in the rest of this chapter. There are two features of X-First views as constitutive explanations I want to stress. The first:

> **Explanatory Completeness**: assuming x and y are normative categories, if x is explanatorily prior to y, then x must be able to explain all of the normatively significant aspects of y.

Although this is not a widely held principle, there is some precedent for it in the recent literature. When referring to the 'single, fundamental normative notion' represented by X in the X-First debate, Simon Kirchin says that he means, "and others may mean that notion that all other candidate normative notions can be analysed in terms of, or at least that notion that can analyse the normative aspect of those other notions (Kirchin, 2025, p. 13). This is what I mean by the explanatory completeness condition on X. Are there any more specific reasons why the X-Firster should accept what I have called Explanatory Completeness above? Well, it's in the name: X-*First*. Recall the nature of the constitutive question being asked, and the distinction between X-First views and The Mark of the Normative. When we ask 'what is goodness?', we are looking for a constitutive explanation. Reasons-First, for instance, claims that reasons can provide this explanation. If there is a normatively significant aspect of goodness that cannot be explained in terms of reasons, then Reasons-First would be false. This is because reasons would then be *Joint-First*, where both reasons and whatever it is that explains this other aspect of value come first relative to goodness.

Furthermore, Explanatory Completeness is compatible with there being a different answer to the question 'what is normativity?'. Perhaps the nature of the class of normative categories to which goodness belongs is explainable in terms of some natural property such as human desire, or maybe it is a matter of divine decree, either way this is independent of X-First. According to Reasons-First, reasons explain everything there is to explain about the good, *assuming* that both the good and reasons are normative categories, even though what it is for goodness or reasons or anything else to be normative may be grounded in some other way.

One reason an X-Firster might reject Explanatory Completeness is that grounding is overdetermined. Suppose we want to know what makes it the case that there are concrete objects in the room I am sitting in. This fact is fully grounded by the fact that there is a table in the room. But it is also fully grounded by the fact that there is a bottle of water in it, or a sock, or a particular strand of DNA and so on (Ricki Bliss, 2023, p. 3). The thought would then be that there are aspects of the fact that there are concrete objects in the room left unexplained by the table fact. These aspects are to be explained by one of the other facts, and so Explanatory Completeness is false. Similarly, if normative categories are overdetermined, then the X-Firster could claim that although x fully grounds y, there may be some normatively significant aspect of y left unexplained by x, and that this aspect is to be explained in terms of another thing z, which also fully grounds y.

This argument does not work. For one thing, it fails to respect the 'firstness' of X-First views. As we have already noted, X-Fist is distinct from X-Joint-First, where this latter view would allow that X explains some of the normatively significant aspects of y, while some other category explains some of y's other normative features. Allowing for overdetermination at the level of the constitutive explanation of normative categories in terms of other normative categories contravenes Explanatory Completeness and can only deliver X-Joint-First views, as opposed to X-First. Furthermore, it is not even clear that overdetermination entails explanatory *in*completeness. Take another paradigmatic example of a constitutive explanation: 'losing a game of chess' is fully grounded by 'being in checkmate' (Dasgupta, 2020, p. 75). Even if losing a game of chess is overdetermined, there does not seem to be anything about this state of affairs that is left unexplained by being in checkmate. Of course, *how* I lost will need further explanation, perhaps in terms of the fact that I made such and such a blunder, but this is a causal explanation of my losing, rather than a constitutive one.

The same is true for normative categories. While in one sense they can be massively overdetermined, this is compatible with Explanatory Completeness. Suppose Stephen Jay Gould was a good writer of popular science. Lots of different things may ground this normative fact: his conversational yet authoritative tone, a knack for simplification without dumbing-down, and the ability to construct overarching narratives about life and death from seemingly insignificant details of the natural world. While each of these features fully grounds Gould's goodness, and there may be aspects of his goodness which any individual feature cannot fully explain on its own, this is a matter of *causal* explanation; a question of what *makes* Gould a good writer. It is consistent with the constitutive nature of goodness being fully explainable in terms of reasons, or fittingness, or the view that goodness is primitive. ‘What features of A make it good?’ and ‘what is goodness?’ are two distinct questions. The first is clearly overdetermined, while the second, if some X-First view is true, may not be. Here is the second feature I attribute to X-First views:

> **Normative Phenomenology and Explanation**: a complete explanation of x, where x is a normative category, must account for the phenomenology of x.

Section 4 proposes a general objection to all X-First views which says that no one normative category can account for the phenomenological variation among the four kinds of categories. However, even if the X-Firster agrees that her view involves the kind of grounding described above, and even if she accepts that this commits her to Explanatory Completeness, she might still deny that the experiential ‘feels’ of normative categories must be included in her explanation. The complete constitutive explanation of ‘losing a game of chess’ in terms of ‘being in checkmate’ makes no reference to how losing feels, and this is quite obviously as it should be.

Likewise, the X-Firster could argue that the metaphysical account of normative categories she is providing need not capture the phenomenology of her analysandum. This would be an easy way to block the sort of argument that I am going to present in 2.4. Crucially however, normative categories are not like losing a game of chess; it is plausible that the phenomenology of normative categories is essential to what they are, metaphysically speaking. That is, the phenomenology of y, where y is a normative category, is a normatively significant aspect of y, and so by Explanatory Completeness, X-First views which are to account for y must account for y’s phenomenology. I am going to argue for this claim by drawing attention to three substantive issues about the constitutive natures of value, fittingness, and reasons, each of which partly turns on phenomenological considerations. If such considerations are at the heart of debates about the nature of normative categories, then we have some reason to think that normative categories are partly constituted by a particular ‘what-it’s-likeness’, or phenomenal quality, just as being in pain is partly constituted by a particular phenomenal quality. If this is right, then a full constitutive explanation of a normative category must account for its phenomenology.

Begin with value. John McDowell kicks off his influential discussion of value and secondary qualities by agreeing with J.L. Mackie (1977, pp. 30-35) that “ordinary evaluative thought presents itself as a matter of sensitivity to aspects of the world" (McDowell, 1998, p. 131). On this view, one aspect of the phenomenology of value is that it is *objective*. When I witness a good deed – someone rescuing a cat stuck up a tree, say – I experience the goodness as an objective feature of the action, just as I experience the ginger colour of the cat's fur, or the triangular shape of its ears, as objective features. Here is where the two writers part ways. For Mackie, only the triangular shape, which is a primary quality, actually exists. Colours and values may appear to us as primary qualities, but for various reasons, including the fact that they evade empirical testing, they are not in fact primary qualities and do not exist. McDowell is more permissive. He allows that even though colours and values present themselves to us as primary qualities, they are in fact secondary qualities – dispositions of objects to appear a certain way to certain observers – and such qualities both exist and are perfectly objective.[7] Which of these views is closer to the truth is not important. The point I want to stress is that phenomenological considerations are playing a central role in a debate between the two writers about a normative category. The 'objective' phenomenology of value acts as a shared starting point on top of which McDowell and Mackie construct alternative arguments about the existence and nature of evaluative properties.

When it comes to practical reasons, consider the popular thesis that reasons have weights. This is a claim about the metaphysical nature of reasons; they are things that are weighted. One reason can be stronger than another, while some reasons are so insignificant they barely affect the scales. This thought is highly plausible. Consider an all-too-familiar case of practical reasoning: a recent graduate needs to make a career choice. She can either take a well-paid, stable job in the civil service, or pursue postgraduate study in philosophy. How does she decide? She deliberates. She *weighs* the various pros and cons of each option; the reasons for and against.

The fact that reasons are routinely weighed against each other in this way plays a role in both meta and normative ethics. Some people think that a notion of weighted reasons is all we need to explain the nature of the all-things-considered ought: what the graduate ought to do is just what she has *most* reason to do when all the relevant

---

[7] This way of framing the debate between Mackie and McDowell is indebted to Maximilian de Gaynesford's discussion (de Gaynesford, 2004, pp. 171-174)

considerations have been weighed (Snedegar, 2016, p. 158), while Errol Lord and Barry Maguire (2016, pp. 3-8) have argued that some weighted category is indispensable to all normative theorising, be it of a Kantian or consequentialist bent. Referring to deontological and consequentialist accounts of what one ought to do respectively, they write:

> "Whether one ought to promise-keep or pleasure-maximize in some situation may depend on further facts. In particular, it might depend on the presence of other normatively relevant act-tokens, and on the relative importance of the promise-keeping or pleasure-maximizing in that situation. Talk about the 'importance' of considerations is usually talk about their weight" (Lord and Maguire, 2016, p. 5).

Thus, both consequentialist and Kantian answers to questions about which actions are right and wrong require reference to some process of weighing, and weighted reasons are the best candidate to fulfill this role. What's important for us is that this proposed feature of reasons, that they have weights, is in part a phenomenological feature. The graduate's dilemma is adapted from Schroeder, where he describes what's going on like this:

> "As Abby thinks through each of the relevant aspects of the options that she is considering, they phenomenologically strike her as counting in favor of one or another of the two options. This is why she finds it so natural to use expressions like 'but' and 'on the other hand' as she shifts between considering features that feel like they point or weigh in different directions. This phenomenology—the phenomenology of different considerations pointing or weighing in different directions in a choice situation—is the phenomenology of reasons." (Schroeder, 2021b, pp. 3-4).

If the weight of reasons is partially a matter of the phenomenological *feeling* of certain considerations pointing or weighing in different directions, and we think that being weighted is partly constitutive of what a reason is – and perhaps what the all-things-considered ought is – then the phenomenology of reasons is essential to what reasons are and at least relevant to the question of what oughts are.

Finally, does fittingness belong within the deontic family of categories? As with values and reasons above, this is a question about the metaphysical nature of fit. Selim Berker (2022, pp. 36-40) argues that fittingness is not deontic, and part of his argument hinges on the phenomenology of fit. For Berker, if fittingness were deontic, we should be able to say whether it entails a kind of permission, or a kind of requirement. However, we cannot do this, because in some cases of fittingness calling it a permission is too weak, while in other cases calling it a requirement is too strong. For example, if I am caught stealing, then shame is a fitting response. To say I am merely permitted to feel shame seems too weak: it would be appropriate for you to tell me that I *should* be ashamed of myself. On the other hand, when a cat is lovable, this means it is fitting to love. Here it seems too strong to say I am required to love the cat. Plausibly, part of what it means to say that a particular interpretation of fittingness 'seems too weak' or 'seems too strong' is that in the first case it *feels* more like a requirement than a permission, and in the other the opposite is true. James Fritz (2023, p. 2600) takes this idea forward and offers an account of fittingness according to which it is sometimes 'demand-like' and sometimes 'permission-like', which I read as entailing that fit sometimes *feels like* a demand or requirement and sometimes like a permission. Once again, we have a case where phenomenological considerations are relevant to theorisation about the nature of a normative category.

Although these phenomenological factors bear on their respective categories in different ways, the general implication should be clear: phenomenological considerations are relevant to metaethical issues about the constitutive explanation of normative categories. If this is true, then it is reasonable to expect that a complete explanation of a particular normative category must account for the phenomenology of that category. Even if this expectation is unwarranted, at the very least the preceding discussion has made an initial case for the claim that there are important differences between the phenomenologies of different normative categories, and it is these differences that I am going to exploit in my argument against X-First views in section 2.4.

2.3. Two Motivations for X-First

In this section, I argue that two standard motivations for X-First do not in fact support X-First over No-X-First. At the outset, it should be noted that most writers focus on motivating local X-Priority claims, such as the view that value is grounded in reasons for pro-attitudes. In this regard, Schroeder makes a useful distinction between indirect and direct arguments for X-First, with indirect arguments being those that focus on trying to prove some X-Priority claim, and direct arguments, which focus on showing that all normative categories are grounded in X (Schroeder, 2021a, p. 3108). In what follows, I consider two indirect motivations from a direct perspective. One reason often given for the explanation of value in terms of some other category is 'demystification' (McHugh and Way, 2016, p. 578) and Rabinowicz and Rønnow-Rasmussen, 2004 p. 400).

> **Demystification**: certain normative categories are mysterious. The fact that explaining mysterious categories in terms of less mysterious categories will demystify the former is a reason that counts in favour of such an explanation.

It might be thought that *good simpliciter* is a mysterious category; what could it mean for something to be just good, rather than say good for A, or good as a B? Many find the generic, unqualified notion of goodness quite strange. A prominent example of this sort of argument can be found in Philippa Foot's (1985), where she argues that talk about good outcomes or states of affairs is incoherent. If the strange category of goodness simpliciter could be fully grounded by some more intuitive normative category, say the category of a reason, then this would speak in favour of such an explanation. This is clearly a local issue, in that it motivates Reasons-Priority with respect to value. However, the point generalises to global X-First theses as well. Some people find fittingness a bit odd, and the recent spike of interest in the category mystifying. The guiding metaphor can seem more unhelpful than edifying: we have a very clear grasp of what it is for a shoe to fit a foot, mainly through our senses of touch and vision, but what could it be for an attitude to fit its object? Perhaps fit can also be explained in terms of reasons. A powerful motive behind Reasons-First would then be that we can demystify both good simpliciter and fittingness in one fell swoop. In general, if there is one basic normative category that explains all others, such a view would be attractive insofar as explaining the nature of our mystifying normative world is a central aim of normative theory.

One problem with this approach is the issue of which normative category is up to the task of demystification. If grounding y in x is supposed to demystify y, then x cannot be at least as mysterious as y. Of course, it may be that the proper explanation of y *does* require reference to some equally mysterious phenomenon, or perhaps something even more mysterious, but if this is the case, then demystification cannot be given as a primary reason to accept the grounding of y by x. To bring back an example mentioned in the introduction, suppose we are living in pre-modern times and we want to explain the nature of earthquakes. If plate tectonics is offered as an explanation, and nobody has any idea what plate tectonics is, then the mere fact that this explanation demystifies earthquakes can hardly be given as a reason to accept it, since we have simply replaced one mystery with another, and this is true even if there are lots of other reasons to accept the explanation of earthquakes in terms of plate tectonics. The difficulty for views that seek to explain one normative category in terms of another is that a large part of what makes categories such as good, fit, and reasons so mysterious is their very normativity, which is something they all share. Mackie famously argued that if categorical or authoritative normative categories ('objective values' as he calls them) really existed, then they would be very strange - utterly different from anything else in the universe (Mackie, 1977, p. 38). This is because authoritative normative categories have a kind of motivational force built into them. Mackie refers to Plato's form of the good which *makes* the person who knows it pursue it, and the 'fitness' between situations and actions as discussed by the 18th century writer Samuel Clarke, which has a *demand* for some particular action built into it (Mackie, 1977, p. 40).

While the Platonic good and Clarke's fitness may not align exactly with the categories we have been discussing, the general point is the same: goods, oughts, fit, and reasons all have a kind of motivational force built into them – they are supposed to be things that entail categorical imperatives – and since this is a strange property, they are all strange to the same extent. What the X-Firster is doing then is proposing an explanation of one mysterious category in terms of another, equally mysterious one. In response, one might say that if the motivational force in question is just what we mean by 'normativity', then X-First views need not take a stand on this mystery, which is instead the concern of defenders of The Mark of the Normative thesis. As we saw in section 2, the X-Firster does not owe us an explanation of the normativity of normative categories. However, if Mackie is right, and the normativity of normative categories is one of the most mysterious things about them, then citing demystification as a reason to be an X-Firster is at best an odd strategy. If we are serious about demystifying normative categories, then we had better have a way of explaining the strange motivational force they all share as normative categories, and if X-First views are silent on this, then they cannot honestly claim to be demystifying. A second motivation for X-First views is Qualitative Parsimony:

> **Qualitative Parsimony**: we should not posit the existence of new kinds of categories in addition to the ones there must be unless positing these new kinds of categories allows us to explain something. If theory A entails the existence of a new kind of category which is not entailed by theory B, and both theories adequately explain their subject, then theory B is preferable to theory A (adapted from Cosker-Rowland, 2019, pp. 70-71).

What is the nature of fire? One theory appeals to a substance called phlogiston contained in combustible materials which is released through the process of burning. Lavoisier's experiments leading to the isolation of oxygen could explain fire without reference to phlogiston. This latter category was therefore rendered explanatorily redundant. Qualitative Parsimony explains why we are justified in discarding the phlogiston theory in favour of the oxidation theory: when deciding between theories A and B, the theory which is less qualitatively parsimonious should be rejected, other things equal.

Cosker-Rowland applies this reasoning to argue for Reasons-Priority with respect to value against No-Priority. Reasons-Priority with respect to value is the view that value is grounded in reasons for pro-attitudes. No-Priority is the view that neither reasons nor values are grounded in the other. For Cosker-Rowland, we should favour Reasons-Priority over No-Priority because No-Priority is less qualitatively parsimonious than Reasons-Priority. This is because, according to Reasons-Priority, value can be fully explained in terms of reasons for pro-attitudes, so we only need one kind of thing in our ontology. No-Priority on the other hand entails the existence of evaluative categories *in addition to* reasons for pro-attitudes. This forces us to admit two distinct kinds of things into our ontology. No-Priority is therefore less qualitatively parsimonious than Reasons-Priority, which gives us a reason to reject the former in favour of the latter, other things equal.

This argument also generalises from Reasons-Priority to Reasons-First, or indeed any X-First view. When deciding which of X-First and No-X-First to accept, the thought would be that X-First is preferable, others things equal, insofar as it explains all normative categories in terms of X, with the other categories being explanatorily redundant. If we go for Fittingness-First for example, we can explain everything there is to explain about reasons, oughts and value in terms of fit, without having to posit the existence of any further categories. If Qualitative Parsimony really is a theoretical advantage, this gives Fittingness-First an edge over No-X-First: we should favour a simple X-First view over the metaphysically profligate No-X-First.

I am not convinced that qualitative parsimony is the coup X-Firsters think it is, and for two reasons. First, in normative theory, considerations of parsimony are most at home when it comes to the ontological question. Anti-Realists deny the existence of normative categories. Realists think that normative categories exist, but they differ on what *kind* of categories they are; some Realists are Naturalists, while some are Non-Naturalists. Very broadly speaking, for Naturalists, normative categories are identical with some property or categories studied by the natural sciences, such as physics, biology, social sciences etc, and are generally known a posteriori. Non-Naturalism is then defined negatively: normative categories are not natural, they are different in kind from the things studied by the sciences, and can be known a priori (Shafer-Landau, 2003, pp. 58-63).

Now, Qualitative Parsimony is clearly relevant to the question of whether the Realist should be a Naturalist or a Non-Naturalist. Naturalism is more qualitatively parsimonious, as it promises to explain the ontological status of normative categories in terms of the kinds of things we are already committed to, i.e. natural categories, whereas Non-Naturalism commits us to the existence of an additional kind of thing, i.e. non-natural categories. We should therefore favour Naturalism over Non-Naturalism, other things equal. But what does all this have to do with X-First views? Well, nothing really, and that is the problem. X-First views are compatible with both Naturalism and Non-Naturalism. If we say that normative categories are natural categories, then each of reasons, values, oughts and fittingness are the same kind of thing. They are also the same kind of thing if they are all non-natural categories. On both views, qualitative parsimony is irrelevant to X-First vs.

No-X-First, since neither view posits some additional *kind* of thing[8]. Of course, X-First is more 'quantitatively parsimonious' than No-X-First, on both Naturalist and Non-Naturalist readings, insofar as it is a theory that explains everything with just one category, rather than several, but this is a separate issue, and as Alan Baker notes, it is qualitative rather than quantitative parsimony which drives the intuition that simplicity is a theoretical virtue (Baker, 2022).

Secondly, even if we accept that reasons, oughts, fit, and value are different *kinds* of things, such that qualitative parsimony does matter, it is far from clear how we would determine which of our four normative categories is the *additional* thing over and above what we are antecedently committed to. In the case of the phlogiston vs. oxidation theories of combustion, it is quite clear that phlogiston is the additional thing, since we have evidence for the existence of oxygen unrelated to questions about the nature of fire, such as the role it plays in the respiratory system of most organisms. When it comes to normative categories, what sort of evidence could the X-Firster appeal to for the claim that we have independent grounds for believing in the existence of X, while the other categories are additional? The onus is on the X-Firster who appeals to Qualitative Parsimony to provide such evidence.

Finally, even if my attempts to show that qualitative parsimony does not favour X-First over No-X-First are unconvincing, the highly defeasible nature of arguments from parsimony means that the advocate of No-X-First need not concede defeat. I am going to argue that there is no x First because there is no single normative category which can fully explain the diverse phenomenology of all other normative categories. If this argument goes through, we will have good reason to accept No-X-First even if X-First would otherwise be the more qualitatively parsimonious account.

[8] It might be the case that X is a natural category while the other three are non-natural. On this sort of view, X-First would be more qualitatively parsimonious than No-X-First. As far as I know, nobody has defended this sort of view in print. Errol Lord (2023) in his "Everything First" defends the view that reasons are First in the normative domain, but that reasons are grounded in oughts, fittingness and value, which are non-normative. This is potentially more qualitatively parsimonious than No-X-First, although it is an idiosyncratic position, given that it denies the normativity of things like fittingness and oughts.

2.4. Against X-First

I now offer a general objection to all X-First views: whichever of the four categories one takes to ground the other three, that category cannot fully explain all the others. Here is why:

The Argument from Phenomenology:

1. If X is First within a set of normative categories S, then X must fully explain the phenomenological profiles of all the other members of S.
2. There is no normative category which fully explains the phenomenological profiles of all the members of S, where S is made up of reasons, oughts, goods, and fittingness.

C. Therefore, there is no X First.

By the phenomenological profile of a normative category I mean the distinctive 'what-it's-likeness' of the category. Normative categories present themselves to us in distinctive ways when we encounter (or think we encounter) them; having a reason or an obligation, and noticing the goodness of a thing or the fittingness of an attitude, are all accompanied by a particular phenomenological experience. If the two features of grounding as constitutive explanation introduced in section 2.2 hold, such that X-First views must fully explain their targets, and a full explanation of a normative category must account for its phenomenology, then the fact that the phenomenology of X is very different from the phenomenology of Y provides some reason to doubt that X is First with respect to Y.

One important caveat before continuing. In what follows, I am going to describe aspects of the phenomenology of normative categories. It must be acknowledged from the outset that this is very far from an exact science, and I make no use of the Phenomenological methods developed in the tradition of Husserl. What I do is rely quite heavily on my own intuitions about normative judgements. However, I also provide justification and corroborating evidence to back up these intuitions, mainly in the form of theoretical considerations and the commitments of established authors in the field which suggest the phenomenology I assign to particular categories. If the reader introspects and finds that my description of the phenomenology of a given normative category does not accord with his or her experience, this should not be taken as a barrier to accepting my argument. I do not need to prove that my way of carving up the phenomenological space in the normative domain is the only way to do so, I just need to draw out the fact that there is enough variation amongst the categories we are interested in to prevent any particular X from fully explaining all the normatively relevant features, including the phenomenology, of the entire set. So, to get a sense of the phenomenological variation I have in mind, consider the following situation:

> **The Concert and the Funeral**: Aisling, a teenager, has been beyond excited for an upcoming concert. Her all time favourite band are coming to play in her hometown on what is likely to be their final tour. She has never seen them before, and she has had her tickets for months. A week before the show, she receives some devastating news. Her close friend's mother has passed away. The funeral will be held on the day of the concert.

There are a number of normative judgments about Aisling's situation which, for the sake of argument, we can suppose are true:

1. Aisling has some reason to go to the concert, and some reason to go to the funeral.
2. Aisling ought to go to the funeral. She ought not to go to the concert.
3. In response to a tragic death, grief is a fitting response. In response to a great concert, enjoyment is a fitting response, other things equal.
4. The band is very good. The concert will be very good. If Aisling were to go to the concert, it would be one of the best experiences of her life.

What are the normative categories employed in these judgments like, phenomenologically? We noticed in section 2.2 that reasons point and pull the agent in different directions. The fact that this will likely be Aisling's last chance to see her favourite band perform live is a reason she feels as pointing or weighing in the direction of going to the concert, while the fact that being there will support and comfort her friend is a reason she feels as pointing or weighing in the direction of going to the funeral. The greater the weight of a reason, the greater the pull to do what it favours.

Furthermore, encountering reasons of this sort is ‘objective’ in the Mackie-McDowell sense, they at least appear to be facts out there in the world to be discovered. This way of characterising the phenomenology of reasons has explanatory pay-offs. For instance, consider the fact that reasons can be ‘disabled’ (Dancy, 2004, p. 41). Suppose Aisling really admires the band’s frontman insofar as he has a great voice and plays the guitar like Hendrix, and that this is a reason for her to go to the concert. Now suppose that he comes out as a neo-fascist and begins vocally advocating “ethnic cleansing” in his home country. Let’s also say that Aisling ought not to go to the concert. One way to think about this is in terms of the frontman’s nasty politics outweighing the quality of his music, giving Aisling more reason not to go than to go. However, it is also plausible that the quality of his musicianship, even if it remains a reason, is for all intents and purposes disabled; there is no possible increase in the brilliance of his guitar work that would tip the balance in favour of going. The functioning of the frontman's musical abilities as normative reasons is disabled by his politics. This disablement can be understood in terms of the nullification of the pull these would-be reasons exert on Aisling. Their weight evaporates.

Whereas reasons pull the agent in different directions, ought in the all-things-considered or overall sense pulls the agent in one particular direction. When an agent recognises that she ought to do some action A, she feels this as a demand to do A that exerts a force on her in the direction of A[9]. Consider again Aisling’s situation, and suppose that what she really wants to do is to go to the concert. We can imagine her asking another friend for advice about what she ought to do, to which her friend replies ‘you already *know* what you ought to do’. Supposing, as seems plausible in such an exchange, that Aisling says ‘yes, you’re right’, and she does know what she ought to do, and what she knows is that she ought to go to the funeral instead of the concert, this would strike her as an external demand insofar as it is not dependent on her desires or preferences, given that she really wants to go to the concert. Oughts thus present themselves to us as ‘objective’ in the Mackie-McDowell sense; they are “experienced as emanating from ‘outside’ us, and as being directed against us” (Mandelbaum, 1955, p. 54).

---

[9] This is a common way of characterising the phenomenology of ought. It appears in Maurice Mandelbaum (1955, p.54), which is later discussed in Terry Horgan and Mark Timmons (2005, p.60)

Next, we come to fittingness, which is potentially more complicated with respect to the phenomenological considerations discussed so far. As we saw in section 2 above, fit does not feel as demanding as ought. Recall the fittingness of love toward a loveable cat. No matter how loveable a cat I encounter is, it seems quite plausible that I can recognise that love would be a fitting response without feeling any associated demand to love the animal. However, we also saw that fitting shame does seem to demand being ashamed.

One possible explanation for this is that fittingness sometimes feels like a demand, and sometimes does not, depending on the attitude in question. However, this is not the only explanation. Notice how shame, the paradigmatic example of an attitude the fittingness of which supposedly exerts a kind of demand, is often connected to the deontic categories in interesting ways. Generally, it is fitting for an agent to feel shame when she does something shameful. This can be an act or an omission. When is an action or omission shameful? Plausibly, when the agent does something she *ought not* to have done, or fails to do something she *ought to* have done. It is fitting to be ashamed of stealing, and you ought not steal. It is fitting to be ashamed of breaking a promise, and you ought to keep your promises. If this is right, it may be the case that the demand associated with fitting shame is in fact the experience of an obligation of some kind being broken. With that said, there do seem to be cases where shame can be fitting even though the agent has done nothing wrong: perhaps it is fitting to feel ashamed in a sense that demands the emotion when somebody walks in on you unclothes in a compromising position, even though you have not violated any oughts. So, fittingness is either sometimes demand-like, depending on the attitude in question and the particular circumstances, or not really demand-like at all. In any event, it is certainly not always demand-like in the way that oughts are, as evidenced by the lovable cats case.

Furthermore, we may have reason to doubt that fittingness presents itself as objective in the way that oughts do. It might be thought that fit feels like obligation insofar as it emanates from outside the agent: the experience of fitting amusement, for instance, is just the experience of funniness, which presents itself to the agent as an object feature of the object of amusement, a joke, say. However, this may be overly simplistic. In many cases, whether some attitude fits an object in a particular situation arguably depends on various background mental states of the agent, and experiencing fit in these contexts is a subjective experience.

Imagine there was no funeral for Aisling to attend, and she goes to see her favourite band. The performance is great, and enjoyment is fitting; the concert is enjoyable. Now consider the case where there is a funeral she should go to, but she goes to see her favourite band instead. As in the previous case, the performance is great, the music fantastic. In a sense, it is still 'enjoyable', however, it would be unfitting for Aisling to enjoy it. This is because she skipped the funeral to be there. Perhaps guilt would be the more appropriate response. This is some reason to think that Fittingness is not a purely objective relation 'outside' of the agent, but one that is mediated by the particular agent's subjective circumstance[10]. Another example is fitting pride. For something to be prideworthy is for it to be the fitting target of pride. Pride is clearly an agent-relative attitude, with agent-relative fittingness conditions. When I encounter the prideworthiness of a family member's achievement, this is not experienced as something purely objective, but intimately tied to my particular position. This is what partly explains the peculiarity of taking pride in a stranger's achievements: we can suppose that scoring the 'goal of the century' is a prideworthy feat, but this does not mean that it is fitting for everyone to be proud of it - it is only fitting for Diego Maradona, his close friends and family, and perhaps Argentinian football supporters to feel proud of the goal. These examples speak against a purely objective phenomenology of fit.

Finally, let's look at value. I think Mackie and McDowell are basically right that goodness (and badness) does present itself to us as an objective feature of objects, outside and independent of the agent. When Aisling attributes goodness to the band she loves, she takes the quality to be a feature of the musicians themselves, rather than a mere expression of her mental states. If she didn't, what would be the point of music criticism and reviews? The same is true for her act of going to the funeral to support her friend; the goodness of the act typically feels independent of anyone's mental states, just like reasons and oughts.

But does the goodness or badness of an object impose the same kind of demand that reasons and oughts do? There are reasons to doubt it. There are many things whose value I

[10] For a detailed argument that supports this sort of view, see Rachel Achs and Oded Na'aman (2023), where they propose that the widely held fittingness biconditional, in our example - enjoyment of x is fitting if and only if x is enjoyable - may not hold in all cases.

can recognise but which are so distant from me in time and space that judging them to be of value imposes no pull on me whatsoever: I can judge the sinking of the Titanic to be a very bad thing, without thereby feeling any demand to feel or do anything about it. Another example involves 'solitary goods'. In the context of arguing against various forms of the fitting-attitude analysis of value, Krister Bykvist has us imagine a situation in which there exist happy egrets, but no past, present or future agents (Bykvist, 2009, p. 5). Plausibly, this is a good situation, and at the very least it is a better situation than one in which there are no agents and no happy egrets. Here again we have something good – the existence of happy egrets – and yet no demands felt by any agents, since no agents exist. The phenomenology of value thus seems to be objective, but not demanding in character.

There is a further phenomenological aspect of value I think does not apply to the other categories. This is the feeling of a positive or negative valence. Typically good things *feel* good, and bad things *feel* bad. There may seem to be exceptions, as in the case of the child who hates the taste of broccoli even though it is good for her, but the bad feeling is associated with the taste, rather than the goodness of the broccoli, which if the child were in a position to apprehend, would plausibly feel good. Watching a great band perform feels great. Learning of the death of a loved one feels very bad. There is at the very least a close correlation between the apprehension of valuable and and dis-valuable things on the one side, and positive and negative affect on the other. Reasons, oughts and fittingness do not seem to be positively or negatively valenced in this way; I will show why this is so below. We are now in a position to summarise the phenomenological profiles of our four normative categories:

Category	**Demandingness**	**Objectivity**	**+/- Valence**
Reason	Yes, in different directions	Yes	No
Ought	Yes, in a single direction	Yes	No
Fit	No (?)	No	No
Good	No	Yes	Yes

X-First views are very ambitious. They seek to explain all normative categories in terms of just one. This means that X-Priority with respect to each of the other three categories must be true. I am now going to consider two popular positions - Reasons-First and Fittingness-First - and show how the phenomenological variation outlined in the table above causes problems for particular Reasons-Priority and Fittingness-Priority views.

Reasons-First is the view that ought, value and fit can all be fully grounded in reasons. Reasons-Priority with respect to value is the view that for x to be good is for there to be sufficient reason to respond to x with pro attitudes. This is the buck-passing account of value. There is good reason to think that reasons for pro attitudes cannot fully explain the phenomenology of value. Although Aisling feels an overwhelming pull to do as she ought and go to the funeral, she might still feel conflicted owing to the value of the concert she is missing out on. Because it would have been so good, Aisling may well feel a twinge of disappointment or regret. If Reason-Priority with respect to value is true, we should be able to explain this aspect of the value of the concert in terms of reasons for pro attitudes. However, Aisling is not disappointed because she has reasons to have pro attitudes in response to the concert. She is disappointed because the concert would have been a once-in-a-lifetime experience, and she can't be there. We do not miss out on reasons, we miss out on the value of potential experiences. Value thus has a positive valence that reasons cannot explain on their own.

Fittingness-First is the view that all normative categories can be explained in terms of fit. Fittingness-Priority with respect to ought fails if our phenomenological profiles are accurate. As we have seen, fittingness is either sometimes demand-like, and sometimes not, or it is never demand-like, inheriting what seems like a demand from an associated ought. If this is true, then the feeling of a categorical demand imposed by an ought would be left unexplained by fit. To take another example, Fittingness-Priority with respect to value also fails on phenomenological grounds. This is the view that for x to be good is for it to be fitting to have a pro-attitude in response to x. This fitting-attitude analysis of value is subject to the very same worry as the buck-passing account of value. The judgement that it is fitting for me to enjoy a particular performance of my favourite band does not capture the actual joy which is an essential aspect of the judgement that the performance is good. There is something about evaluative properties, namely a positive or negative valence, that neither fittingness nor reasons can fully account for, and this is because the latter two categories have no valence to speak of.

If this is on the right track, then two potential X-First views - Reasons-First and Fittingness-First - have an achilles heel: there are aspects of the phenomenology of the categories to be reduced to reasons or fit that the two candidates for X cannot explain. What's more, the same sort of argument can be run for each X-First view: given the phenomenological variation across the four categories we have been discussing, we can point out aspects of some category Y that another category X cannot fully explain. The argument from phenomenology thus constitutes a general problem for all X-First views.

I will conclude this chapter by briefly considering the most obvious line of defense the X-Firster can take against this sort of argument. Assuming it is incumbent on X-First views to account for the phenomenological profiles of the various normative categories they purport to explain, and assuming also that the phenomenological profiles represented in the table above are more or less accurate, the X-Firster can insist that her view can in fact capture the phenomenologies of the categories grounded in X. Take Reasons-First for example. One component of this view involves the grounding of values in reasons. As we have seen, the buck-passing account of value – which is the claim that for something to be good is for there to be sufficient reason to respond to the thing with a pro-attitude – is a popular way of grounding values in reasons. I have argued that the buck-passing account fails to fully explain the nature of goodness because the good is positively valenced, while reasons are not. In response, the Reasons-Firster could say the following: the buck-passing account can explain the positive valence of value, even if reasons are not positively valenced, because it contains reference to pro-attitudes, which do have a positive valence. The pro-attitudes that are a component of the buck-passing account of value can explain the phenomenology of goodness, or so the Reasons-Firster might argue.

Perhaps the Reasons-Firster will want to make this sort of move, but I think it comes at the cost of significantly watering down the view. If it is the pro-attitudes we have reasons for that explain the positive valence of goodness, then reasons are not doing any work in the explanation of this aspect of value. A naive kind of subjectivism about value, which says that x is good just in case I in fact have a pro-attitude toward x, would fully explain this phenomenological feature much more directly than Reason-First. On this front then, naive subjectivism does a better job than Reasons-First. In general, if it is the pro-attitudes referred to in the buck-passing account that gives us an explanation of the positive valence of goodness, then I have trouble seeing how reasons are First in the sense that they themselves explain all the normatively significant aspects of value.

2.5. Conclusion

Normative categories are curious things which call out for explanation. A popular method in recent years has been to explain one in terms of another. X-First is the view that all normative categories can be explained in terms of a single, master category. This essay has voiced scepticism at the very idea of an X-First view. I argued that two potential motivations for X-First views - Demystification and Qualitative Parsimony - do not in fact favour X-First over No-X-First. I then proposed an objection to X-First views based on two conditions: (1) if X is first with respect to Y then it must fully explain Y, and (2) such a full explanation must account for the phenomenology of Y, if X and Y are normative categories. The fact that each normative category has a distinctive phenomenological profile gives us reason to doubt that any one category can fully explain all the others. This is not meant to put a nail in the coffin of all X-First views. It is instead a challenge to the X-Firster to show how their favoured X does in fact ground each of the other normative categories despite the phenomenological variation among them.

Chapter 3. Fittingness: Dramatis Personae

3.1. Introduction

Feet fit shoes, rucksacks fit below the seat in front of you, and round pegs fit into nice, round holes. This physical relation of fit between two things serves as a metaphor for a normative relation between a response and its object[11]. The normative relation of fit has the following structure: ‘φ is a (un)fitting response to x’. In this chapter, I ask what can be substituted for φ. Or in other words, what are the sorts of responses that are evaluable in terms of fit? The paradigmatic case is attitudes, which can be fitting or unfitting in response to their intentional objects. Attitudes are always about something, they have intentional content. When I am angry, afraid or ashamed, I have these attitudes in response to things in the world. For instance, respectively: *that my bike has been stolen*, *the vicious bear in front of me,* and *the cruel way I behaved toward you*. These are affective attitudes, but propositional attitudes such as beliefs and intentions also have objects: I believe *that Paris is the capital of France* and I intend *to visit Notre-Dame*.

There are at least three reasons why attitudes are the current paradigm of φ in ‘φ is a fitting response to x’. The first is that attitudes are intentional, as described above, and even without yet specifying the sort of intentionality necessary for fit-evaluability, the metaphor we began with suggests fit is a relation between two things, and attitudes and their objects are clearly two very closely related entities. Secondly, it accords nicely with a widely-accepted way of elucidating the nature of the normative fittingness relation: a response is fitting when its object merits or is worthy of that response, as we saw in the previous chapter. Christopher Howard (2019, p. 216) takes this to be a straightforward definition of fittingness, while Selim Berker (2022 and 2024) places ‘being worthy of φ’ and ‘meriting φ’ alongside fittingness in an *aptic* family of normative categories, which also includes appropriateness.

The notion of an object meriting a certain response is quite clear when we think about attitudes. I can admire just about anything, but not everything merits admiration. For instance, I can admire the architecture of a soulless, grey office block, or I can admire a majestic, gothic cathedral. Intuitively, the architecture of the latter building is worthy of my

[11] I am going to use ‘response’ as a generic term for any kind of human reaction and/or engagement with the world, so that pain, fear, belief, intention, crossing the road, and protesting a government policy all count as responses to various kinds of stimuli.

admiration, while the former is not. And this is familiar for a whole host of affective attitudes: funny jokes merit amusement, doing something wrong merits shame, and the bared teeth of a mother bear merit fear. If φ is fitting whenever the thing it is a response to merits φ, then φ can be an attitude.

The third reason attitudes dominate contemporary discussions of fittingness is that a great deal of work involving the category revolves around the way in which fit might be used to explain or analyse other normative categories. The classic example is the fitting-attitude analysis of value. This is the view that x is good if and only if it is fitting to respond to x with a pro-attitude. Versions of this view have been defended by Ewing (1939 and 1947), Schroeder (2010), and McHugh and Way (2016 and 2022) among others. Theories of value like this are clearly interested in the fittingness of attitudes, rather than anything else.

In light of these considerations, my investigations of the nature of fit in the following chapters will focus for the most part on attitudes of various kinds. However, there are at least two other candidates for x in the literature: actions and bodily episodes. The first of these was arguably as central to discussions of fittingness in the 1930s and 40s as attitudes are to debates in the 21st century. Representative authors include C.D. Broad (1930) and W.D. Ross (1939), both of whom were interested in explaining the rightness of an action in terms of the fittingness of that action to a given situation. The thought that actions can be fitting or unfitting in response to situations has also been suggested by more recent writers, e.g. Philip Stratton-Lake (2022) and Selim Berker (2022), although there has been little systematic defence of this view. That bodily episodes can be fitting or unfitting has been proposed by Oded Na’aman (2022). He argues that not only can we assess whether fear fits its object, but the characteristic bodily processes that often accompany fear, such as a racing heart and sweaty palms, can also be assessed for fittingness.

This chapter has two central aims. The first is to answer the question of what kinds of responses, along with attitudes, are fit-evaluable. In the course of offering an answer, my second aim is to shed some light on the nature of fittingness. Section 3.2 sets out a series of criteria responses must meet in order to be fit-evaluable. Five of these conditions are due to Na’aman, to which I add three of my own. Section 3.3 asks whether bodily episodes meet the criteria outlined in section 3.2. I argue that Na’aman does not convincingly show that they do. Section 3.4 then asks the same question about actions. This section constitutes an argument for the claim that actions are fit-evaluable. Section 3.5 concludes.

3.2. Rationality and the Criteria of Fit-Evaluability

In the abstract to his *Philosophy Compass* paper on fittingness, Christopher Howard (2018) pointed out that the category of fittingness "remains strikingly undertheorised". I think this is still true today, although a good deal of progress is being made in the area. As we have already seen, a lot of this work has focused on what seems to be a close connection between fittingness and other normative categories with far more extensive theoretical pedigrees. Two such categories are normative reasons and rationality. How these three categories relate to one another is a matter of some debate. It is very plausible that whenever a response is fitting, there is at least some reason for that response: when admiration fits an admirable piece of architecture, I also have some reason to admire it. Some writers argue that fittingness is grounded in normative reasons for attitudes (D'Arms and Jacobson, 2000, and Schroeder, 2010), while others go in the opposite direction, arguing that reasons are essentially reducible to fittingness (McHugh and Way, 2022). It also seems likely that whenever a response is fitting, that response is rational: it is rational for me to admire a building whenever it has features that make it worthy of admiration.

The precise nature of the relationship between normative reasons and fittingness, and between rationality and fittingness, will be explored in detail in chapters 4 and 5 respectively. For present purposes, I am going to follow Na'aman on the relation between rationality and fittingness. He takes the position that "rationally evaluable items are necessarily evaluable as justified with respect to their objects", which means that "rationally evaluable items are necessarily fit-evaluable, because for an item to be fit-evaluable is for it to be evaluable as justified with respect to what it is about." (Na'aman, 2022, p. 94). In this chapter, unless otherwise stated, by rationality I will mean 'substantive rationality', according to which a response is rational to the extent that it is supported by normative reasons, as opposed to 'structural rationality', according to which you are rational to the extent that the set of attitudes you hold is coherent (Kiesewetter, 2017 and Lord, 2018). So, for Na'aman, rational evaluation in the normative sense – in the sense of the item, or response, being supported or favoured by the features of the object it is directed at – is essentially the same thing as fit evaluation. This suggests that thinking about what sorts of responses are rationally evaluable should be a fruitful approach to the question of what responses are fit-evaluable.

Na'aman (2022, pp. 82-85) provides a list of 7 features he claims to be widely agreed upon characteristics of rationally evaluable responses. For the most part, these appeal to intuitions about our evaluative practices. I am going to eliminate two of them. The first is simply the principle stated above, that if a response is rationally evaluable then it is also fit-evaluable. The other is the claim that only attitudes are fundamentally rationally evaluable. Although this claim has its defenders, (e.g. McHugh and Way, 2022b), I am going to exclude it on the grounds that it is a substantive view about the kind of thing that is rationally evaluable, rather than a schematic condition on rational evaluability. So, here are the five conditions:

(A) Rationally evaluable responses are responses to which we normally apply norms of rationality and justification.
(B) Rationally evaluable responses are responses for which we normally seek and offer reasons-for-which explanations.
(C) Rationally evaluable responses are responses we normally expect to align with the agent's judgements.
(D) Rationally evaluable responses are responses we can normally reason to.
(E) Rationally evaluable responses are normally attributable to the agent.

To lay the foundations for our investigation of what sorts of responses are fit-evaluable, it will be helpful to see, following Na'aman, the way in which the kind of responses we have assumed can be fitting, i.e. attitudes, meet these conditions. Belief is the least controversial attitude in this respect. (A) We normally apply norms of rationality and justification to beliefs: when you express your belief that it is going to rain tomorrow, I can ask for justification, which you can attempt to provide in the form of testimony from a meteorologist, for instance. I can also criticise you as irrational for believing simultaneously that it is not going to rain tomorrow, or for intending to hang your washing out to dry. (B) We normally seek and offer reasons-for-which explanations for beliefs: on what basis do you believe it is going to rain tomorrow, I ask. I believe it because the weather forecast of the national broadcaster said so, you reply. A reason-for-which explanation provides a 'motivating' reason for a response. A motivating reason is a reason *in light of which* an agent has a particular response.

(C) We normally expect beliefs to align with the agent's judgments: my belief that it is going to rain tomorrow should, if I am fully rational, be 'sensitive' to my judgments about the reasons there are for this (Scanlon, 1998, p. 20). If I accept your testimony and form the belief that it is going to rain tomorrow, I should revise this belief if I later discover that you never actually watched the weather forecast and were lying to me. (D) We can normally reason to beliefs: following Pamela Hieronymi (2006), to reason to a response is to settle a 'deliberative question' by forming that response on the basis of some considerations. I reason to the belief that it is going to rain tomorrow when I settle the question 'whether it is going to rain' by forming the belief on the basis of the fact that the weather forecast says so. (E) Beliefs are normally attributable to the agent: beliefs *belong* to the person who holds them, and they reveal something about the kind of person they are. For instance, when we learn that Kate believes in God, we learn quite a lot about Kate as a person.

I think affective attitudes also meet these conditions, although this is somewhat more controversial. Let's take admiration. (A) We normally apply norms of rationality and justification to admiration: If you tell me you admire Keir Starmer, I can ask you for justification, which you can attempt to provide by pointing out some of his admirable qualities, and I can criticise you for being irrational if you believe that he has no admirable qualities, and yet admire him nonetheless. (B) We normally seek and offer reasons-for-which explanations of admiration: A. Why do you admire the pyramid? B. Because it was built so long ago, without the aid of modern technology. This is a simple conversation illustrating the seeking and provision of a reason-for-which explanation for admiration.

(C) We normally expect admiration to align with the agent's judgments: if I admire you for the insightful paper you wrote, we would expect my admiration to cease when I discover that it was not actually you who wrote it. This norm is not belied by the phenomenon of recalcitrant emotions, most familiar from the case of fear: we often continue to be afraid even when we judge that there is nothing to be scared of. The point is that we *normally expect* one's attitudes to reflect one's judgments. (D) we can normally reason to admiration: we can settle the question 'whether to admire this piece of art' by forming the attitude on the basis of a reason, that it makes beautiful use of perspective, for instance[12]. (E) Admiration is normally attributable to the agent: my admiration for a piece of art *belongs* to me, and it reveals something about me as a person; think of what you might learn about me and my other attitudes and beliefs if you knew I admired the work of Leni Riefenstahl.

If these five features really are necessary for rational-evaluability, as Na'aman suggests, then our prediction that attitudes are rationally evaluable looks secure. To these conditions on rational evaluability I am going to add three further conditions that may or may not be essential to rationality, but are very plausibly necessary for fit-evaluability. These are:

(F) If φ is fit-evaluable, φ has an object.
(G) If φ is fit-evaluable, it must be possible for φ's object to merit or be worthy of φ.
(H) If φ is fit-evaluable, φ must have constitutive standards of correctness.

(F) and (G) express the widely accepted features of fittingness mentioned in the introduction to this chapter. (F) can be interpreted in a number of different ways. Two dimensions are especially important. Firstly, we can distinguish between a response's 'particular object' and its 'formal object', and secondly, the manner in which a response relates to its object can be glossed as the response being 'directed at' its object, or the response being 'about' its object. In both cases, I think the former, less theoretically loaded interpretations are all we need. A response's 'particular object' is the thing in the world the response is a response to. For instance, *that I was cruel to you* can be the object of my shame, and *the vicious bear in front of me* can be the object of my fear. The so-called 'formal object' of a response is the evaluative property we "implicitly ascribe" to the particular object of the response by virtue of having that response to it (Scarantino and de Sousa, 2018). Thus, the formal object of shame is *shamefulness*, for fear it is *fearsomeness*, and for belief it is *truth*. The formal objects of attitudes are what Berker refers to as

[12] This specific point is disputed by Nathaniel Sharadin (2016), and at length by Conner Schultz (2025). I respond to the latter's arguments in the next chapter.

‘thick-fittingness categories’ (2022, pp. 33-34); they are properties definable in terms of fitting attitudes: x is fearsome if and only if it is fitting to be afraid of x. Lots of different particular objects can have the same formal object: lots of different things can be fearsome. Some writers, such as Christine Tappolet (2016, p. 87), think that affective attitudes represent particular objects as instantiating certain formal objects. Specifically, Tappolet thinks that emotions are perceptions of evaluative properties - fear is a perception of fearsomeness – such that an emotion is fitting when it accurately represents its particular object as having a particular evaluative property – for fear to be fitting is for it to be a veridical perception of fearsomeness.

At this point I am going to introduce a methodological consideration that will guide the present chapter as we proceed. I want to ask which responses can be evaluated in terms of fit whilst taking on as few theoretical commitments about the nature of fittingness as possible. We should build our theory of fit around the plausible candidates for fit-evaluability, rather than deciding which responses can be fitting on the basis of a premade theory of fittingness. If we take attitudes to be representations of formal objects, i.e. fear is a perception (or other kind of representation) of fearsomeness, we have made a significant theoretical commitment, since such a view implies that fittingness is a matter of accurate representation (Tappolet, 2016, p. 88). If instead we leave the nature of attitudes open, and think of them as responding directly to particular objects – *the vicious bear in front of me,* for instance – then we avoid presupposing an accurate-representational account of fittingness.

Secondly, if we interpret a response’s object as being that which the response is ‘about’, we may be unduly setting the bar for fit-evaluability too high. To say that attitudes are *about* their particular object is felicitous: I’m annoyed about the weather, I’m sad about the ongoing war, I’m happy about your engagement etc. However, on the face of it at least, it seems quite infelicitous to say that your heart rate is about anything, or that getting on the bus is about anything. Taking the intentionality central to fit-evaluability as directedness rather than aboutness is less demanding: we can say that my sadness is directed at the ongoing war, and there is room for arguing that the crying which accompanies my sadness – a bodily episode – is directed at the war, and that my marching in protest – an action – is directed at the war as well[13]. Since both aboutness and directedness are often used to describe intentionality, we have reason to take it as the latter so as not to beg the question against those who believe that actions and/or bodily episodes are fit-evaluable.

(G) Seems right given that phrases such as x 'being worthy of φ' and 'meriting' φ are often taken to be synonymous with x being the 'fitting object' of φ, and the former two locutions bring out the key difference between a response being fitting and there merely being sufficient reason to have that response. If you hold a gun to my head and threaten to pull the trigger unless I admire you, I have very good reason for admiration. However, you do not thereby *merit* admiration. There is nothing admirable about threatening to kill someone unless they do what you say; you are not worthy of my admiration. This is to say that I can have sufficient reason to admire you, namely avoiding being shot, even though it is not fitting to admire you.[14] With respect to (H), here is Errol Lord on the fittingness of fear:

> "Fear in response to the fearsome is correct; it is also a constitutive fact about fear that this is so. It is part of what fear is that fear is correct in response to the fearsome. This helps to explain why fittingness is so important to the things that can be fitting. It is part of what they are that they are correct when they are fitting" (Lord, 2023, p. 253).

This forms the bedrock of a substantive theory of fit according to which fittingness just is constitutive correctness.[15] Part of the nature of fear is that it fits fearsome situations, just as part of the nature of chess is that it has certain rules etc. On the sort of view we are considering, φ is a fitting response to x just in case it is partly constitutive of φ that it is correct in response to x. Belief is another paradigmatic example. Belief is fitting in response to true propositions, and this can be explained in terms of the fact that it is partly constitutive of believing a proposition that the belief is correct or 'successful' whenever the

[13] In section 3.3 I suggest that bodily episodes in fact lack directedness, while section 3.4 claims that actions are directed at objects.

[14] Incentives to have certain attitudes, like the one discussed in this paragraph, are often called the 'wrong-kind of reason' for a response. It might be argued that 'right-kind reasons' are sufficient for fit: the fact that you produced some great work of art is the right kind of reason to admire you insofar as this fact makes you *worthy of* admiration.I deal with this in the next chapter, where I argue that right-kind reasons do not determine fit.

[15] For a defense of this view, see Lord (2023). For objections, see McHugh (2023).

proposition is true. Belief, it is often thought, *aims* at the truth (e.g. J. David Velleman, 2000, p. 244). I include the possibility of having constitutive standards of correctness as a condition on the kind of responses that can be evaluated for fit because, as mentioned above, I want to ask the question of this chapter without making any theoretical presuppositions about what fittingness is. By accepting (H), I have not thereby committed myself to the claim that fittingness just is constitutive correctness, I have merely allowed for the possibility of such a view, which is a live option in contemporary debates.

3.3. Are Bodily Episodes Fit-Evaluable?

Do bodily episodes meet Na'aman's criteria for rational evaluability and my further three conditions on fit-evaluability? First, the term 'bodily episode' needs to be clarified. All sorts of things fall under the umbrella of bodily episodes: a pang of hunger, a headache, goosebumps, butterflies in your stomach, sweaty palms, a racing heart beat, the warmth of the sun on your skin, and a mild itch behind your ear. Na'aman refers to the thought that such physical sensations are never rationally evaluable - and by extension never fit-evaluable - as "the common view". Intuitively, bodily episodes are just not the sorts of things we apply norms of rationality (and fit) to: we do not demand or offer justification for headaches (A), they do not look like the conclusion of a deliberative process (D), and they do not seem to reveal anything interesting about the person who has one (E).

Na'aman's aim in his (2022) is to reject the common view. However, it is important to recognise what this does not entail. Rejecting the common view does not commit one to the claim that all bodily episodes are always rationally evaluable. Na'aman does not want to argue that random headaches can be more or less rational, fitting or unfitting. The claim is that some bodily episodes are sometimes rational and fit-evaluable. Specifically, the bodily components of fit-evaluable attitudes are evaluable for fit. For instance, fear is uncontroversially fit-evaluable. Na'aman's claim is that the bodily reactions that commonly accompany fear, such as an increased heart rate, sweaty palms, and trembling extremities, can be fitting or unfitting. Here is the first pillar of the common view that Na'aman wants to attack, which focuses on the five criteria of rational-evaluability:

> **R1**. Paradigmatic examples of rationally-evaluable responses such as attitudes have a set of characteristics not shared by bodily episodes such as headaches and heart rates.

Na'aman argues that R1 is false: certain bodily episodes do have the relevant characteristics. For some of these, attributing them to bodily episodes is very reasonable. For others, it is something of a stretch. Let's take each condition in turn, and stick with Na'aman's favoured example of fear, which often involves an increased heart rate and trembling extremities. (A) We normally apply norms of rationality and justification to the bodily manifestations of attitudes. Suppose it is irrational to be afraid of a fruit fly. If you are terrified of the tiny insect flying around the room, I can criticise you for being irrational. Your fear is unjustified. Furthermore, if you are visibly trembling with fear, this bodily episode is equally criticisable, and for the very same reason as your fear, namely that the tiny fruit fly is not remotely dangerous.

Next, (B) we normally seek and offer reasons-for-which explanations for the bodily manifestations of attitudes: just as we can offer reasons in light of which we feel afraid, such as the fact there is a vicious-looking bear in front of me, our associated bodily reactions can be explained in just the same way: my heart rate shot up in light of the fact that there was a vicious-looking bear in front of me. I think such reason-for-which explanations are possible. However, there is an important disanalogy here rooted in Na'aman's appeal to our ordinary evaluative practices: although there may be reasons-for-which explanations of the bodily manifestations of emotions like fear, do we 'normally seek and offer' such explanations? If I see you displaying all the physical signs of fear – you are trembling and breathing in a way that indicates a racing heart – and I have good reason to believe that you are frightened by something, then ordinarily I would simply infer fear and ask 'why are you afraid?', rather than 'why are you trembling'?

Of course, if I am really not sure what is going on and do not make this inference, I might ask 'why are you trembling?' A sufficient answer here would be 'because I am afraid'. But this is not a reasons-for-which explanation: you do not first consider the fact that you are afraid, and then start trembling in light of this fact. It is more like a constitutive explanation - the trembling is a part of your fear, and that is why it's happening.

Furthermore, after receiving this constitutive explanation of your trembling, I may then go and seek a reason-for-which explanation with a question like ‘why are you afraid?’, to which you might give a reason-for-which explanation like ‘because I thought I saw a ghost’. What I typically seek and you typically offer is a reason-for-which explanation of attitudes, rather than the bodily episodes that accompany them. The point is that if we tie the question of rational-evaluability to our typical evaluative practices, as Na’aman does, it speaks against the rational-evaluability of bodily episodes if we do not ordinarily seek reasons-for-which explanations of them.

To my mind, the next condition is uncontroversially met, (C) we normally expect the bodily manifestations of attitudes to align with the agent’s judgments. We noticed that with attitudes, recalcitrance means that our responses do not always align with our judgments - I can continue to be afraid even though I judge that there is nothing to be afraid of. However, this does not impugn the fact that we ‘normally expect’ attitudes to align with our judgments. In the case of the bodily manifestations of emotions, I suspect that recalcitrance is less pronounced, and they are thus even more likely to align with our judgments.

If my heart rate rockets upon encountering a bear in the woods, we would expect it to fall back down when I judge that there is nothing to fear - the bear turns out to be my friend in a costume. Of course, we can have recalcitrant bodily episodes in the same way that we can have recalcitrant emotions: my heart might continue to race despite the fact that I have judged that there is no immediate danger. However, once we have recognised that there is no danger, though my fear may persist, my heart rate will usually die down significantly, if not to a resting rate. If my heart rate does not slow down upon discovering that what I took to be an angry bear was in fact my friend in a costume, we would think that something has gone seriously wrong, and if it continues to persist even longer, we would be justified in seeking medical attention. So, we strongly expect bodily episodes to align with our judgments.

The relevance of (D) I think is the weakest point in Na’aman’s case for the rational-evaluability of bodily episodes. I will examine it in detail below. First, consider (E) the bodily manifestations of attitudes are normally attributable to the agent. On one level this seems right. Imagining an agent's response to an encounter with a bear in the woods, Na'aman writes:

> "We are inclined to view your racing heart in this case as an expression of your evaluative point of view and to attribute it to you in the same way we attribute your fear to you. Just like your fear, so your perspiration, shivers, and racing heart express your understanding and appreciation of your predicament" (Na'aman, 2022, p. 90).

Although bodily episodes are not mental states, they are clearly the reactions *of* an agent, and they reveal something about her evaluative perspective. However, this 'something' is merely a *clue*. In a robust sense of the term 'attributable', a response is attributable to the agent insofar as we can read directly off it the agent's appreciation of her predicament, as well as what this says about her as a person. Bodily episodes never meet this threshold on their own. This is because knowing what specific attitude a bodily episode expresses is indispensable to learning the relevant information about the agent, and so the bodily manifestations of attitudes are at best attributable derivatively. Take tears for example. This bodily episode can express at least two attitudes: tears of sadness and tears of joy. Tears may also express no attitude at all, as when they result from hay fever or chopping onions. If I see someone crying, I cannot accurately infer anything about her evaluative point of view unless I know which attitude, if any, her tears express. It would therefore be an exaggeration to say that crying is attributable to her in the robust sense described above. If I ask this person why she is crying, and she tells me it's because she saw a dead badger on the motorway earlier that day, I can be quite confident she is crying tears of sadness, and I have learned that she feels above-average levels of empathy for animals. This shows that sadness is robustly attributable to the agent in a way that tears are not.

Now for (D). Recall that this condition reads: rationally evaluable responses are responses we can normally reason to. In support of his claim that bodily episodes meet this condition, Na'aman simply offers the following:

> "it is arguable that one's racing heart in response to the bear is *also* an embodiment of one's answer to the question of whether the bear poses a danger. If that is so, then

> you can reason to your racing heart in the very same way that you reason to your fear." (Na'aman, 2022, pp. 89-90).

This formulation exposes a flaw in Na'aman's position; even if we can settle deliberative questions by forming bodily episodes, it is not the case that you can reason to your racing heart *in the very same way* that you can reason to your fear. There is a crucial discrepancy between the two rational processes: when we settle a deliberative question by forming an attitude on the basis of certain premises, the attitude risks being *unjustified* when the premises are purely causal. In contrast, the settling of a deliberative question by forming bodily episodes, assuming such a process is possible, can always be done on the basis of merely causal considerations without sacrificing rational justification.

Take the attitude of belief. Suppose a child called Sally believes the following proposition: 'that stick is bent'. Sally is standing next to a lake with her father, who has dipped a small, straight branch into the clear water. If we were to ask Sally why she believes this, she might say something like 'well look at it, it's bent!'. Sally has settled the question 'whether the stick is bent' by forming the belief that it is on the basis of her visual experience. Importantly, this kind of reason is an *explanatory* one - it explains the cause of her possession of the belief.

Now, it is clear from this example that reasoning to beliefs on the basis of purely causal premises can lead to the formation of unjustified responses. While Sally may be justified in believing as she does, the belief itself is not adequately supported by reasons. In this particular case, the cause of the attitude, a visual experience, is liable to be deceptive. This is because refraction makes us see things illusively. The cause of Sally's belief about the shape of the stick thus cannot justify it. This aspect of the rational evaluability of beliefs is well-documented. In a discussion of justification from his work on theories of knowledge, Keith Lehrer (1990) highlights the same distinction between the causal explanation of belief and the normative evaluation of it. He writes: "what a person originally believes as a result of prejudice may later be accepted on the basis of scientific evidence." This entails that it is a mistake to think "it is what originates a belief that converts it into a justified belief and knowledge. This is, in effect, to confuse the *reason* a person has for believing something with the *cause* of his believing it. The confusion is such a common one that we might name it the *casual fallacy"* (Lehrer, 1990, p. 168-169).

To see how this works, consider another example. Suppose I believe that Brendan Behan was an alcoholic. Suppose also that I formed this belief immediately upon learning of his existence, after a friend introduced me to the man and his work, including the fact that he was a 20th century Irish writer and Republican. Now imagine that this belief is based on my unreflective prejudice, grounded in my blind acceptance of certain stereotypes and assumptions about Irish writers. As it happens, Behan was an alcoholic, who is said to have described himself as 'a drinker with a writing problem'. Was my belief justified? If it is based on nothing but its cause, then the answer is no: it is clearly fallacious to take the cause of my belief – a stereotype about Irish literary figures – to provide normative support for the belief that one such figure was an alcoholic. The kinds of reasons which do support this belief involve facts about the writer's drinking habits, and perhaps trustworthy testimony. It is thus a feature of reasoning to rationally evaluable attitudes like belief that the process is susceptible to this causal fallacy.

A question we might then ask is this: is the process of reasoning to bodily episodes also susceptible to this causal fallacy? I want to suggest that it is not, and that this is because if we assume that it is sensible to talk about the justification of our physical reactions, they can always be fully justified solely with reference to causal explanations. If the sort of *reasoning to* a response which indicates that the response is rationally evaluable involves the possibility of committing the causal fallacy, then the absence of such a possibility in the case of 'reasoning to' bodily episodes suggests that such responses may not be rationally evaluable.

Let's return to Na'aman's case of an encounter with a bear in the woods. I can justify my fear of the bear to an interlocutor by recounting how I reasoned to it: I felt afraid because the bear was big and vicious looking, it had sharp claws and powerful jaws and it appeared to pose a threat to my life. One might think, along with Na'aman, that my racing heart and sweaty palms can be justified in just the same way; the bodily reactions which accompanied my fear are normatively supported by reasons involving the viciousness and threat of the bear. However, contrast this with a variant case that Na'aman introduces in a slightly different context, but which I think illustrates the distinction I want to draw: "a bear approaches you, unseen, and steps on a button that triggers a mechanism that causes you to see a hologram of a bear, which in turn causes your heart to race." (Na'aman, 2022, p. 88).

In this case, it is plausible that fear is not a fitting response. Although I might feel fear and assert a proposition like 'I am afraid', fear is not supported by the relevant facts; as a hologram, the bear is neither vicious nor a threat to my life. Fear in this situation seems unfitting, if we think that fittingness is determined by the objective facts alone, rather than what the agent thinks or believes about her situation. This distinction will be discussed in detail in chapter 5. While recounting the way I reasoned to my fear, an interlocutor might reasonably accuse me of committing the causal fallacy if I stipulate that fear was justified simply because the hologram scared me. The fact that the hologram caused me to feel afraid is not enough to secure substantive rational justification for the emotion. However, if for some reason my interlocutor challenges me to justify my increased heart rate, it seems enough to simply state the fact that the hologram *made* it jump. The bodily reaction is justified by the cause. What we see is a divergence between the processes of reasoning to fear and reasoning to an increased heart rate. With respect to the former, one must be wary of the causal fallacy, as for the latter, there is no such trap to avoid. Considering a couple of non-attitudinal examples might help to clarify the point:

1. 'Why are you hungry?' - 'because I skipped breakfast'
2. 'Why does your foot hurt?' - 'because I stood on a piece of lego'

In both of these cases, I cannot think of any further reasons which might be required to justify the empty belly and the sore foot. The facts of the matter, which causally explain the bodily sensations, exhaust the range of reasons relevant to their justification. With respect to fear, it makes sense to ask, 'well I know the hologram scared you, but was fear an appropriate response?'. However, asking the same of a sore foot would sound odd: 'well, I know that standing on the piece of lego hurt, but was pain an appropriate response?'. The answer is yes, of course it was, because when I stood on the piece of lego, it hurt.

This difference tracks the fact that while the rational justification of the former response is susceptible to the causal fallacy, the latter is not. This reference to non-attitudinal bodily episodes might seem slightly unfair, as Na'aman rightly recognises that physical sensations are not always evaluable for fit. In fact, on his view, it is only insofar as a bodily episode can be understood as part of an attitude which itself is intentionally related to an object that the bodily episode is fit-evaluable. However, I appealed to these examples only to illustrate the general distinction between the justification of attitudes and bodily episodes, which I believe holds true even if we consider bodily episodes that are components of fit-evaluable attitudes:

3. ‘Why did your heart rate shoot up?’ - ‘because I saw a fly.’

Here your heart rate is fully justified by the mere fact that the fly caused it to shoot up, even though the fear of which your heart rate is one component is not itself justified because the cause alone does not suffice to justify the attitude.

To sum up the degree to which bodily episodes meet Na’aman’s 5 conditions on rational evaluability: (A) and (C) are quite plausibly met, we have reason to doubt whether (B) and (E) are fully met, and bodily episodes do not satisfy (D), at least insofar as reasoning to rationally evaluable responses entails susceptibility to Lehrer's causal fallacy, and reasoning to bodily episodes is not so susceptible. Na’aman also rejects a second pillar of the common view:

> **R2**. Paradigmatic examples of rationally evaluable responses such as attitudes are necessarily intentional; they are directed toward some object. Bodily episodes are not intentional, therefore they are not rationally evaluable.

This speaks directly to my first condition on fit-evaluability: (F) if φ is fit-evaluable, φ has an object. Intentionality refers to the way in which a response is directed at a particular object, and the response is fitting when that object has features which merit the response. So, fear might be directed at a bear, or it might be directed at a fruit fly. Intuitively, in the first case fear is fitting, while in the second case it is not. Na’aman recognises the fact that physical, bodily states are not always intentional. He gives the example of taking a pill which increases your heart rate. In this case, the physical reaction is not directed at anything, it simply happens to you. However, he insists that bodily episodes which are part of intentional attitudes are themselves intentional. He notes that an emotion is a complex of “patterns of sensations, imaginings, thoughts, and motivations” (Na’aman, 2022, p.97), including bodily episodes such as an increased heart rate in the case of fear, and goes on to assert that "emotion-patterns inherit their fittingness from the fittingness of the narratives in which they are embedded" (Na'aman, 2022, p. 98).

Working with a generic notion of a narrative as a representation of a situation – the story an agent tells which reflects her understanding of what is going on – Na'aman points out how "narratives can misrepresent in various ways: they might distort facts and causal connections, fail to note relevant information, etc." (Na'aman, 2022, p. 99). A fitting narrative is one that does not misrepresent; it is a representation of the situation by the agent which gets things right. This gives us the *narrative account of fit-evaluability*. For the kinds of physical sensations we are scrutinising, "bodily episodes and sensations can have intentional objects and be fit-evaluable when and because they are explained as elements of fit-evaluable emotional reactions by the agent's fit-evaluable narrative representation of the situation." (Na'aman, 2022, p. 100). The best way to make sense of this is to consider Na'aman's example:

> "Stuck in traffic on my way home from work, I scratch my chin. A fellow driver mistakes my hand movement for an insult and responds with anger. I cannot hear him but I see his facial expressions and hand movements through the windshield. If I didn't know better, I might think he is having some kind of seizure. However, given my understanding of the story he must be telling himself, his frantic movements strike me as elements of anger. Thus, the fact that he represents the situation as he does explains his sensations and bodily episodes as elements of anger. Moreover, his bodily episodes and sensations have intentionality and are fit-evaluable because they are explained as such by his fit-evaluable narrative, according to which I intentionally insulted him" (Na'aman, 2022, p. 100-101)

If I understand the view being proposed, the three fit-evaluable elements share a single object: the narrative the angry man tells himself is about my 'rude' gesture, which is what his anger is also about, and it is in virtue of these facts that the bodily episodes which express his anger are about my gesture too. Since these three elements misrepresent what they are about, they are all unfitting. We thus have an explanation for why an increased heart rate caused by the effects of a pill has no intentional object, but an increased heart rate as a component of anger does: the former does not form part of a fit-evaluable narrative representation of the world, while the latter does. There are a number of reasons to resist Na'aman's account of the intentionality of bodily episodes here. The first involves the methodological consideration introduced in section 3.2. As I have been attempting to do throughout this chapter, it is worth asking what sorts of responses are evaluable for fit without taking on too many theoretical commitments about the nature of fittingness. The claim that fittingness is fundamentally a matter of accurate narrative representation involves a substantial theoretical commitment. Perhaps narrative representation is essential to fittingness, but perhaps it is not. Either way, to explain the intentionality of potential candidates for fit-evaluability by appealing to a narrative account of fittingness is in a sense to put the cart before the horse.

Suppose we set this worry aside and allow substantive views of fittingness to colour our approach to the question of fit-evaluability, do we have good reason to accept Na'aman's account? I think we have good reasons to reject it. It is unclear to me how this narrative account directly addresses the issue of intentionality for which it was introduced. It is true that on the interpretation of intentionality that stresses 'aboutness', fit as accurate narrative representation is well placed. As we saw in the example above, the angry man is telling himself a story about my apparently rude gesture, and insofar as they form part of this narrative, both his anger and his bodily expressions of anger are about my gesture too.

However, if we think of intentionality as 'directedness', the metaphor breaks down. Are narratives directed at anything? One way they might be is in a didactic sense, or as entertainment: Sally tells Tom a story in order to teach or entertain him; the story is directed at Tom. But clearly Tom is not the object of the story in the same way that my gesture is the object of the driver's anger. My gesture may be the object of the driver's narrative, but only in the sense that the narrative is *about* my gesture. It would be strange to say that the story is *directed* at my gesture. Narratives do not have objects in the directedness sense. Bodily episodes therefore cannot inherit directedness from the directedness of narratives, since the latter does not exist. Furthermore, the claim that bodily episodes have objects in the directedness sense relevant for fit-evaluability is implausible in its own right: while the driver is angry *at* me and my gesture, surely his heart is not racing *at* me.

In light of these considerations, I am not convinced by Na'aman's effort to show that bodily episodes satisfy (F), even if we go for his narrative representation account. Although he does not explicitly discuss (G) and (H), I am going to conclude this section by briefly discussing whether bodily episodes can meet these two conditions. Staying with anger and the increased heart rate which often accompanies it, satisfying (G) would entail: it is possible for the object of an increased heart rate to be worthy of the bodily episode. As we have seen, it is not obvious that a bodily episode like an increased heart rate has an object, especially if having an object entails that the response be directed at something, and given that 'x merits φ' implies that x is the intentional object of φ, then φ cannot be a bodily episode.

However, if we accept something like Na'aman's narrative account, such that bodily episodes have the same object as the attitudes they partly constitute, might there be a sense in which this shared object can merit the bodily episodes just as it can merit the attitude? I still think it would be strange to say that a rude gesture *merits* a racing heart rate. One reason for this is that being merited seems to imply a cognitive element. It implies *recognition*. When φ is merited by x, φ-ing must involve the recognition of the features of x in virtue of which x merits the response: when a piece of art merits admiration, to admire it fittingly is to recognise its admirable features. Although it is debatable whether affective attitudes really are cognitive in this way – my conditions on fit-evaluability, along with the assumption that all attitudes are fit-evaluable, presupposes that they are – I do not think even Na'aman would argue that bodily episodes involve cognitive recognition; they are purely non-cognitive events. Sweaty palms may be indirectly caused by cognitive processes, like the perception and recognition of impending doom, but they are not themselves a cognitive process.

Finally, according to (H) bodily episodes must have constitutive standards of correctness. This looks more promising. One way to think about the constitutive standards of bodily episodes is in terms of their proper functioning. A normal, resting heart rate, for instance, is between 60 and 100 beats per minute. In other words, it is constitutive of the nature of the human heart that it is working correctly when the rate at which it beats falls within a certain range. This is enough for heart rates at least to satisfy (H), since all this condition requires is that fit-evaluable responses have constitutive standards of correctness.

There is a further question though: is it part of the proper functioning of the heart that the rate at which it beats increases when certain emotions are fitting? On Na'aman's view, when a racing heart is a component of anger, the heart rate inherits its object from the emotion, and so if anger is fitting in response to its object, then so is the increased heart rate. This entails that if it is constitutive of anger that it is correct in response to an offensive gesture, then it must be constitutive of the human heart beat that it correctly increases in response to an offensive gesture. However, whether it is a part of the proper functioning of the heart that the bpm should increase whenever anger is fitting is to a large extent an empirical question, and some recent neuroscientific research suggests that the heart may not in fact accelerate during *typical* bouts of anger (2019, Wu et al, p. 5). If this is right, then although it is part of the 'proper functioning' of anger that it is felt in response to offensive gestures, it may not be part of the proper functioning of the heart that the rate at which it beats increases in response to offensive gestures. Whatever the case, the philosopher in her armchair cannot simply help herself to claims about the proper functioning of biological responses, meaning it is unclear whether the constitutive standards of bodily episodes track fittingness, as they plausibly do in the case of attitudes. So, even if bodily episodes do have constitutive standards of correctness which can be understood in terms of proper functioning, these standards may have nothing to do with the fittingness of the attitudes they commonly accompany.

3.4. Are Actions Fit-Evaluable?

We have just seen that while there may be some reason to think that bodily episodes are fit-evaluable, there are also lots of reasons to doubt it. What about actions? Do actions meet our criteria for rational-evaluability and fit-evaluability? I will not offer any theory of what actions are, or how to differentiate them from attitudes and bodily episodes. I will simply proceed with an intuitive notion of what an action is, and a minimal description: actions are causally efficacious interventions in the world. Actions meet our initial five conditions for rational-evaluability even more comfortably than attitudes. Let's consider them very briefly in turn. The action I am going to discuss is this: marching in protest against your country's involvement in an ongoing war. This is quite a complex action: it is often collective, symbolic, and includes various subactions among other things. I am going to discuss the potential fittingness of protesting rather than a more simple action – like turning on the lights – because complex actions like protesting are of more obvious normative significance than isolated simple actions like flicking a light switch.

(A) We normally apply norms of rationality and justification to protesting war: when I tell you that I am heading out to join the anti-war demonstration in the city centre, you can demand reasons and justifications. Is it going to achieve anything? What about our right to self-defense? These are the sorts of questions that force me to rationalise and justify my action. (B) We normally seek and offer reasons-for-which explanations for protesting war: just as I can attempt to rationalise and justify my protesting, so you can seek and I can offer reasons in light of which I protest; I can give you my motivation: our government is supporting the murder of innocent children, and this is impossible to justify. This is the reason for which I am protesting.

(C) We normally expect protesting war to align with the agent's judgements: if someone convinces me that the deaths of these children is in some way justifiable in wartime, other things equal, we should expect that I would stop protesting. In other words, if I judge that there is no reason to protest, we would expect that I would not protest. And this *expectation* remains despite the fact that I can (and perhaps often do) act irrationally. (D) We can normally reason to protesting war: 'whether to protest' is a deliberative question I can settle by joining the demonstration on the basis of certain considerations. (E) Protesting war is normally attributable to the agent: my protesting my country's involvement in a war reveals my evaluative perspective from which you can infer important things about who I am. You can see that I am opposed to the war, and that I am prepared to try to do something about it.

So, actions are clearly rationally evaluable, according to Na'aman's criteria, which gives us some indication that they are also fit-evaluable. The question now is whether they meet our somewhat stricter criteria of fit-evaluability? I think there is a case to be made that they do. The first thing to notice is that actions can feature in the kinds of fittingness biconditionals that provide definitions of thick-fittingness properties, or as I will sometimes call them in this section, evaluative properties. For attitudes, examples include:

1. X is admirable iff it is fitting to admire x

2. X is funny iff it is fitting to be amused by x
3. X is fearsome iff it is fitting to fear x.

For actions, a small selection includes:

4. X is 'protestable' iff it is fitting to protest x.
5. X is punishable iff it is fitting to be punished for x.
6. X is useful iff it is fitting to use x (for a given purpose)
7. X is pursuitworthy iff it is fitting to pursue x[16].

On the face of it, there can be features of a government policy that make it fitting to protest just as there can be features of some other policy that make it fitting to admire. This similarity gives us prima facie reason to treat actions as fit-evaluable. It is important however to notice, as Berker does, that not all able/ible adjectives connected to actions are definable in terms of the fittingness of those actions. Two examples:

8. X is visible iff it is possible to see x
9. X is stoppable iff it is possible to stop x.

For something to be visible or stoppable is not for it to be worthy of being seen or stopped, but rather that it *can* be seen or stopped. And then there are some evaluative properties which admit of both fittingness and possibility analyses. Again following Berker (2022, p. 34):

10. X is printable iff it is fitting to print x
11. X is printable iff it is possible to print x.

Where P is a particular sentence, it might be that we are physically able to print P insofar as we have a printer at our disposal, which entails 11, or it might be that the content of P does not violate certain standards set by the publisher and/or society at large, entailing 10. I think this second analysis is the intended meaning behind the famous *New York Times* slogan: "all the news that's fit to print", and it is especially clear in the negative - a sentence might

[16] The pursuitworthiness of scientific research programmes is an area of ongoing debate in the philosophy of science; see for example DiMarco and Khalifa (2022) and Fleisher (2022).

be unprintable, meaning it is unfitting to print it, perhaps insofar as it refers to some taboo subject. Which sorts of actions are fit-evaluable, and when, is thus a matter of some complexity. We therefore ought to investigate whether and how actions meet our three conditions on fit-evaluability. Here is the first: (F) if φ is fit-evaluable, φ has an object. As we have seen, for a response to meet this condition, it must be intentional. In section 3.2, we distinguished between the 'aboutness' and 'directedness' interpretations of intentionality. We went for the latter, more coarse-grained interpretation so as not to beg the question against the fit-evaluability of actions, many of which are not obviously *about* anything: having a drink does not seem to be about something; it's not about the water in my glass, nor is it about my thirst, although both the water and my thirst are plausibly objects at which my drinking is *directed*.[17] This reveals two different ways to think about the directedness of actions:

1. Actions are directed at some particular thing in the world (drinking is directed at the water in my glass).
2. Actions are directed at some goal or end (drinking is directed at quenching my thirst).

Consider protesting again. When I protest a government policy, this action is directed at a particular thing in the world, namely the government policy. The policy is quite literally, as opposed to metaphorically, the object of my action insofar as it appears as the object of the verb in the sentence 'I am protesting the government policy'. The other object of my protest is my aim or goal in protesting: my protest may be directed at causing the government to reverse course and abandon the policy. So, not only do actions have objects, meaning they meet condition (F), they appear to have two. This invites a question: which of the two objects is the one against which we measure the fittingness of an action? Interestingly, depending on which object we choose, we might get different results. It is conceivable that a policy can be protestable, meaning fitting to protest, because it will directly result in the impoverishment of millions of people, while it is unfitting for me to

[17] It is worth noting that some actions can be 'about' things: I can *talk about* x, and I can *protest about* x. This indicates the sense in which aboutness is a more fine-grained gloss on intentionality than directedness: to talk about or protest about x is to represent x in some way, but not all actions, including drinking water, are obviously representational.

protest in relation to my goal, perhaps because marching in the streets will simply make the government strengthen their resolve.

This potentially mirrors the connection already discussed between fittingness and rationality: we can think of the fittingness of protesting in response to the policy considered in itself as ‘substantive fittingness’, while the fittingness of protesting as a means to my end – bringing about a government U-turn – as a kind of ‘structural fittingness’. That is, for my act of protesting to be substantively fitting is for it to be merited by the features of the particular object it is directed at, like a government policy, just as protesting is substantively rational when the policy provides reasons to protest against it. On the other hand, for my act of protest to be structurally fitting would be for it to be merited by my aims, like getting the government to capitulate, just as protesting is structurally rational if it displays means-end coherence.

If the implication that there are in fact two kinds of fittingness is troubling, and if the idea of an action being fitting solely in response to one’s aims has echoes of the wrong-kind of reason problem – a kind of ‘structural fittingness’ may entail that if I have the goal of becoming a notorious mass murder, then it is fitting for me to kill 10 people, but unfitting to kill zero – we can turn to C.D. Broad for a solution. Writing almost a hundred years ago, Broad gave an account of the object of an action which promises to unify the two dimensions just distinguished. He says:

> “Whenever a man is called upon either to act or to abstain from action he is in presence of a highly complex situation, composed of pre-existing persons, institutions and things, in various relations to each other and to himself. Let us call this the "initial phase". Whether he acts or abstains from action this phase will be succeeded by others. The initial phase, together with its subsequent developments, may be called a ‘total course of events’” (Broad, 1930, p. 218).

He goes on to say that “fittingness or unfittingness is a direct ethical relation between an action or an emotion and the total course of events in which it takes place” (Broad, 1930 p. 219)[18]. Presumably, among the elements of the ‘total course of events’ we can include both

[18] Broad and his contemporaries conceived of fit as an ethical or moral category, whereas nowadays, partly under the influence of D’Arms and Jacobson’s ‘moralistic fallacy’ (2000b), fit is not typically understood to be exclusively ethical, given the thought that immoral responses can be fitting: amusement can be fitting in

the particular thing a given action is directed at as well as the action's aim or goals. On this sort of view, the complex situation which is the object of my act of protest is partly constituted by the government policy and partly constituted by my aim of getting the government to reverse course, among other things. My protest is fitting, if it is fitting, when it is merited by all the variables that make up the total course of events into which the action is an intervention.

Since actions evidently have objects, we can ask whether these objects can merit or be worthy of their actions. I think they can, thus satisfying (G). This is especially clear with respect to the object that determines what I have tentatively called substantive fittingness. It is intuitively plausible that a government policy which will impoverish millions *merits* protest; this dreadful aspect of the policy makes it a worthy object of protest. In section 3.3, we doubted whether bodily episodes can really be merited (assuming they have objects) on the grounds that in order for a response to be merited, that response must involve a kind of cognitive recognition of the features of the object that make it merited. Bodily episodes do not involve such recognition. Actions do. When I decide to protest a particular government policy, and I join a march through the streets of the capital city, included as part of this complex action is a recognition of the injustice of the policy, in a similar way to how the anger I feel on my way to the protest involves the recognition of the injustice.

Finally, do actions have constitutive standards of correctness (H)? Well, there is a venerable tradition within metaethics which seeks to explain normative claims about actions in exactly these terms (see for example J. David Velleman, 2000 and Christine Korsgaard, 2009). The guiding thought for constitutivist accounts of practical normativity is that action itself, considered in the abstract, has some constitutive aim which generates standards of correctness or success conditions (Katsafanas, 2018, p. 369). What exactly the constitutive aim of action is will be a complicated issue: for Velleman, it is something to do with 'integrative knowledge of what we are doing' (Velleman, 2004, p. 234), such that normative claims about actions are grounded in how conducive they are to this integrative knowledge while for Korsgaard the aim of action is 'self-constitution', such that "what makes actions good or bad is how well they constitute you" (Korsgaard, 2009, p. 25).

---

response to an offensive joke, even though it would be morally wrong to be amused. I will bracket the issue of whether or not fit is an ethical relation as I do not think it affects the present issue.

Even if we reach a settled view about what exactly constitutes action in the way that is relevant to grounding normative claims, writers such as David Enoch (2006) have argued that any version of constitutivism as a metanormative theory is bound to fail. This is because, as Enoch sees it, constitutivism cannot solve one of the main problems it was designed to address, namely answering the sceptical 'why be moral?' question. If you tell the moral sceptic that she is not really an agent unless she behaves in way x, where behaving this way is constitutive of agency, which in turn is what grounds normative claims about action, then the sceptic can continue to deny that she has any reason to be moral and go on behaving in way y while accepting that she is not an agent, but a *schmagent.*

How this quite intricate debate between constitutivists and their opponents pans out need not concern us too much. For our purposes, all we need to see is that actions are plausibly the sorts of things that can have constitutive standards of correctness or success, and considering particular action types which have been used to motivate the constitutivist strategy in metaethics can help with this. Once again, playing chess is the classic example: part of what it is to play chess, to perform the action, is to aim at checkmating your opponent. As Katsafanas (2018, p. 368) points out, it is of course possible to 'play at playing' chess while aiming to lose, perhaps as a way of cheering your opponent up, but in such a case you are not really doing the simple action *playing chess*, but a rather more complex action such as *playing chess with the intention of losing to cheer someone up*. It is in virtue of being partly constituted by the aim of checkmating your opponent that playing chess has constitutive standards of correctness. Here is a plausible principle offered by Katsafanas:

> "**Success**: If X aims at G, then G is a standard of success for X, such that G generates normative reasons for action" (Katsafanas, 2018, p. 369).

So, the standard of success for playing chess is determined by the goal of checkmating one's opponent, and we can make normative judgments about one's *playing chess* insofar as what she does is conducive to achieving this end: some move can be better than another, and you can have reason to move this piece here, but not to move this other piece there, in virtue of the degree to which these subactions conform to the constitutive standards of chess playing. To return to the case of protesting, this action plausibly has a constitutive aim too: what it is to protest is to aim at bringing about some change with respect to the thing you are protesting about. If we adopt a constitutivist account of fittingness, then some token act of protest would be fitting insofar as it is conducive to this constitutive aim. So, actions can have constitutive standards of success or correctness, just like attitudes, and if the possibility of telling such a constitutivist story about the nature of fit imposes a condition on what counts as a fit-evaluable response, then it looks as though actions are to this extent fit-evaluable.

To sum up the results of this section, where do actions stand with respect to our eight conditions for fit-evaluability? There are few who would deny that actions are the sorts of things that can be more or less rational, and so we ticked off Na'aman's five conditions on rational-evaluability without much fuss, giving us prima facie reason to believe that actions are fit-evaluable when we grant the close connection between rationality and fittingness. We then considered three further conditions on fit-evaluability, and I have argued that actions meet each of them.

3.5. Conclusion

In this chapter, we explored the nature of fittingness in some detail by asking what sorts of responses are evaluable in terms of this normative category. Attitudes were taken as the paradigm of fit-evaluable responses, and we saw why this should be so by showing how they meet a series of conditions that plausibly must be met by fit-evaluable responses. Five of these conditions were derived from the close connection between fittingness and rationality: if a response is fitting, then it is also rational, and attitudes arguably satisfy the five criteria for rational evaluability proposed by Na'aman – (A) we normally apply norms of rationality and justification to attitudes, (B) we normally seek reasons-for-which explanations for them, (C) we expect them to align with our judgments, (D) we can reason to them, and (E) they are attributable to the agent. Attitudes also meet three additional conditions on fit-evaluability: (F) they are directed at objects, (G) these objects can merit or be worthy of them, and (H) they have constitutive standards of correctness.

We then asked whether bodily episodes and actions, two further candidates for fit-evaluability, meet these eight conditions. The bulk of the chapter attempted to show how

Na’aman’s defense of the claim that bodily episodes are fit-evaluable leaves much to be desired. Although (A) and (C) plausibly apply to an increased heart rate which may accompany a bout of anger, there are reasons to doubt that your heart rate meets conditions (B), (D) and (E). Furthermore, while your heart rate does have constitutive standards of correctness (even though it is unclear whether this standard tracks fittingness), it is difficult to accept that bodily episodes meet conditions (G) and (H) without taking on unwarranted assumptions about the nature of fit.

The final section of the chapter applied the test to actions, responses which are often assumed to be fit-evaluable, but rarely argued to be so. We saw that, unsurprisingly, actions satisfy (A)-(E). I then argued that there is good reason to think that they satisfy (F)-(H) as well: actions, such as protesting a government policy, plausibly has an object which can merit that action and be assessed according to constitutive standards of correctness. If all this is on the right track, we have an answer to the question we began with: attitudes and actions are fit-evaluable, bodily episodes are not.

Chapter 4. Why Fit-Making Considerations Are Not Normative Reasons

4.1. Introduction

An attitude is fitting when its object merits that attitude; for example, admiration is fitting when the object of admiration has features which are worthy of that emotion. But what determines whether an object merits a certain response? One answer is that normative reasons do[19]. The view that fittingness is ultimately to be understood in terms of reasons has been defended by Mark Schroeder (2010). Normative reasons are considerations that count in favour of a response. For instance, you might think that it is fitting to admire Rory McIlroy, and that this is because there are reasons which count in favour of admiring him: he plays beautiful golf, and he recently achieved a career grand slam. However, there can be reasons to admire a person that do not seem to bear on whether that person merits admiration. If someone threatens to kill me unless I admire McIlroy, this looks like a very good reason to admire him, but not the kind of consideration that makes it the case that he merits admiration. Incentives for attitudes such as these have come to be known as the 'wrong kind of reasons'. In contrast, the 'right kind of reasons' for an attitude are those that intuitively do make that attitude fitting; when a person achieves some difficult, highly valued record, that person is to that extent *admirable*, and to be admirable is to be the fitting object of admiration.

In this chapter, I argue that right-kind reasons are not normative reasons at all. This view has been developed by Barry Maguire (2018), where he argues that the considerations which make attitudes fitting are not normative reasons because normative reasons can be weighed against each other - they are gradable, and they combine and compete in the determination of deontic facts - whereas fit-making considerations are not gradable, and neither combine nor compete. A number of authors have attempted to rebut Maguire's arguments, including David Faraci (2018), Alexander Heape (2020), Shane Ward (2022), and Christopher Howard and Stephanie Leary (2022). Conner Schultz (2025) has recently argued for a similar conclusion to Maguire, but from a slightly different direction. He claims that there are no reasons for affective attitudes, right or wrong, because affective

[19] When I talk about reasons in this chapter, I am talking about normative reasons - considerations which count in favour of a response - even when I drop the 'normative' qualifier for shorthand.

attitudes are not the kind of things we can reason to: if there are reasons for x, one must be able to deliberate to x, but an emotion can never be the conclusion of deliberation.

I show that neither Schultz's nor Maguire's arguments are fully convincing. I then offer a new argument for the claim that fit-making considerations are not normative reasons. This argument builds on a key insight of Maguire's, namely that a weighing process is essential to the nature of the reason relation. However, rather than denying that fit-making considerations can be weighed, I argue that even if they can, the way in which they weigh is fundamentally different from reasons. In a word, the process of weighing normative reasons has a tripartite structure of normatively significant verdicts: a response can be either forbidden, permissible, or required. In the case of fit-making considerations, there are only two normatively significant verdicts: a response is either fitting or unfitting. There is no third intermediate fittingness category that resembles permissibility. This gives us reason to reject the assimilation of fit-making considerations to normative reasons.

There are four more sections to this chapter. Section 4.2 lays out the contours of the debate and why it matters. Section 4.3 evaluates some arguments against fit-making considerations as reasons that have been advanced in the literature. I conclude that these arguments are inconclusive. Section 4.4 advances my own argument. Section 4.5 concludes.

4.2. Setting the Scene

While I consider both Schultz and Maguire allies, I also want to distance the claim I am going to defend from some of the terminology they use. Here is my claim:

> **The Independence of Fit**: the fit-making relation is not the normative reason relation.

Barry Maguire calls his paper "There Are No Reasons for Affective Attitudes", while Schultz advances what he calls "Strong Eliminativism" about reasons for emotions. First of all, I will also focus on affective attitudes. I use the terms 'emotion' and 'affective attitude' synonymously. There is one assumption about the nature of emotions which makes them fertile ground for theorising about fittingness: emotions have intentional objects.

When I admire, feel proud, or feel ashamed, these responses are all directed at someone or something. This is their intentional object; the thing in the world the emotion is about. I admire a painting, or feel proud of a particular achievement. Responses are fitting when their intentional objects have features which merit or are worthy of that response; when a painting merits admiration, or an achievement is worthy of pride, those affective attitudes are fitting responses to their objects. So, if we are interested in investigating the kind of normative support that features of objects provide when they make those objects worthy of particular responses, then it makes sense to focus on responses with this kind of intentional structure, like emotions. In the previous chapter, we demonstrated that actions are plausible candidates for fit-evaluability too, and that this is partly because they may also be directed at objects. Selim Berker (2022, pp.33-34) offers the following sorts of examples of thick-fittingness properties definable in terms of fitting actions:

x is punishable =df it is fitting to be punished for x
x is persuasive =df it is fitting to be persuaded by x.
x is seaworthy =df it is fitting to sail x.
x is memorable =df it is fitting to remember x

If we think that pride is fitting in response to x whenever x is prideworthy, why not think that it is fitting to remember x whenever x is memorable? The existence of fitting actions will feature as a possible objection to Maguire's case against fit-making considerations as normative reasons in subsection 4.3.3. Nevertheless, mainly because affective attitudes are the focus of most contemporary literature on the subject, I will generally work with them in this chapter.

As for the connotations of Maguire's and Schultz's views from which I want to distance myself, it is not strictly entailed by The Independence of Fit that there are no reasons for emotions. Maguire's title is somewhat hyperbolic. He says himself that his argument only targets so-called *right-kind reasons*. But if wrong-kind reasons can be genuine reasons for affective attitudes, then it is not really the case that there are *no* reasons for affective attitudes. For example, if we think that incentives provide some normative reason for affective attitudes, then it will be true that there are normative reasons for emotions even if intuitively right-kind reasons are not normative reasons at all: the fact that I will give you a tenner if you feel sad right now can be a normative reason to feel sad, even while the fact that your long-term romantic relationship is ending is not a normative reason to be sad, but is rather a fit-making consideration, where normative reasons and fit-making considerations are distinct entities[20].

This is in contrast to Schultz, who argues that the very notion of normative reasons for emotions is misguided. Schultz's view is Strong Eliminativism, which is the view that there are no reasons for emotions whatsoever. The Independence of Fit does not entail Strong Eliminativism about reasons for emotions, since The Independence of Fit is compatible with wrong-kind reasons being genuine normative reasons for attitudes, but Strong Eliminativism does entail The Independence of Fit. If Strong Eliminativism can be established, and it is true that there are no reasons for emotions, then whatever it is that makes an affective attitude fitting, it cannot be normative reasons.

Maguire tends to talk about 'fit-making facts', with the claim being that they are not reasons, although this can be misleading. As McHugh and Way (2022, p. 189) point out, a given fact can be both a reason for a response and a fit-making fact, even if the fit-making relation is distinct from the reason relation: the fact that there are dark clouds overhead is a reason to take an umbrella, and a consideration that makes it fitting to believe that it might rain. Furthermore, there is some reason to doubt that the units of normative support are factive. Perhaps false beliefs can count in favour of a response; if by a trick of the light, the sky *appears* to be full of heavy rain clouds, but in fact is not, do I have a normative reason to take an umbrella? And is my reason my *belief* that there are dark clouds overhead, or what that belief represents, the dark clouds themselves?[21] To avoid prejudicing these issues unnecessarily, when referring to the things that provide normative support in the case of fittingness, I will speak of 'fit-making considerations' instead of facts, as well as the 'fit-making relation'.

Now, why should we care whether the fit-making relation is the same as the normative reason relation? Well, a couple of issues hang in the balance. The first concerns

---

[20] There is some debate over whether wrong-kind reasons are genuine reasons for attitudes; see Christopher Howard (2016) for discussion. Since this chapter is primarily concerned with whether right-kind reasons are genuine normative reasons, we can leave this matter to one side.

[21] Dancy makes much of this distinction in chapter 5 of his (2000).

the debate discussed in chapter 2. This is the X-First debate. A number of authors think that the category of reasons is a kind of silver bullet that can explain everything there is to explain about the nature of all normative categories. This is the Reasons-First view, set out in the work of Thomas Scanlon (1998), John Skorupski (2010), Derek Parfit (2011), and Mark Schroeder (2021b). Among the driving forces behind the Reasons-First programme is a perceived need to provide a unified account of normativity: we find issues about what's right and wrong, or good and bad, in ethics, aesthetics and epistemology. What do these distinct normative issues have in common? The Reasons-Firster claims they can all be understood in terms of reasons. And what exactly are 'rightness' and 'goodness'? Here again the Reason-Firster offers a unifying answer, they are each definable, analysable, or somehow grounded in reasons. Assuming that fittingness is a normative category, the Reasons-Firster also owes us an account of fit in terms of reasons.

A related issue concerns the fundamentality of fit. Christopher Howard (2019) and Selim Berker (2022) have argued that fittingness is a *sui generis* normative category, irreducible to any of the others. Thomas Hurka (2022) has offered arguments against this thesis. Defenders of the view that fittingness cannot be explained in terms of other normative categories clearly have a stake in the present debate. If it is true that the fit-making relation is distinct from the reasons relation, one component of fit-fundamentalism would be vindicated: fittingness is independent of normative reasons. The outcome of the present debate also has a number of broader philosophical implications. The concept of fit features prominently in aesthetics. For instance, Uriah Kriegel (2023) proposes a fitting-attitude analysis of aesthetic value, while Berys Gaut (2007) argues that artworks which endorse and elicit immoral attitudes are to that extent aesthetically defective, and this is because immoral attitudes are always unfitting, and the endorsement of unfitting attitudes in a work of art is an aesthetic flaw. These lines of argument clearly have a vested interest in the nature of fittingness.

More specifically, the claim that the fit-making relation is not the normative reason relation has direct application in the theory of emotion. Some people think that emotions are perceptual experiences (e.g. Doring, 2007; Tappolet, 2016; Milona, 2016). One objection to this view is that, unlike emotions, there are no normative reasons for perceptions. In order to respond to this, the perceptualist can do one of two things. She can try to show that there are in fact reasons for the kinds of perceptual experiences that constitute emotions, although this is a tall order, given the intuitiveness of the claim that nothing counts in favour of perceptions per se, or she can deny that there are reasons for emotions, despite appearances. This is also problematic, since the fact that there are normative reasons for emotions is often taken as a data point in theorising about them (see Naar, 2022, pp. 109-110).

However, there is a third option open to the perceptualist. Part of the difficulty in denying that there are reasons for emotions is the very familiar experience of the objects of our emotional responses providing some kind of normative support for them: if your paper, which you worked very hard on, is continuously rejected by journals, feeling disappointed is *justified*, and if someone hurts the people you love, anger is in some sense the *right* response. These notions of justification and rightness might be cashed out in terms of normative reasons, but they do not have to be. They may refer to a different kind of normative support, namely fittingness. If fit is indeed a kind of normative support distinct from reasons, and the present chapter aims to show that it is, then the perceptualist can concede that there are no normative reasons for perceptions without thereby accepting that there is a fatal disanalogy between perceptions and emotions, because like emotions, perceptions may be subject to a norm of fit. Whether perceptions can in fact be fitting will depend on how the distinctive normativity of fit is spelled out, but at the very least this provides the perceptualist with some room for maneuver.

4.3. Arguments Against Fit-Making Considerations as Normative Reasons

In this section, I canvass some arguments against the view that the fit-making relation is the reason relation. My aim is neither to reject nor defend these arguments. Instead, I draw attention to some of their shortcomings with a view to offering a more robust argument in section 4.

4.3.1. The Wrong Kind of Reason Problem

The most obvious reason to deny that fit-making considerations are normative reasons involves the distinction between the right and wrong kinds of reasons. This is essentially the line Christopher Howard takes in his (2018, p. 3). It cannot be the case that an attitude is fitting just in case there is sufficient normative reason for that attitude because such a view does not meet what Howard calls the 'crucial desideratum' for an adequate account of fit. For Howard, whatever fittingness is, the fittingness biconditionals must come out as true. This is because they are highly plausible: if x is admirable, then it is fitting to admire x, and if it is fitting to admire x, then x is admirable.

The problem with a simple reasons-based account of fit is that the second half of the biconditional comes out as false. If an evil demon threatens to destroy the world unless I admire him, then I have a very good reason to respond with admiration. Plausibly however, such a demon would not be admirable. If Howard is right that our best theory of the nature of fit should predict the truth of the fittingness biconditionals, then an unqualified reasons-based account is a non-starter. There is however an extensive literature on qualifying the notion of a reason in order to get around this problem. One approach would be to specify what kinds of considerations count as the wrong sort. Reasons widely agreed to fall within this category include incentives, such as the threats of evil demons, and moral and prudential concerns; although there may be very good moral and prudential reasons for me not to laugh at a joke told during a funeral, these do not necessarily bear on whether the joke is amusing, i.e. the fitting object of amusement (Maguire, 2018, p. 786).

In a footnote to his (2022), Garrett Cullity provides a helpful list of 10 proposed solutions to the wrong kind of reason problem, all of which involve attempts to provide necessary and sufficient conditions for when a consideration favouring an attitude counts as a reason of the right kind. For instance, Derek Parfit thought that (2011, p. 45) right-kind reasons for a response can only be given by the object of the response, while Cosker-Rowland (2017, p. 216) argues that right-kind reasons for a response are those that are neither provided nor enabled by facts about the additional consequences of having that response. For our purposes, it should be noted that whatever the right way to cash out the 'rightness' of right-kind reasons turns out to be, settling the issue will not be enough to secure the thesis that the fit-making relation is the same as the reason relation. This is because we might doubt whether right-kind reasons for affective attitudes, or fit-making considerations as I prefer to call them, are really reasons at all. The following two subsections discuss some of these doubts.

4.3.2. Arguments from Guidance

There are reasons to be sceptical of the very notion of a reason counting in favour of an affective attitude. Since fittingness is a normative relation between emotions and their objects, if there are never any normative reasons for emotions, of the wrong or right kind, then it cannot be normative reasons that make emotions fitting. One source of scepticism derives from certain purported conditions on the kinds of things that can be subject to reasons. For instance, it might be thought that in order for x to be a normative reason to φ, φ-ing must be under our *voluntary control.*[22]

Plausibly, normative reasons guide our responses. We often respond in light of them. When I recognise the fact that you are in trouble as a reason to help you, and I act on that basis, I help you *because* you are in trouble. That normative reasons guide agents in their responses has been defended by the likes of Kiesewetter (2017) and Lord (2018). This guidance condition provides a neat explanation of the distinction between the things we have normative reasons for, and the things we don't - there are no reasons for things like itches and headaches insofar as we are not generally guided by reasons to have itches and headaches[23] - and it accords with the view, typified by Jonathan Dancy (2000), that normative reasons and motivating reasons - the reasons in light of which we act - are the same sort of thing, they just explain different aspects of actions.

The thought is then that in order for a response to be guided by reasons, that response must be voluntary. On this view, some consideration can be a normative reason for me to help you only insofar as I can help you voluntarily - as an act of will - at least partly on the basis of that consideration. There are no normative reasons for itches because I cannot typically make myself itch on the basis of some consideration through an act of will. Thus:

The Voluntary Control Argument:

1. A consideration C is a reason for agent A to have (or refrain from having) an attitude X only if A can be guided by C to have (or refrain from having) X.
2. In order to be guided by C to have (or refrain from having) X, X must be voluntary.
3. Emotions are not voluntary.

[22] The following discussion of the voluntary control argument is informed by Hichem Naar (2022, p. 111).
[23] Although see chapter 3 where I discuss Na'aman's (2022) arguments against this line of thought.

∴ Therefore, there are no reasons for emotions.

The defender of the claim that there are reasons for emotions can reply in a number of ways: she can (a) deny that there is a guidance condition on reasons, (b) accept the guidance condition but deny that there is a condition of voluntariness on guidance, or (c), accept both the guidance and voluntariness conditions but deny that we have no voluntary control over our emotions. Option (a) is not promising given the reasons cited above. Option (c) is highly controversial at best. An argument to the effect that emotions are under our voluntary control can be found in Robert Solomon's (1973). The thought here is that emotions are not merely passive feelings, but active cognitive judgments made by agents who are thereby responsible for them in a way that implies voluntariness or choice. However, the intuitive pull against this implication is strong: when we lose a loved one, do we choose to cry? And when the plane we are on unexpectedly loses altitude and oxygen masks drop from above, do we choose to be afraid? Even if I do exert some voluntary control when I try to 'calm myself down' in these situations, this does not entail that I can simply will a given emotion on the basis of some consideration.

Luckily for the opponent of the voluntary control argument against reasons for emotions, this issue can be avoided, because option (b) will do the trick. We can be guided by reasons to form attitudes even though these attitudes are not under our voluntary control. The obvious example is belief. It is widely accepted that there are normative reasons for beliefs. The fact that I can hear a rain-like patter on the roof counts in favour of the belief that it is raining. What's more, I can be guided to the belief on the basis of this fact. But beliefs are not generally considered to be states we will or bring about voluntarily (see Hieronymi, 2006). I cannot simply choose to believe that I have 15 fingers, nor can I voluntarily suspend judgment on the matter, or disbelieve that I have 10. If this is right, we have a case where there are reasons for a response that meet the guidance constraint but are not voluntary. The defender of the reasons for emotions thesis can then say that voluntariness is not a genuine constraint on the sorts of responses that are subject to normative reasons, and so the argument from voluntary control does not go through.

However, it seems clear that the guidance constraint on normative reasons does entail *some* kind of control over the relevant responses. After all, how could I φ *on the* *basis of* x if φ-ing is completely outside of my control? In a recent paper, Conner Schultz (2025) has argued that if we can have a reason for φ-ing, then φ-ing must be under our *deliberative control.* By deliberation he means the settling of a question the answer to which involves forming or refraining from forming an attitude. Practical deliberation involves questions about what to do. Doxastic deliberation involves questions about what to believe. To settle deliberative questions like these is to form an intention in the first case, and a belief in the second. We deliberate *to* an attitude just in case we settle these questions *by* forming the relevant attitude.

So, for example, suppose the question is whether or not it is likely to rain this evening. I deliberate. I look outside the window and see a dark, overcast sky. I look at the weather app on my phone and see "70% chance of rain". On the basis of these considerations, I form the belief that it is likely to rain this evening. These considerations count as normative reasons insofar as they can feature in this kind of deliberation, from premises about the way the world is to the formation of an attitude. We can say that beliefs are under my deliberative control insofar as I can form them in this way. Deliberative control is thus weaker than voluntary control – it admits reasons for beliefs as well as intentions and actions – yet strong enough to exclude itches and headaches, which at least intuitively are not subject to reasons: we clearly do not settle a question about whether to be itchy by deliberating from premises about the world to the formation of an itch. Schultz then proposes the following argument:

The Deliberative Control Argument:

1. A consideration C is a reason for agent A to have (or refrain from having) an attitude X only if A can deliberate to X (or to refraining from having X) at least partly on the basis of C.
2. No one can deliberate to any emotion.

∴ Therefore, there are no reasons for emotions (Schultz, 2025, p. 73).

Again the defender of the claim that there are normative reasons for emotions has a number of options here. She can deny that there is a deliberative constraint on reasons, or she can insist that we can in fact deliberate to emotions. Let's accept premise 1, given that the guidance constraint on reasons implies some kind of control, and *deliberative* control aligns with widely-held intuitions about what reasons there are; it predicts that there are normative reasons for actions, intentions, and beliefs, but not for itches and headaches. Let's also accept, for now, the above story of deliberation, in broad outline: to deliberate to (or away from) an attitude is to settle a certain deliberative question by forming (or refraining from forming) that attitude.

What about premise 2? Is it possible to deliberate *to* an emotion? Schultz does not think so. His main reason for this is that however we formulate the deliberative question that might be settled by forming or refraining from forming an emotion, no such questions can ever actually be settled in this way. One might think that deliberating to, and away from, affective attitudes is a ubiquitous feature of our emotional lives. When we're angry we often try to calm ourselves down, and when we're sad we might try to cheer ourselves up, and both of these by deliberating with reasons for calmness or cheerfulness. Schultz (p. 78) takes the example of deliberating away from my fear of an upcoming conference presentation, to, say, a state of calm, or maybe even excitement. There are different deliberative questions I may be trying to settle here, many of which are in fact *doxastic*:

> (a) Whether I will do poorly at this conference talk
> (b) Whether it's bad that I have to give a conference talk
> (c) Whether I should be afraid that I have to give a conference talk
> (d) Whether giving a conference talk is something to be afraid of (Schultz, 2025, p. 78).

To take the last one, I reason as follows: the audience is very likely to be friendly and supportive, and I am well prepared. On the basis of these reasons, I form the *belief* that there is nothing to be afraid of. If I come to feel calm after deliberating in this way, feeling calm merely follows indirectly as a result of deliberating to the belief. However, it might be enough to deliberate to an emotion if the emotion *necessarily* follows deliberating to a certain belief.

The issue Schultz takes with this move is that emotions do not necessarily come along with associated beliefs due to the phenomenon of recalcitrance. An emotion is recalcitrant just in case it remains, or fails to be had, despite one's better judgment. In our example, it is perfectly possible, perhaps even likely, for me to continue feeling scared about my upcoming talk despite having reasoned to the belief that there is nothing to be afraid of. Schultz concludes that since recalcitrance entails that emotions do not accompany the formation of beliefs necessarily, we cannot deliberate to an emotion by deliberating to an associated belief (Schultz, 2025, p. 79).

I think this is too quick. Deliberation is fallible. Sometimes we settle deliberative questions by forming the 'wrong' attitude. For example, suppose our question is: 'whether it is raining'. I deliberate. I hear a rain-like patter on the roof. On the basis of this reason, I form the belief that it is raining. However, that rain-like patter is in fact static from an old TV out of sight. I settled a deliberative question by forming an attitude, which is to say I deliberated to that attitude, but it has the wrong content. This failure to form the 'right' attitude can explain what is going on in cases of recalcitrant emotions.

My deliberative question is 'whether I should be afraid', I weigh the considerations for and against fear, and form the belief that there is nothing to be afraid of. Nonetheless, my fear persists. Here I have settled the deliberative question by forming two attitudes, a belief and an affective attitude, the first directly, and the second indirectly. These are 'that there is nothing to be afraid of' and 'fear' respectively. The latter is wrong, given that there is nothing to be afraid of. However, the fact that it is mistaken does not entail that I have not deliberated to it. Furthermore, although emotional recalcitrance is a common occurrence, it is far from inevitable. Consider a more direct way of deliberating to an emotion involving what Schultz calls "sui generis emotional deliberative questions", such as:

(e) whether to be cheerful.

This is not like (a)-(d) in being settled by the formation of beliefs accompanied by emotional states. It is to be settled directly by forming some emotion. Suppose I am feeling a bit down, and you and I are looking for ways to cheer me up. Our deliberative question is (e). Suppose we both know that whenever I know it is going to be warm and sunny tomorrow, I feel cheerful. You show me the weather app on your phone: 90% chance of clear skies and balmy temperatures. On the basis of this consideration, I feel cheerful. Have we deliberated to cheerfulness?

It may seem a bit awkward to think of this as deliberation, but this is only because that term connotes a slow process. Thinking about the prospect of fine weather simply *makes* me happy. However, this 'making' me happy is the settling of a deliberative question about whether to be cheerful, just as the question 'whether it is raining' is settled when I am made to believe that it is by seeing the rain lash the pavement. The very possibility of cheering someone up in this way suggests that we can deliberate to emotions. For Schultz, this will not do, because my gloomy feelings may persist despite your efforts to cheer me up. However, the fact that we often fail to cheer each other up, due to recalcitrance, does not entail that cheerfulness cannot be formed via deliberation. All recalcitrance shows is that we sometimes fail to deliberate successfully to particular affective attitudes, not that it is impossible to do so.

4.3.3 Arguments from the Weight of Reasons

So far, we have seen that neither the wrong kind of reason problem nor Strong Eliminativism about reasons for emotions are entirely convincing routes to the claim that fit-making considerations are not normative reasons. I am going to consider one more possible route, which starts from the following premise: reasons have weights. The fact that I must finish this chapter by the end of the month is a reason to keep working on it. The fact that it is warm and sunny is a reason to spend time outside instead. My reason to keep working is stronger, weightier than my reason to quit and go outside. So I continue working. Thinking of reasons as weighted in this way provides certain theoretical payoffs. Most prominently, it gives us a neat account of what I *ought* to do.

When I have a toothache, the fact that getting dental treatment now will solve the problem is a reason to go to the dentist. The fact that the dental treatment will be painful and unpleasant in the short term is a reason that counts against going to the dentist. If the reasons in favour of going to the dentist outweigh the reasons against, then plausibly I ought to go to the dentist. This example is from Schmidt (2024), where he develops and defends a version of this 'balancing' account of ought, such that I ought to do whatever is favoured by the balance of reasons, a view endorsed by many other contemporary authors, including Schroeder (2007), Parfit (2011) and Snedegar (2016).

Maguire (2018) isolates three elements of the weight of reasons and argues that they do not apply to fittingness. These are competition, combination and gradability. Begin with the latter. The claim that reasons are gradable is uncontroversial. There can be more reason to do x than to do y, and less reason to do a than to do b. Can responses be more or less fitting?

As Maguire acknowledges, the evaluative properties analysable in terms of fittingness are clearly gradable. For instance, McIlroy's *Masters* win is impressive iff it is fitting to be impressed by it, and it is clearly the case that something can be more or less impressive. An obvious explanation of this is that the fittingness relation itself is gradable. But Maguire thinks this is mistaken. Imagine I have just won a round of pitch and putt against a few friends, and suppose, for the sake of argument, that this win is impressive, but that McIlroy's is more impressive. According to Maguire, there are two ways of understanding 'more impressive', only one of which is right:

1. It is *more fitting* to be impressed by McIlroy's win than mine
2. It is fitting to be *more impressed* by McIlroy's win than mine.

Maguire thinks 2 is the only correct interpretation. That is, when we say x is more impressive than y, this means that it is fitting to be more impressed in response to x, not more fitting to be impressed. This is because fittingness is *strict*, like correctness: an answer to a question is either correct or incorrect; it does not make sense to say of a particular answer that it is a little bit correct, or to say of two answers that one is slightly more incorrect than the other. Suppose our question is this: what is the capital city of the United Kingdom? You say Paris, and I say Kabul. Both of us are incorrect, and fully incorrect. Neither is more incorrect than the other. This is true even if we are inclined to *say* that I am *more* incorrect, because this inclination can be explained in terms of an implicit shifting of the standard of assessment: when we say that I am *more incorrect,* we really mean something like my answer is geographically further from the correct answer, where being geographically further away is of course a gradable notion. Maguire uses a mathematical example to illustrate this point at (2018, p. 791). On Maguire's view, fittingness is like correctness in this respect. Being impressed, for instance, is either fitting or unfitting, and if we are inclined to say that it is *more fitting* to be impressed by x than by y, we are really tracking the degrees of being impressed, rather than the fittingness of being impressed, which is non-gradable. There are a number of reasons to reject this picture of the non-gradability of fittingness. Berker offers one. He suggests that although it is possible to interpret 'more impressive' in terms of *fitting to be more impressed*, this does not work for certain non-affective actions which are plausibly subject to a norm of fit:

> "Shall we say that x is more memorable than y when it is fitting to 'more remember' x than y? That x is more forgivable than y when it is fitting to 'more forgive' someone for x than for y?" (Berker, 2022, p. 42).

The answer is no. Remembering and forgiving are not gradable, even though being memorable and being forgivable are gradable, and these two thick-fittingness categories are plausibly analysable in terms of the fittingness of remembering and forgiving respectively. It might be objected that remembering and forgiving are in fact gradable: I can remember x *well*, or *not so well*, or I can remember x *better than* I remember y, and I can *completely* forgive you, or *just about* forgive you. I do not think this shows that the actions themselves are gradable. In the first case, what comes in degrees is the quality of the memories, not the act of remembering, and in the second case, once you cross the threshold of forgiveness, then you have forgiven the person completely, and if you are teetering somewhere on the edge below this threshold, then you have not really forgiven the person at all; it is hard to see how this leaves room for degrees of forgiving. This is essentially to use Maguire's own move against him: if we sometimes speak as though actions like remembering and forgiving are gradable, we are really tracking some other gradable feature of the expression. So, if there are potentially fitting responses which are not themselves gradable - like remembering and forgiving - then Maguire's explanation of the gradable 'more A-able' as 'fitting to more A' is not applicable in all cases, and to the extent that we are interested in a unified account of fittingness, this is a reason to reject it. This may not faze Maguire too much though, as he explicitly formulates his view in terms of affective attitudes, and it is thus open to him to reject Berker's claim that the fit in 'fitting to remember' is the same as the fit in 'fitting to admire'.

Here is another reason to reject the claim that fit is non-gradable. Notice that when we talk about the gradability of reasons, this is often in the context of there being more reason to do some action rather than another, where these two actions are two distinct responses to a situation. However, when Maguire (2018, p. 790) makes the case that *more fitting* seems infelicitous, this is in the context of one response being 'more' or 'less' fitting toward two different objects. The intuition Maguire relies on - that *more fitting* to be impressed looks like the wrong analysis of 'more impressed' - does not have the same bite when we compare two different responses to the one object. Suppose the final shot that won me the game of pitch and putt was both impressive and cringeworthy, and that this is because the shot was from well outside the green, and I did a little dance as I took it. It is fitting to be impressed and it is fitting to cringe.

Let's assume that the shot was more cringeworthy than impressive. Is this best understood in terms of the fact that it is fitting to *cringe more*, or the fact that cringing is *more fitting*? I think both interpretations hold water. 'It is more cringeworthy than it is impressive' can mean it is fitting to *cringe more*. However, the following three propositions seem perfectly consistent: it is fitting to be impressed to degree n, it is fitting to cringe also to degree n, and my shot is more cringeworthy than impressive. The best explanation of this is that it is *more fitting* to cringe than to be impressed. Whether some response can be more or less fitting is ambiguous at best. This ambiguity is more than enough reason to conclude that fittingness is not strictly non-gradable.

Maguire's second argument is that unlike reasons, fit-making considerations are not contributory: they neither compete nor combine in the determination of the fittingness of attitudes. Reasons for and against available actions do seem to compete in the determination of what an agent ought to do in a particular situation. For instance, imagine I am a political leader who must decide whether or not to cut the social welfare budget. What should I do? On the one hand, cutting welfare is the only way to direct funds to much-needed infrastructure projects which will benefit the entire country, while on the other hand doing so will sever a financial lifeline for the most vulnerable members of society. These are reasons for and against the same action. They are in competition insofar as they pull the agent in opposite, mutually exclusive directions: the first plausibly contributes to the case for cutting the welfare budget, while the second counts against doing so.

Maguire (2018, pp. 787-788) argues that fit-making considerations do not compete in this way. He does so by means of examples. Here is one: your good friend Andrew gets a promotion that you wanted and that you deserved. How is it fitting to feel? On the one hand, the fact that you didn't get the promotion plausibly supports feeling disappointed. On the other hand, the fact that your good friend got the promotion supports feeling pleased. Do these considerations compete, such that the latter counts against feeling disappointed? Maguire doesn't think so. This is because it can be both fitting to be disappointed and fitting to be pleased at the same time. There is no competition between the two considerations because they simply make different attitudes fitting.

This argument is unconvincing. As Faraci (2020, p. 227) indicates, Maguire has cherry picked an example in which there is far less tension between the alternative responses, and so of course there is no obvious competition. In our example of normative reasons for actions, we saw two incompatible options: cutting social welfare vs. not cutting social welfare. I cannot do both. Opposing reasons for the opposing actions compete. In Maguire's example, being disappointed *that you didn't get the promotion* and being pleased *that your good friend got the promotion* are not incompatible; it is possible to feel both of these emotions at the same time. There is no tension and no competition. The problem is that normative reasons for actions do not compete in cases where there is no tension either: I can have some reason to vote yes on motion 1, and a different reason to vote no on motion 2. If these two actions are compatible, then the reasons favouring them do not compete. Maguire has not shown that fit-making considerations are distinct from reasons insofar as they do not compete because reasons also do not compete in cases where there is no tension between incompatible actions. What he needs is a case where fit-making considerations do not compete even though there is a tension between incompatible alternative attitudes.

For Christopher Howard and Stephanie Leary (2022, p. 234-236), two attitudes are incompatible in the relevant sense if they cannot both be fitting responses to the same object, or in other words, if x cannot simultaneously be a-able and b-able. Inspired by Howard and Leary's discussion, consider the following example:

> **Kanye West**: the artist currently known as Ye is a brilliant musician. He produced the album *My Beautiful Dark Twisted Fantasy* in 2010, commonly cited among the

best hip hop records of the first part of the 21st century. In more recent years, he has taken to spouting antisemitic diatribes and apparently endorsing Nazism.

What sorts of attitudes are fitting in response to this man and his work? Plausibly, it is fitting to admire his music, and fitting to deplore his antisemitism. The former is admirable and the latter is deplorable. Admiring and deploring are opposites, like cutting welfare and not cutting welfare are opposites. However, as Howard and Leary acknowledge, we do not necessarily have a case of incompatible alternative attitudes here, because there is no tension between admiring West's musicianship and deploring his antisemitism. I can have both attitudes toward these different qualities. What *would* be incompatible, according to Howard and Leary, is simultaneously admiring and deploring one and the same particular quality: *that West is a great lyricist* cannot be both admirable and deplorable, which is to say it cannot be fitting to both admire and deplore this particular aspect of the rapper. Maguire's argument is in trouble if we can show that in cases like this, competition between the considerations favouring the incompatible responses occurs.

At this point, Howard and Leary make a weird move. Instead of showing how different considerations might compete in determining which of two incompatible attitudes is fitting to one particular aspect of a person, they shift focus to the question of which attitude toward the person is fitting *overall:* recognising that Kanye West has both admirable and deplorable features, we can ask 'but taking everything into consideration, overall, is he admirable or deplorable?' They then argue as follows: when determining whether Kanye West is an admirable or deplorable person, all things considered, we weigh the various considerations supporting the alternative attitudes. The weight of a deplorable-making consideration is the degree to which it is deplorable, and the same for admirable-making considerations. These considerations are weighed against each other. If the deplorable qualities outweigh the admirable ones and surpass some threshold, then Kanye West is deplorable, and vice versa for the fittingness of admiration. And if neither weights meet the given threshold, then neither attitude is fitting. If this is right, we have a plausible account of how fit-making considerations compete with one another, contra Maguire.

There are a number of problems with this picture. I am going to focus on two problems pertinent to the dialectic with Maguire. Firstly, recall that in order to have genuine competition between fit-making considerations, we were looking for a case where two incompatible attitudes cannot be both fitting responses to one particular object, or in other words, where one object cannot be both a-able and b-able. Howard and Leary (2022, p. 236) assume that this is the case for the overall nature of a person, i.e. someone cannot be both admirable and deplorable. But this is far from obvious.

I can accept that some particular aspect of a person, say *that they are antisemitic*, cannot be both admirable and deplorable at the same time, but if a person has lots of deplorable qualities, as well as lots of admirable qualities, why can't they be both admirable and deplorable? A Big Mac can be both delicious and disgusting. It can be delicious insofar as it tastes good, and disgusting in light of what it's made of. There need be no further question of which one it is, overall. It is just a fact that the burger is both delicious and disgusting, which is to say it is both fitting to feel gustatory pleasure and fitting to be disgusted by it, even though these two responses are plausibly incompatible alternatives.

Secondly, Maguire anticipates the objection from overall fittingness by denying the existence of such a thing, a move Howard and Leary seem to overlook. His argument for this is a bit vague, but the thought seems to be that we never really have just one, overall emotional response to some object or state of affairs, and it makes little sense for there to be normative standards bearing on responses we never actually have (Maguire, 2018, p. 796-799). Instead, what we tend to have is a complex set of emotional responses to complex states of affairs. This is the core of Maguire's alternative to the view that fit-making considerations compete and are weighed against each other in determining which attitude is fitting, if any. Some consideration either makes a particular attitude fitting, or it does not, and this can be iterated for any number of attitudes and any number of considerations, as opposed to some consideration counting for or against one overall attitude.

While the denial of overall fitting attitudes may provide Maguire with the resources to see off Howard and Leary, if we focus on the original challenge to his view – that of showing there is no competition between fit-making considerations even in clear cut cases of incompatibility – this defense is unconvincing. Take the fittingness of amusement. A question we might ask of a joke: is it funny? Answer: it either is or it is not. Here we have a case of one object, and two clearly incompatible attitudes; to be amused or not to be amused. This mirrors more closely the incompatibility between cutting welfare or not cutting welfare: a joke cannot be both funny and not funny in the way that a person can be both admirable and deplorable.

The question is, do the considerations bearing on the fittingness of amusement toward a particular joke compete? I think they do. A plausible way of determining the funniness (or not) of a joke is to weigh competing considerations. For instance, that it is crass and reductive plausibly counts against amusement, while the fact that it lampoons its target in an unexpected way counts in favour of amusement. If both being crass and being the lampooning of someone in an unexpected way bear on the funniness of a joke, and if the one tends toward making amusement unfitting while the other tends toward making that same attitude fitting, then it is hard to see how this would not be a case of competition between potential fit-related considerations.

Maguire will want to say that these considerations just make some more complex set of emotions fitting; amusement at the unexpected lampooning and maybe disgust at the crassness. But the question remains whether or not the joke is funny. It cannot both have the property of being funny and not have this property, and so the considerations that tend toward making amusement fitting and those that make it unfitting are plausibly in competition.

Finally, this example reveals how fit-making considerations can also combine in the determination of fittingness, which Maguire denies. When deciding whether or not to cut the social welfare budget, we can imagine a pros and cons list representing these reasons on each side, with five pros and five cons. Suppose I discover an additional reason to do it, say that it will put me in good stead with a certain portion of the electorate. This sixth reason combines with the other five, such that they now outweigh the five cons, and I ought to cut welfare. A similar story can be told about the considerations determining the fittingness of amusement in response to a joke. N+1 funny-making features can combine to outweigh n features that speak against the funniness of the joke.

4.4. Why Fit-Making Considerations Are Not Normative Reasons

I have argued that neither arguments from guidance nor arguments from the weight of reasons convincingly show that fit-making considerations are not normative reasons. In this section, I offer a new argument for this claim:

The Normative Verdicts Argument

1. The weighing of normative reason issues in three distinct, normatively significant verdicts: forbidden - permissible - required.
2. The weighing of fit-making considerations, assuming such a process is possible, only issues in two distinct, normatively significant verdicts: fitting - unfitting.

∴ Therefore, the fittingness relation is not the normative reason relation.

At the beginning of 3.3, we saw that reasons have weights. We also noted the plausibility of the view that ought facts are determined by the balance of reasons. I ought to φ whenever the reasons that favour φ-ing outweigh the reasons that favour any alternative to φ-ing, and I ought not to φ whenever the reasons against φ-ing outweigh all reasons that favour φ-ing. I assume that φ-ing is required whenever it is the case that the agent ought to φ, and φ-ing is forbidden whenever the agent ought not to φ.

Now, there is also a third possible outcome of this weighing process. Φ-ing is *merely permissible* whenever it is not the case that φ-ing is required and it is not the case that φ-ing is forbidden. In the language of weighted normative reasons, the balance neither tips in favour of nor against φ-ing. I am going to give two reasons to think that permissibility is a distinctive, normatively significant deontic category, which I outline in terms of moral reasons for action[24]. I then defend premise 2 by showing how neither of these considerations apply to fittingness[25]. This move will be prefaced by two further reasons in support of premise 2.

The first consideration that reveals the normative significance of permissibility concerns demandingness. Shelly Kagan refers to permissible actions as 'options', and

[24] Philip Stratton-Lake (2025, p. 154-155) discusses these two considerations in a roughly similar context.
[25] Though I illustrate the two considerations in moral terms, they plausibly apply in non-moral normative domains such as the epistemic and prudential as well (see McElwee, 2017). Therefore, the fact that fittingness is non-moral, if this is a fact, is not a good reason to deny that they apply to it.

thinks that they are an integral part of "ordinary morality" (Kagan, 1991, p. 3)[26]. Part of the reason for this is that there seem to be limits on what moral reasons can demand of agents. Plausibly, there is a standing, very weighty reason to promote the good. If I can either save one person or save five, I ought, other things equal, to save the five. And this is because doing so will promote the greater good. But what about cases where to save five I would have to sacrifice my life or my own projects in some way. Here it seems it is not required to do what's best, but nor is it forbidden. I am not morally blameworthy for failing to sacrifice myself to save five people. Refraining from doing so is permissible. The existence of permissible action lets me off the hook, morally speaking, and this fits with the common-sense assumption that moral reasons should not be overly demanding.

Relatedly, the permissible partially grounds the supererogatory. An action is supererogatory when it is neither required nor forbidden, but would be a very good thing to do. If I throw myself on top of a live grenade to save five comrades, I have gone beyond the call of duty and done something saint-like. If there were no permissible actions, and sacrificing myself to save five people were either required or forbidden, then we would lose the sense in which the supererogatory constitute a class of highly admirable acts, and this is because part of what makes them admirable is that they go *beyond* what is required (Stratton-Lake, 2025, pp. 154-155).

So, the balance of normative reasons for and against φ can be such that φ-ing is neither required nor forbidden, and this constitutes the distinctive category of permissibility which is normatively significant insofar as it is part of the explanations of optionality and supererogation. What about fit? Assuming for the sake of argument that some weighing story can be told about the determination of fittingness, is there a distinctive, normatively significant category instantiated when a response is neither fitting nor unfitting? I do not think there is.

Firstly, there is no term in English for 'neither fitting nor unfitting', beyond simply 'not-fitting', in the way that 'permissible' is a distinctive term for 'neither required nor forbidden'. Although a linguistic fact like this does not settle the metaphysical issue, it does shift the burden of proof toward those who defend the existence of a category for which we

[26] Kagan of course goes on to argue at length against the existence of options. It is beyond the scope of this thesis to respond to these arguments. In any event, the fact remains that the existence of options is an intuitive component of common-sense morality, even if common sense is mistaken.

have no word. Secondly, as mentioned in chapter 2, there is some debate over whether fittingness itself is more like a requirement or a permission. When shame is fitting this is generally thought to be requirement-like, while fitting love seems more like a permission.

On the basis of this data, Selim Berker (2022, pp. 36-40) thinks that fit is something more than a permission, but less than a requirement, while James Fritz (2023, p. 2600) argues it is sometimes permission-like and sometimes requirement-like. Fritz's position entails the binary structure of fittingness verdicts I am arguing for. If fittingness is sometimes itself a kind of permissibility, then a permission-like category between the fitting and unfitting is not a third, distinct verdict. Berker's position here is more difficult to pin down, but if fittingness is neither a kind of requirement nor a permission-like status, then a view which says that fittingness verdicts map onto normative reason verdicts in a straightforward way is ruled out.

It might be suggested that for all this, there are three distinct fittingness verdicts: fitting - non-fitting - unfitting. In his (2024, pp. 1073-1077), Berker argues against this possibility. He thinks that unfitting is identical to non-fitting, as opposed to 'anti-fitting'. This closes off the possibility of an independent non-fitting category between the fitting and the unfitting. If he is right, this supports the disanalogy I am advocating between a tripartite normative reasons structure and a binary fittingness one. However, let's suppose that he is wrong, and non-fitting is distinct from unfitting. There are a couple of reasons to think this is so. First, certain examples are suggestive. It is plausibly fitting to be amused by a funny joke, unfitting to be amused by the death of a loved one, and neither fitting nor unfitting to feel slightly apathetic during a philosophy talk. A mild kind of apathy while listening to someone list the axioms of expected utility theory may not be fitting, but it does not seem particularly unfitting either.

Another reason to think there is non-fittingness involves a sorites-like case, where we imagine a spectrum with 'reprehensible' at one end and 'not reprehensible' at the other. Suppose that some action is reprehensible whenever it is fitting to have censorious attitudes such as blame and disapproval toward an agent who does the action, and an action will not be reprehensible just in case it is unfitting to blame or disapprove of the agent. Plausibly, stealing £100,000 is reprehensible, whereas stealing 1p is not reprehensible; it is fitting to blame the agent in the former scenario, but unfitting in the latter. It also seems plausible that there will be some point between these two extremes where stealing that amount of money is neither reprehensible nor not reprehensible, and thus where blame is neither fitting nor unfitting[27].

These two considerations indicate how there might in fact be a distinctive 'non-fittingness' category which would give fit a tripartite structure. I now want to argue that even if we allow for this third category between the fitting and the unfitting, 'non-fittingness' is not *normatively significant*, and this is because we find no issue of demandingness and no supererogation, the two phenomena which characterised the normative significance of a distinct category of permissibility in the case of normative reasons for action. If this is right, then non-fittingness would be normatively inert, if it is normative at all, and so the structure of normatively significant fittingness verdicts will turn out to be binary.

One potential source of demandingness in the fittingness case is the phenomenon of 'forever fitting feelings', discussed by Howard in his (2023). If you wrong me, anger is plausibly a fitting response. Since the features of the wrong that make anger fitting never change, given that the wrong is in the past (and assuming none of the relevant features are time-indexed), then anger must remain fitting forever, given that fit is a match between an attitude and the features of an object that make it fitting. This is quite a counterintuitive result. If you drank the last of my milk in the fridge after I had explicitly asked you not to, it looks as though it is fitting for me to be angry at you. Anger is fitting in virtue of the fact that you drank the last of my milk. Suppose nothing is done to make amends; you do not apologise or replace the milk and so on. Six months later, I am angry at you for drinking the last of my milk. This seems like an entirely unfitting way to feel about that particular incident. However, if the fact that you drank the last of my milk makes anger fitting, and nothing about this fact has changed, then nothing about the fittingness of anger can have changed either. So my anger must remain fitting six months down the line.

That it is fitting for me to go on feeling angry in this way is also counterintuitive insofar as it appears to be awfully demanding on my emotional capacities. If I am to respond as it is fitting for me to respond, I would go on feeling angry at you for drinking the last of my milk for the rest of my life. Anger is a famously debilitating emotion: it often

[27] This point and example were suggested to me by Alex Gregory.

makes us act rashly, and it is difficult to focus on anything else whilst angry - the proverbial red mist. Furthermore, think of all the things I can be forever fittingly angry about, like the time my bike was stolen, and the decisions of various governments, or all the other countless emotions that were made fitting by events in my past and remain fitting insofar as these unchanged events continue to be fit-makers. To feel as it is fitting to feel would demand of me the impossible.

There have been some attempts to explain how emotions can fittingly fade over time (e.g. Oded Na'aman, 2021), but perhaps we could deal with this apparent overdemandingness by making room for a category of *fitting options*. Just as options in the realm of practical reason let us off the hook in the sense that it can be permissible to not-φ even when there are very strong reasons to φ, so fitting options, the fittingness correlate of permissibility, could let us off the hook emotionally even when there are fit-making considerations which seemingly go on making certain emotions fitting forever.

However, there is a simple reason why there is in fact no issue of demandingness here. Fit does not imply should. This is essentially Howard's preferred solution to the problem of forever fitting feelings (2023, p. 95). As we saw in the previous section, a joke can be the fitting object of amusement even though I should not be amused, because it is told during a funeral, for instance. What's more, this is relatively independent of the question of whether fittingness is itself requirement-like or permission-like. Even if fittingness is more like a requirement than a permission, this 'requirement of fit' is very easily outweighed by practical considerations.

Suppose that various funny-making features of a joke impose a requirement of fit on me to respond with amusement. If someone offers me a tenner to refrain from laughing, then I should not laugh. Or to take another example, when you and I are standing in front of the Sagrada Família, it is fitting for me to admire the building, but when you indicate that you are hungry and ask what time we are going to have lunch, I really ought to quit admiring and think about our afternoon meal. Practical considerations readily trump fit-making considerations when it comes to what I ought to do all-things considered, and so even if anger can be fitting forever in a requiring sense of fit, this will never imply that I should feel angry until the day I die, since the fit-making considerations are bound to be outweighed by practical reasons, including basic reasons for getting on with other things in my life. This implies that the phenomenon of forever fitting feelings, assuming such a phenomenon exists, exerts no excessive demand on my emotions, and this in turn dissipates the normative need for a third category of fitting optionality to let us off the hook. As for the possibility of a kind of aptic supererogation, consider:

> **Grief**: Lucy learns that her good friend's pet hamster has passed away. She breaks down in floods of tears, feeling as though the ground has fallen away from beneath her feet. She is inconsolable.

Is Lucy's response fitting? Although some degree of sadness plausibly fits the tragic story, Lucy seems to have gone *beyond the call of fittingness* in a way analogous to how she goes *beyond the call of duty* when she runs back inside a burning building to rescue others. There may also be a sense in which Lucy's extremely empathetic response to the poor hamster's fate is highly admirable, even saint-like. Feeling intense grief is much better than a very mild sadness in the same way that sacrificing yourself to save five people is much better than standing by to save your own skin. It might then be argued by analogy that Lucy's response is supererogatory along the fittingness dimension. And if this is so, then we may need a normatively significant notion of non-fittingness to explain it: intense grief at the death of a friend's pet hamster is neither fitting nor unfitting, leaving room for aptic supererogation.

However, whereas the supererogatory act of sacrificing oneself to save five people is permissible, which is to say neither required nor forbidden, it is not the case that Lucy's intense sadness is neither fitting nor unfitting. It is simply unfitting. To see this, consider the widely recognised fact that the strength of an emotional response bears on its fittingness; the *size* of the attitude matters (Achs and Na'aman, 2023, p.2532). Here is an example:

> **Size**: "Holly meets her colleague and they are both wearing the exact same wardrobe. This is surprising, but Holly is so surprised that she falls out of her seat and cannot recover for several minutes. Her surprise doesn't seem fitting." (Achs and Na'aman, 2023, p. 2529).

The degree of surprise that Holly experiences is unwarranted. If we put numbers on it, with 0 being no surprise, and 10 being the shock of your life, the outfit coincidence plausibly merits a 2, it does not merit an 8. To say that a surprise of degree 8 is not merited is to say that it is unfitting. Returning to Lucy, when she learns of the passing of her friend's rodent, a mild kind of sadness is plausibly fitting. Amusement would be unfitting, assuming there is nothing funny about it. But intense grief seems just as unfitting, since the incident does not merit intense grief any more than it merits amusement. If the degree to which an emotion is felt were irrelevant to the fittingness of the response, then intense, inconsolable grief would be just as fitting in response to the death of a friend's pet hamster as it is to the premature death of a parent, but that is absurd.

This means that when an emotional response goes beyond the call - when the degree to which it is felt is greater than the degree that is merited by the object - the response is unfitting, rather than non-fitting. This is in contrast to an action which goes beyond the call in the normative reasons case. Such actions are not thereby forbidden. They are permissible and supererogatory. If this is right, there is in fact no fitting supererogation, and combined with the absence of demandingness concerns, this indicates that even if there is a third, non-fittingness category between the fitting and the unfitting, it is not normatively significant in the way that permissibility is. This gives us good reason to think that fit-making considerations only issue in two normatively significant verdicts - fitting or unfitting - as opposed to the three normatively significant verdicts that can be reached via the weighing of normative reasons, and if this tripartite structure of normatively significant verdicts is essential to what normative reasons are, then fit-making considerations are not normative reasons.

4.5. Conclusion

In this chapter, I have argued that fit-making considerations are not normative reasons. We began by examining some arguments for that claim in the literature, concluding that they are ultimately unsuccessful. Arguments from guidance, including the voluntary and deliberative control arguments, attempt to show that there are no reasons for emotions whatsoever, and since emotions can be fitting, fit-making considerations cannot be reasons. The first of these fails because it imposes a condition on what there are reasons for which does not exist, while the second imposes a condition which plausibly exists, but which emotions can arguably meet.

Barry Maguire's arguments from the weight of reasons are also unconvincing, given that there is at least some reason to think that fit-making considerations are gradable, and both compete and combine. I then offered a new argument for the claim that fit-making considerations are not normative reasons: normative reasons issue in three normatively significant verdicts. Fit-making considerations only issue in two. This disanalogy in the structure of normative support - tripartite vs. binary - entails that the fit-making relation is distinct from the normative reason relation.

Chapter 5. Epistemic Perspectivism about Ought, Fit, and Value

5.1. Introduction

In this chapter, I argue that epistemic perspectivism about fittingness – a view roundly rejected in the literature (McHugh and Way, 2022, pp. 78-9) – is supported by the same considerations that support epistemic perspectivism about ought. I then show how these considerations do not support epistemic perspectivism about good. This represents a problem for Value-First theories of normativity (e.g.Cosker-Rowland, 2022a and Maguire, 2016): the fact that 'ought' and 'fit' are plausibly constrained by the agent's epistemic position, while goodness is not, will need to be accounted for by theorists who seek to explain all normative categories in terms of value. More specifically, I argue that we have reason to reject any view which says that fittingness is reducible to value, and this is because, while the instantiation of evaluative properties is determined solely by non-epistemic features of the world, there is some reason to think that the instantiation of the fittingness relation is in part determined by facts about an agent's epistemic position.
Epistemic perspectivism is a popular view about the 'deliberative ought' (Zimmerman, 2008 and 2014, Kiesewetter, 2011 and 2017, Mason, 2013):

> **Epistemic Perspectivism about Ought**: A ought to Φ iff A ought to Φ relative to A's epistemic perspective.

A stands for the agent, any agent. By ought I mean the deliberative, or all-things-considered ought; what is essentially the answer to a question of the form 'what should I (the agent) do now?'. I use the term 'should' as well as 'obligation' and 'right' interchangeably with 'ought'. 'Wrong' I use interchangeably with 'ought not'. There may be lots of important differences between these categories, but I will assume that nothing hangs on these differences, mainly because different writers on epistemic perspectivism couch their views in different terminology. Φ stands for a response, namely an action or an attitude.

There is some debate over what exactly an agent's 'epistemic perspective' amounts to. W.D. Ross (1939) and H.A. Prichard (2002) are generally interpreted as holding that it is the agent's beliefs that count. Michael Zimmerman (2008) rejects this view and defines the relevant epistemic perspective in terms of prospects, or reasonable expectations. We might also think of the agent's epistemic perspective in terms of her evidence, or her knowledge. The present chapter is neutral on this issue: I am interested in the viability of epistemic perspectivism about ought, fit, and value generically understood, however the relevant epistemic position is best defined. Objectivism about ought is the rejection of epistemic perspectivism, a view defended by Peter Graham (2010):

> **Objectivism about Ought**: A ought to Φ iff A ought to Φ relative to the way the world is, irrespective of A's epistemic perspective.

Another important normative category is fittingness. To reiterate what we have said about fit in previous chapters, fittingness is a match between an attitude and its object. It is fitting to admire x when x is admirable. Fittingness matters insofar as we care about having the *correct* emotional and attitudinal responses to the world around us. Adapting the above view about ought, we get:

> **Epistemic Perspectivism about Fittingness**: Φ is a fitting response to x iff Φ is fitting to x relative to A's epistemic perspective.

A third normative category is value. I am interested in goodness and badness *simpliciter,* which pertain to states of affairs. I use the term 'value' interchangeably with 'good'. Our final view:

> **Epistemic Perspectivism about Good**: x is good iff x is good relative to A's epistemic perspective.

Objectivism about fittingness and objectivism about value are the rejections of these two views. The primary aim of this chapter is *not* to provide a full vindication of epistemic perspectivism about ought, much less of epistemic perspectivism about fittingness, but rather to establish that the very same reasons which support the former, also support the latter. If you are sympathetic to epistemic perspectivism about ought, then you ought to be sympathetic to epistemic perspectivism about fit as well. Furthermore, these same considerations do not support epistemic perspectivism about value to the same extent. With respect to the agent's epistemic position, there is therefore a significant gulf between ought and fit on the one side, and value on the other. The chapter thus defends the following three claims:

1. Epistemic perspectivism about fittingness is supported by the same considerations that support epistemic perspectivism about ought.
2. Epistemic perspectivism about value is not supported by these same considerations.
3. If 1 and 2 are true, then we have reason to doubt the fundamentality of value within the normative domain.

This is the plan. Section 5.2 motivates a debate between epistemic perspectivists and objectivists about ought and extends this motivation to a parallel debate about fittingness. It is argued that a third, 'sense-splitting' position which promises to dissolve the debate fares no better than epistemic perspectivism or objectivism with respect to the primary motivation behind it. Section 5.3 offers one argument against objectivism about ought and fit, and one in favour of epistemic perspectivism about the two normative categories. I do not pretend that these arguments are conclusive. The point is rather to show how the very same moves which can be made by the epistemic perspectivist about ought can also be made by an epistemic perspectivist about fit. Section 5.4 turns to value and argues that the considerations brought to bear in sections 5.2 and 5.3 have little to no purchase when it comes to the good. Section 5.5 concludes.

5.2. Epistemic Perspectivism vs. Objectivism: Ought and Fit

There are a number of ways of making the distinction between epistemic perspectivism and objectivism about ought salient, including what seems to follow from Frank Jackson's (1991) famous three-option case (see Fassio, 2022 for references). Here we imagine a sick patient and a doctor with three pills, A, B, and C. A will partially cure the patient, B will kill her, and C will fully cure her. While the doctor knows that A is the partial cure, she does not know which of B and C is the complete cure and which is fatal to ingest. What should the doctor do? The distinction between objectivist and perspectivist answers is glaring: according to all of the facts, irrespective of the doctor's epistemic position, she ought to give the patient pill C. However, according to her epistemic perspective, she ought to give pill A, the partial cure, given her ignorance about the contents of B and C.

Because Jackson-type cases like the one above are not as clearly relevant when it comes to fittingness – a confrontation with a selection of options we need to actively choose between is less familiar in the case of attitudes than actions – I follow Kristian Olsen's (2017) discussion and structure the debate between epistemic perspectivists and objectivists around a distinctive tension between equally intuitive though contradictory answers to a question about what the agent ought to do, and the attitude it is fitting to have, in certain situations. To begin with ought, consider the following:

> **Twin Towers***:* "Following the crash of an airplane into a skyscraper, security guard Tom, believing that the elevators will cease working, tells office workers to evacuate the building via the stairwell rather than the elevators. In this case, using the stairs takes too long and all the office workers are killed when the building collapses, whereas the elevators remain operational long enough for the employees to have used them to evacuate safely." (Smith, 2010, p. 64)

Does Tom act as he ought? Two contradictory answers suggest themselves: on the one hand, it seems as though he does. After all, he fully believes that taking the stairs affords everyone the best chance of survival, and for good reason. In light of his understanding of the situation then, he ought to direct the workers to take the stairs. On the other hand, it seems he ought *not* to do this. Why? Because as it turns out, telling everyone to take the stairs results in their deaths. The tension at issue can be expressed in terms of the feeling of being pulled in different directions by these two answers: we want to say that Tom acted *both* rightly *and* wrongly. The debate gets going when we recognise that one and the same action cannot be both right and wrong; these are mutually exclusive categories. The tension thus demands a resolution. According to the epistemic perspectivist about ought, Tom ought to tell the workers to take the stairs, while the objectivist intuition is simply mistaken and should be abandoned. Objectivism about ought gives the opposite result: if objectivism is true, Tom ought to tell the workers to take the elevator, while the perspectivist intuition is simply mistaken and should be abandoned.

A third method of resolving the tension avoids the awkward task of explaining away the opposing intuition and instead accepts that there are two distinct *senses* of ought:

> **Sense-Splitting about Ought**: there are two distinct senses of ought, an *objective* sense and a *subjective* sense.

On this kind of view, despite appearances, there is no tension or paradox involving contradictory intuitions about what the agent ought to do in the Twin Towers case because saying that Tom both ought and ought not instruct the workers to take the stairs is no contradiction. This is because two distinct deontic categories are being used, i.e. he *subjectively ought* to instruct them to take the stairs, and he *objectively ought* not to do this[28]. Olsen (2017) offers an argument for this sense-splitting position, which can be summarised as follows:

1. Cases like *Twin Towers* generate a tension between two equally intuitive though incompatible moral judgments.
2. While single-sense views may be able to account for this tension in some cases, they cannot handle certain variations which generate a similar tension.
3. Sense-splitting comfortably accounts for the tension in all cases.

The argument trades on the fact that the correct theory of obligation should be able to account for the tension between opposing intuitions about what an agent ought to do. The sense-splitter is well-equipped to do this, single-sensers – like the objectivist and the epistemic perspectivist – struggle to. We should therefore adopt sense-splitting, or so the argument goes.

For Olsen, the sense-splitter's account of the tension in Twin Towers is straightforward: she can simply say that Tom's action really is both right and wrong, just in different senses. He then suggests that the single-senser may also have an explanation. This is because the conflict between contradictory intuitions is not indicative of the fact that Tom's act seems both right and wrong, but rather that it calls for both positive and negative evaluation.

[28] This sort of view can be found in Andrew Sepielli's (2018).

The epistemic perspectivist can provide this: she can say, positively, that what Tom did was right, and negatively, that terrible consequences ensued. Since terrible consequences do not entail that an act is wrong, the epistemic perspectivist can comfortably account for the objectivist intuition without contradiction. Something similar can be said for objectivism. On this view, Tom acted wrongly. This gives us our negative evaluation of the situation. But positively, he is *blameless*; he did what he reasonably thought was best, not being in a position to know that taking the stairs would end in disaster. If this diagnosis of the tension is correct, and it is simply a case of wanting to say both positive and negative things about Tom's action, then both single-sense views appear to be able to account for it. According to Olsen however, this method of accounting for the intuitive tension will not work for the single-senser in other cases. I will focus on the trouble he alleges for epistemic perspectivists. Firstly, Olsen asks us to consider the following situation:

> **Airport**: "Two of my friends, Jesse and Andrew, need to be picked up from the airport. They are at different airports, and I can pick up only one of them. I made a promise to one of the two that I will pick him up, but I made no such promise to the other. I believe, and all my evidence suggests, that I made the promise to Jesse. However, my belief is mistaken: I actually made the promise to Andrew. Suppose, however, that the consequences of my two options are equally good: no matter which friend I pick up, it will be a minor inconvenience to the other. In the end, I choose to pick up Jesse." (Olsen, 2017, p. 355)

Again a similar tension: on the one hand, the agent seems to have acted rightly, believing he was picking up the friend he made the promise to, while on the other hand, he acted wrongly, breaking the promise he actually made. Attempting to apply the same method of accounting for the tension, the epistemic perspectivist can say positively, that the agent acted rightly, but they cannot say, as in the Twin Towers case, that negatively, there were terrible consequences of failing to do what seems objectively right, because as Olsen points out, the value of picking up either friend is equal by stipulation. There are no bad consequences of acting according to subjective rightness, so the epistemic perspectivist can provide no negative evaluation of the situation, and thus no explanation of the opposing intuition. For Olsen, the fact that neither of the single-sense views can account for both intuitions in cases like Airport is a reason to adopt sense-splitting, the view that there are two kinds of oughts, an objective one and a subjective one, which straightforwardly captures both intuitions.

I think there are a number of problems with Olsen's argument for sense-splitting. First of all, while it is true that any adequate theory of obligation ought to be able to explain the intuitive tension between contradictory responses to the cases described above, it is not clear that Olsen's characterisation of the tension is an adequate one. To begin with, the stipulation that there are *no* bad consequences of picking one friend up but not the other is implausible. At the very least, the person you do not pick up will have their feelings hurt; a minor inconvenience is still an inconvenience. Secondly, Olsen frames the issue for single-sense views in terms of whether or not they can say both positive and negative things about certain situations. But is this really the crux of the tension? Olsen himself recognises that it is not, writing:

> "what many people find intuitive about the cases we've been discussing is not merely that the action in question is subject to both a positive and a negative moral evaluation, but rather, that the action seems, from one perspective, to be morally right, but from another perspective, it seems morally wrong" (Olsen, 2017, p. 357).

In light of these remarks, it is puzzling why he should devote a large part of his argument to showing how the epistemic perspectivist (and indeed the objectivist) struggles to say both positive and negative things about cases like Airport. What's more, I think this amended account of the tension at stake is still misleading. Notice the phrase 'from one perspective' in the above quotation. By smuggling this in, Olsen essentially loads the dice.

To characterise the tension itself as a conflict between rightness and wrongness *from different perspectives* is question begging. For the defender of a single-sense view can insist that the tension is not about a particular action seeming morally right from one perspective, and morally wrong from another, but simply that the action seems both morally right and wrong, without qualification. It may well be the case that the single-senser will struggle to account for these appearances, but if no prior assumptions are made about the existence of rightness and wrongness from different perspectives, then the sense splitter also needs a story about this unqualified tension, one which Olsen makes no attempt to provide. On top of all this, and perhaps more importantly, sense-splitting explicitly dodges a question asked by Ross:

> **The Rossian Question about Ought**: "which of the characteristics – objective or subjective rightness – is ethically the more important, which of the two acts is that which we ought to do?" (Ross, 1939, p. 147)

Even if we accept that there is a subjective and an objective sense of ought, the question remains which is ethically the more important, or in other words, what ought we to do, all things considered, when subjective rightness and objective rightness prescribe different actions. In the Twin Towers case, subjective rightness prescribes instructing the workers to take the stairs, while objective rightness prescribes leading them to the elevator. But which should Tom do? When faced with this decision, sense-splitting has no answer, and if we recognise the force of Ross's question, then this is a serious problem for the view.

What I hope to have shown is that there is no easy resolution of the tension which engenders the debate between epistemic perspectivists and objectivists about ought. It is a live issue, and adjudicating between the two views must come down to the strengths and weaknesses of arguments for and against. A parallel debate between epistemic perspectivists and objectivists about fittingness arises in light of an analogous tension. Consider:

> **Credit-Stealer**: John is a budding 19th century chemist. He presents some ground-breaking research to the Royal Society. The fellows greatly admire the man and his achievement. However, unbeknownst to anyone, John's housekeeper, Marie, doubles as his lab assistant. It was in fact Marie who did the bulk of the work, and her creative reasoning which led to the stunning conclusions John presented as his own.

Is it fitting to admire John? There is a tension here between equally intuitive though contradictory answers: on the one hand, it seems as though it is fitting for the fellows to admire John, given their epistemic position. On the other hand, it seems as though admiration is unfitting, given the facts of the situation irrespective of anyone's epistemic perspective. We thus have one and the same attitude appearing both fitting and unfitting: admiration is fitting in light of the fellows' evidence and beliefs about John's contribution to science, and admiration is unfitting because John in fact stole the credit for someone else's achievement. It might be argued that there is in fact no tension here because there are two distinct objects against which we are measuring the fittingness of admiration, rather than one. Admiration may be fitting in response to the fellows' *belief* that John made a great contribution to science, while unfitting in response to the actual situation, wherein he has stolen credit for someone else's work. However, while beliefs can sometimes be the proper objects of attitudes like admiration, I do not think this is what's happening in Credit-Stealer: the fellows are responding to John and his (supposed) work, not to their mental states.

So, the tension stands, and it calls for a resolution. Other things equal, according to the epistemic perspectivist about fit, admiration is fitting, while for the objectivist, it is unfitting. As in the case of the debate over ought, the onus is on each of these views to somehow explain away the opposing intuition, or bite the bullet and accept that there are two distinct *senses* of fit. I think the same reasoning advanced in the above discussion of ought applies to the case of fit: there is no easy way to account for the opposing intuition, and biting the bullet does not really help. Splitting fit gives us:

> **Sense-Splitting about Fit**: there are two distinct senses of fittingness, an *objective* sense and a *subjective* sense.

While we might be tempted by sense-splitting insofar as it promises to account for the tension between conflicting intuitions about cases in a way that neither single-sense views can, it fails to fulfil this promise. This is because sense-splitting about fit proposes that some response Φ is fitting *from one perspective*, and unfitting *from another perspective.* However, it is unclear how this accounts for the tension if the tension is a matter of Φ seeming both fitting and unfitting *without qualification.* Furthermore, sense-splitting about fit is also subject to the same objection to sense-splitting about ought vis-à-vis the question asked by Ross:

> **The Rossian Question about Fit**: which of the characteristics – objective or subjective fittingness – is the more important, which of the two attitudes is the one which is fitting?

Here we can imagine, as seems plausible, that admiration of John is subjectively fitting, while some kind of disapprobation is objectively fitting. Admittedly, the Rossian question may be less forceful with respect to fit than in the case of ought. The force of the Rossian question as an objection to sense-splitting about ought is derived from the fact that an agent must make a decision about what to do; through a process of deliberation, she needs an answer to the question 'what should I do now', and saying 'well, you should Φ in one sense, but not-Φ in another sense' is really no answer at all.

When it comes to our attitudes, we don't generally have to make a conscious decision about which one to have in a given situation, and even when we do, the consequences of that decision are likely to be less pressing than those of our actions. However, fittingness is a measure of the match between attitudes and their objects, regardless of what attitude the agent actually has, and determining whether there is such a match does demand an answer to the Rossian question. Imagine you and I are watching an action film, and we want to determine whether some choice the protagonist made is regrettable or not. To be regrettable is to be the fitting object of regret. Suppose we think that from the character's perspective, regret is a fitting response to the choice he made, but given the facts we know from the privileged perspective of viewers (the facts according to the fiction that is), regret seems unfitting. If we conclude that regret is thus subjectively fitting and objectively unfitting, we are no closer to determining whether the choice is in fact regrettable; that is, whether regret is a fitting response. We need to know which of the two kinds of fittingness – the objective or the subjective – determines whether the choice really is regrettable or not.

The result to take forward is this: if the debate between epistemic perspectivists and objectivists about ought is well-motivated, then a parallel debate about fittingness is too, and for the very same reasons.

5.3. Two Arguments for Epistemic Perspectivism about Ought and Fit

In this section, I provide two arguments, one against objectivism about ought and fit, and one in favour of epistemic perspectivism about the two normative categories. Persuading the reader to wholeheartedly endorse epistemic perspectivism about ought and fit on the basis of these arguments is not the goal here. Instead, the point I want to make is a conditional one. *If* these arguments lend support to epistemic perspectivism about ought, then they should be taken to support epistemic perspectivism about fittingness in the very same way and to the very same degree. There are many more reasons and arguments for and against epistemic perspectivism about ought which will not be discussed in this chapter, most notably those found in Zimmerman's two books; (2008) and (2014). I have chosen the two arguments I focus on because I take them to support the generic formulations of epistemic perspectivism stated in the introduction to this chapter, as opposed to more sophisticated versions, such as Zimmerman's 'prospectivist' epistemic perspectivism which cashes out the relevant epistemic perspective in terms of evidence. I am interested in the general question of whether fittingness and goodness simpliciter are plausibly constrained by an agent's epistemic position in a way analogous to ought. I am not interested, at least not in the present chapter, in developing and defending the best version of epistemic perspectivism about any particular normative category. The first argument I want to consider, derived from W.D. Ross (1939, pp. 156-157), runs as follows:

1. The Argument from Rationality

> P1. If A ought to Φ, then it is rational for A to Φ, and if it is fitting for A to Φ, then it is rational for A to Φ.
>
> P2. It is sometimes *irrational* to respond in accordance with 'objective ought', and it is sometimes *irrational* to respond in accordance with 'objective fit'.
>
> C1. Therefore, objectivism about ought and objectivism about fit are false.

Begin with the first premise. P1 comes out as true when combined with the plausible view that ought, fit, and rationality can all be understood in terms of reasons for responses, broadly construed. The thought that what one ought to do is determined by the balance of one's reasons is as close to orthodoxy as any view in contemporary metaethics. Cosker-Rowland (2019, p.194) points to a marked correlation between what you have reason to do and what you ought to do. If you ought to take out the rubbish, for instance, then you have a reason to do this, maybe because you promised your partner you would. For Cosker-Rowland, this correlation demands explanation. There are two options: ought is grounded in reasons, or vice versa. Cosker-Rowland opts for the former[29]. Further expressions of this view can be found in Dancy (2004) and Schroeder (2007), while Justin Snedegar (2016, p. 158) suggests that a principle along the following lines is 'very widely accepted':

> **Ought As Most Reason**: A ought to Φ just in case A has most reason to Φ.

Something like this has a lot of intuitive appeal. Reasons have weights. My reason to continue writing this paragraph is plausibly stronger or weightier than any reason I might have to scroll through the day's headlines. If I have *more reason* to write than I have to read the news, then plausibly I *ought* to write. In other words, I ought to do what I have most reason to do. When it comes to fittingness, the wrong-kind-of-reason problem (Rabinowicz and Rønnow-Rasmussen, 2000), which tells us that a demon's threat can give you a reason to admire him even though he is not thereby *admirable,* scuppers a straightforward reduction of fit to reasons. However, a little alteration gives us the following thesis:

> **Fit as Most Right-Kind Reason**: it is fitting to Φ just in case A has most reason *of the right kind* to Φ.

How to best articulate the notion of a 'right-kind' reason is the subject of ongoing debate (Gertken and Kiesewetter, 2017). However, in light of the argument of chapter 4, we need not wade into that particular thicket, because the connection between fit and rationality can be made without resorting to the idiom of normative reasons. This is because we can talk about the 'fit-making features' of a situation, or 'fit-making considerations' instead. While on my view fit-making considerations are distinct from reasons, they are similar in many ways in virtue of offering a kind of normative support to responses. Take the fittingness of fear. If it is the case that danger is a feature of a situation which makes fear fitting, as seems

[29] John Broome (2015) notably goes the other way. He explains reasons in terms of oughts. Even if Broome is right, on his view, there is still an essential connection between ought and reasons, and if such a connection also exists between rationality and reasons, this may be enough to secure the 'ought' part of P1.

plausible, we can say that danger also makes fear a rational response. Fear is both rational and fitting in virtue of the danger the agent faces. If you are happy with danger being a right-kind reason, whatever that is, then leave Fit as Most Right-Kind Reason as it stands. If you are convinced by my argument in chapter 4, then take my claim to be that it is fitting to Φ just in case certain fit-making considerations are present, and that these considerations also make it rational to Φ.

Finally, the view that rationality is determined by the balance of reasons, sometimes referred to as 'substantive rationality', has been defended by Benjamin Kiesewetter (2017) and Errol Lord (2018). Substantive rationality is often contrasted with 'structural rationality'. The paradigmatic example of structural rationality is means-end coherence: if I intend to finish writing this chapter by Sunday evening, and I believe or know that the only way to do this is to spend all day Saturday working on it, then I am rationally required to intend to work on the chapter all day Saturday. However, as Lord points out, not only do we criticise people as irrational for failing to have coherent attitudes, we also describe someone as irrational for having particular attitudes. For instance, if you believe that the earth is flat, you may be criticised for being irrational, where this is not merely a matter of the coherence of your attitudes. It may be that you are also irrational in the sense of being incoherent, perhaps the belief that the earth is flat does not cohere with your other attitudes, but then again it might be perfectly coherent; think of an isolated child who has been educated in a certain way. Nevertheless, it still seems that we can ask about the rationality of your having this particular belief, i.e. 'is it rational for you to believe that the earth is flat?', meaning do you have *good reason* to endorse this belief, as opposed to 'is it rational for you to believe that the earth is flat while simultaneously believing x, y and z?' This brings us to:

> **Rationality As Most Reason**: It is rational for A to Φ just in case A has most reason to Φ.

I have been arguing that ought and rationality are determined by reasons, and that fittingness is determined by fit-making considerations, which for present purposes we can think of in terms of right-kind reasons. If this is true, we have reason to believe that the action one ought to do, and the attitude it is fitting to have, are also actions and attitudes which are rational, thus giving us P1. What about P2? The fact that it is sometimes *irrational* to act in accordance with the ‘objective ought' can be illustrated with the following example. Suppose that I promised to return a book I borrowed from you, and that I ought to return it. Suppose further that the facts are these: if I carelessly leave the book on top of a post box, it will find its way to you (perhaps, unbeknownst to me, your address is printed on the inside cover and, also unbeknownst to me, a very conscientious and benevolent postal worker serves that particular post box), while if I carefully post it by any usual method, it is guaranteed to get lost due to some unforeseeable fault in the postal system. Assume that delivering it in person or by any other means is not an option (Ross, 1939, pp. 156-157). According to objectivism about ought, I ought to carelessly leave your book on top of the post box, as this is the only way to return the book to you. However, this would surely be irrational. If I intend to do what I ought to do, the only rational thing for me to do is to carefully post the book, as I am completely unaware of the various unlikely facts that happen to obtain in our situation. Therefore, P2 with respect to ought is true. In the case of fittingness, consider the following situation from Brad Hooker (2024, p. 181):

> **Letters***:* Amir lived several hundred years ago, before the advent of modern communication technology. His sister Nur lives on the opposite side of the world. She regularly sends him letters in which she explains how well things are going for her: she has a happy family, many friends, plenty of free time to pursue her interests etc. Amir takes great pleasure in the fact that his sister is having such a wonderful life. However, the letters are full of lies. In reality, Nur is miserable; she has no family of her own, few friends, and works 12 hour shifts in an unfulfilling job.

Is Amir’s pleasure fitting? Here we have the familiar sort of tension; on the one hand, his pleasure seems fitting, given what he believes. On the other hand, pleasure seems unfitting, given the facts about his sister’s life, facts which plausibly make sadness, regret or pity more fitting. Now, assuming Amir trusts his sister, pleasure is the only rational response: would we not criticise Amir for being irrational, if, while reading about the joys of his trustworthy sister’s life, he broke down in tears of sadness? It looks as though responding in accordance with objective fit is irrational in Amir’s case, and if something like P1 is true, then objectivism about fittingness must be false.

Zimmerman points out that the argument from rationality is only an argument *against* objectivism and does not entail epistemic perspectivism. Using the term ‘reasonable’ in place of our ‘rational’, and referring to epistemic perspectivism about ought while keeping Ross’s postal example in mind, Zimmerman writes:

> “Even if we accept that carelessly doing what is in fact best is not reasonable, this suffices only to preclude the Objective View. We would arrive at the Subjective View only if it were added that only doing that which one believes best is reasonable” (Zimmerman, 2008, p. 9).

While the argument shows that acting in accordance with objective ought can sometimes be irrational, and therefore not what you really ought to do, it does not show that acting in accordance with the subjective ought is always rational, and therefore what you really ought to do. Furthermore, since the reasons which determine what is substantively rational are evidently constrained by the agent’s epistemic position – it is substantively rational to eat the fish in front of you, even if, unbeknownst to you, it is riddled with salmonella (see Broome, 2007 and Lord, 2018 for discussion) – appealing to rationality to show why ought and fit cannot be objective may seem question-begging (McConnell, 1988, p. 85). However, I think this may be a feature of the argument, rather than a bug: responses which are right, and those which are fitting, are also rational responses *because* the three categories are determined by reasons (or fit-making considerations) which are epistemically constrained. With that said, we should have an independent argument for why the reasons and fit-making features which determine ought and fit respectively are in fact epistemically constrained. Here is one such argument:

2. The Argument from Availability

> P3. For R to be a reason or a fit-making feature which determines that A ought to Φ, or that it is fitting for a to Φ, A must be able to respond in light of R.
>
> P4. For A to be able to respond in light of R, R must be *available* to A.
>
> P5. Availability entails an epistemic constraint.
>
> C2. Therefore, there is an epistemic constraint on reasons and fit-making features.

Taking ought first, reflecting on the *deliberative* nature and function of the category lends initial support to this argument. Ought, which as we have seen can be defined in terms of what the agent has most reason to do, is action-guiding. In a given situation, we consider our options, weigh the reasons in favour of and against each, and come to a conclusion. This conclusion gives us practical guidance; it tells us what to do. It is quite natural to think that in order to be guided by reasons we must be aware of them; they must be within our view. By analogy, think of being guided by a map as you move through the streets of a new city. You cannot follow a map that is not currently in your possession and readily legible. So it is with reasons and oughts, which we might think of as a kind of map for behaviour.

Of course, it is possible to do as you ought for reasons other than those that make it what you ought to do, as when Kant's shopkeeper gives the correct change because doing so will bring him further profits by furnishing a good reputation, rather than because it is honest and fair (Kant, 2018, p.13). However, if it is true that a shopkeeper ought to charge reasonable prices, and this ought fact is determined by the fact that charging reasonable prices is honest and fair, then it must be the case that a shopkeeper can charge reasonable prices *because* doing so is honest and fair; he must be able to perform the action on the basis of those right-making features. This is the intuition behind the argument from availability, and it supports epistemic perspectivism about ought insofar as our shopkeeper must be *aware* of the fact that giving the right change is honest and fair in order to act in light of those facts; the facts must be available to her, where availability entails an epistemic constraint.[30]

As in the case of ought, reflecting on the nature of fittingness indicates some reason to accept the argument from guidance. Recall that fit is a match between an attitude and its object. The metaphors often used to illustrate this matching may at first appear to speak in favour of the thought that epistemic availability is irrelevant. Think of two puzzle pieces fitting together, or the way a foot fits into a shoe. It's the size and shape – objective features unconditioned by any agent's epistemic perspective – that determine fit.

However, this overlooks the attitudinal component of the fit relation. Admiration is not like a foot or a puzzle piece. It is a *reactive* affective attitude, as are many of the other

[30] Lord (2015) provides detailed defence of this kind of argument. See Way and Whiting (2016) for objections.

attitudes often discussed in fittingness theory, such as pleasure, amusement and fear. Reactions are always reactions to some feature of the world the agent is aware of; they are necessarily directed at what is epistemically available to us: a joke we hear, a person whose actions we observe, an event we experience or a memory etc. Furthermore, fittingness at its core is a match between an attitude and the thing the attitude is directed at. If attitudes like amusement are always about or directed at some feature of the world available to the agent, and fittingness is a relation between an attitude and whatever it is directed at, then the fittingness of amusement must be determined by features of the world available to the agent.

Relatedly, being a mental phenomenon, an attitude like admiration cannot be considered in isolation: admiration is felt alongside, and in the context of, a whole range of other mental phenomena including perception, background and occurrent beliefs, memories of past experiences etc. If these associated phenomena are in some sense integral to the attitude, and if they are essentially *about* features of the world that are available to the agent - I can only perceive what is in front of me, or believe propositions I have considered - then it is plausible that in assessing the fit of such attitudes the features of the situation taken into consideration must be restricted to those available to the agent. Finally, fittingness is plausibly a guide for our attitudes in the way that ought guides actions. As we saw when discussing the tension between intuitions about objective and subjective fittingness, whether some action is in fact regrettable, meaning fitting to regret, demands an answer, and so it must be within our reach to find one. The features of an object which make a particular response fitting must be the sorts of things that guide our attitudes: we regret not leaving the pub earlier in light of the fact that failing to do so means we miss the last bus home.

We have considered two arguments in this section, one against objectivism about ought and fit, and one in favour of epistemic perspectivism about the two normative categories. Whether these arguments are ultimately successful in establishing epistemic perspectivism is subsidiary to the primary goal of bringing out the parallels between ought and fit on the issue of the potential role played by ignorance in the instantiation of normative categories. What we have seen is that the very same considerations that can be leveraged by the epistemic perspectivist about ought are also available to her counterpart theorist about fittingness.

5.4. Epistemic Perspectivism about Good

In this section, I show how the considerations brought to bear in support of epistemic perspectivism about ought and fit do *not* lend the same degree of support to epistemic perspectivism about good simpliciter. These considerations are: (1) the intuitive tension that motivates a debate between epistemic perspectivists and objectivists in the first place, (2) the argument from rationality, and (3) the argument from availability. Let's begin with the first and consider:

> **Tension about Good***:* in certain cases, when we ask whether x is good, it seems as though one and the same state of affairs can be both good and bad.

Think again about the Twin Towers case, where Tom the security guard reasonably believes that taking the stairs is more likely to save everyone than taking the elevator, so he directs people to take the stairs, when in fact the elevator would have got them to safety while they all die in the stairwell. Is the state of affairs in which everyone dies trying to escape down the stairs good or bad? This question does not obviously raise a tension between conflicting intuitions. The state of affairs in which everyone dies seems very bad indeed, irrespective of Tom's beliefs about the stairs and the elevator. This is true even if it is a good thing that Tom did his very best to save everyone: acting in accordance with his reasonable beliefs may have been better than acting in accordance with the objective facts, but overall, it is a terrible state of affairs, given that Tom's well-intentioned actions led to tragedy. On the face of it then, Tension about Good is a lot less plausible than the corresponding theses about ought and fit. With that said, there is one variety of value which might raise the kind of tension we are interested in. Here's a case from Shelly Kagan:

> "Imagine a man who dies contented, thinking he has achieved everything he wanted in life: his wife and family love him, he is a respected member of the community, and he has founded a successful business. Or so he thinks. In reality, however, he has been completely deceived: his wife cheated on him, his daughter and son were

> only nice to him so that they would be able to borrow the car, (and) the other members of the community only pretended to respect him…" (Kagan, 1994, p. 311).

Did the man's life go well? On the one hand, it seems as though it did. As far as he was aware, he achieved all his goals. On the other hand however, it seems as though it didn't. He hadn't actually achieved these goals, he was just ignorant of this fact. This case echoes Nozick's famous experience machine (1974, p. 48-9), where we imagine being hooked up to a device that simulates a life filled with the most wonderful, pleasurable experiences imaginable. However, this life is just that, a simulation. Is a life lived through the experience machine a good one? Initially, it sounds fantastic, but then again, something is missing, such that living in the real world seems preferable, i.e. plugging into the experience machine would be a bad thing, comparatively speaking (Bramble, 2016, p. 138). Both of these cases are structurally similar to those in which we found a tension between conflicting answers to questions about ought and fit: there is a kind of goodness – agent-relative in these cases – which seems to be instantiated if we take the agents own perspective to be determinant of value, but uninstantiated when we consider the actual facts of the agent's circumstances. Although I am only interested in denying epistemic perspectivism about good *simpliciter*, rather than the kind of agent-relative value depicted in the above cases, it is worth asking whether the arguments from rationality and availability can support perspectivism about the generic goodness category anyway.

Can we apply Ross's argument from rationality to value? On the face of it, rationality has nothing to do with goodness. Ought and fit are both in some sense or another concerned with the regulation of actions and attitudes, just as rationality is. Value is not like this. Good, bad and the other evaluative categories are primarily used to describe things. We say 'the weather is good today', 'the new Tarantino film is better than his last', or 'the ongoing conflict is a very bad situation indeed'. Of course, non-cognitivists and subjectivists of various stripes will want to explain away this aspect of value as a *mere appearance*, but on the surface at least, 'good' and 'bad' function as ordinary descriptive predicates, as opposed to being guides for rationally-evaluable actions and attitudes.

'Good' is an adjective, whereas 'ought to' is a modal verb[31]. One way we might get a version of the argument from rationality for value off the ground is by adopting a similar strategy to the one used for ought, and reduce goodness to reasons. This yields the following view which we have discussed in previous chapters and which is most closely associated with Thomas Scanlon (1998), and more recently defended by Cosker-Rowland (2019):

> **The Buck-Passing Account of Value***:* x is good if and only if there is most reason to approve of x, and x is bad if and only if there is most reason to disapprove of x.

The first thing to bear in mind is that there seems to be a good deal less consensus about buck-passing than there is about the explanation of ought in terms of reasons. The epistemic perspectivist about value must therefore take on more controversial commitments than her ought counterpart. Nevertheless, armed with the buck-passing account of value, the following argument might be made:

The Argument from Rationality (goodness edition)

> P1*. It cannot be the case that x is good unless it is rational for A to approve of x.
> P2*. Approving in accordance with 'objective-value' is sometimes *irrational*.
> C1*. Therefore, objectivism about value is false.

I think the epistemic perspectivist about value is likely to stumble at the first hurdle. P1* is not plausible. This is because counterexamples are readily apparent. Put yourself back in the shoes of Tom the security guard during the attacks on the Twin Towers.

Suppose he has gathered a group of people whom he is trying to help escape the burning building. Suppose also that there is an emergency evacuation staircase on the side of the building accessed through a window. Believing it to be their best chance of survival, Tom directs the group to take it before moving off to try to help more survivors. As he is leaving, he sees the staircase full of people buckle and collapse, he hears screaming and some seconds later, the horrible crash of twisted metal. He knows that nobody could

[31] It's true that we can assess an individual's evaluative description of some object in terms of rationality, but this is an assessment of their act of describing, or perhaps the description's associated beliefs, rather than a rational assessment of the value of the thing itself.

survive a fall from such a height. As he rushes to the window to see what happened, falling steel beams block his way. He peels off in the other direction, wracked with guilt. Unbeknownst to Tom, the staircase only fell a short distance before catching on a ledge, and the entire group was then rescued by firefighters.

Now, is this outcome good or bad? It seems quite obvious that it is very good. However, it seems equally obvious that it would be *irrational* for Tom to approve of the fate of the group. As far as his evidence goes, it is devastatingly likely that they fell to their deaths. Some extremely negative attitude toward the outcome would therefore be the rational response for him to have, or perhaps a suspension of (dis)approval until he knows for sure what happened. If this is right, then P1* is false. It could here be argued that it is always rational for some perfectly informed, ideal agent to approve of x whenever x is good, but this would go against the spirit of epistemic perspectivism, which is intended to explain how conditions of ignorance and uncertainty can affect the instantiation of normative categories. The preceding example also shows why there is no availability constraint on the features which determine the value of outcomes and states of affairs. Briefly, the perspectivist about value might propose:

The Argument from Availability (goodness edition)

P3*. In order for G to be a feature of x which makes x good, A must be able to approve of x in light of G.

P4*. In order for A to be able to approve of x in light of G, G must be *available* to A.

P5*. Availability entails an epistemic constraint.

C2*. Therefore, there is an epistemic constraint on good-making features.

I think the example we have been working with shows that whether or not some agent has epistemic access to a feature of a situation is irrelevant to whether that feature makes the situation good. Plausibly, the fact that the group is perfectly fine makes the state of affairs in question good, and this is true, if it is true, regardless of whether Tom can approve of the situation in light of this fact. To see in more general terms why the epistemic availability of the good-making features of x is not relevant to the goodness of x, we can ask whether any change will come about in the value of a particular state of affairs when we change a particular agent's epistemic position.

Let's say the state of affairs is this: [all 10 people survived the explosion]. Suppose the explosion happened in a remote area, and that all 10 people are unconscious. Is this state of affairs good? Intuitively, the answer is yes. It is good that everyone survived. Now suppose that the 10 people regain consciousness and discover what has happened. Has the value of [all 10 people survived the explosion] changed in any way? I do not think so, it is just as good as when nobody knew about it. What *has* changed is that some further state of affairs has come about and the value of this state of affairs plausibly leads to a greater degree of overall value: [all 10 people survived the explosion] + [the 10 survivors are aware that they survived the explosion] is greater than the value of [all 10 people survived the explosion], but this does not entail that the value of [all 10 people survived the explosion] has changed. It just means that there is some additional value in the world, namely the goodness of the survivor's knowledge of their survival. Epistemic access to a good x brings about further, different goods. It does not change the value of x.

The thought that epistemic access to the good-making features of some object is irrelevant to the value of the object is widespread, and John Gibbons (2013, p. 119) gives another example: if there is some good in me going to the shops, and getting in the car will get me to the shops, then there is some good in getting in the car. According to Gibbons, the value of me getting in the car does not depend in any way on my believing or having evidence that doing this will get me to the shops. It is good regardless of what I think, assuming the goodness of getting to the shops and the relevant facts.

On top of these considerations, there is a more specific problem with the appeal to an availability argument for the good, and this is that the notion of guidance, which provided a reason to accept the claim that we must be able to respond in light of the features that determine ought and fit, does not do the same work for P3*. This is because it is perfectly coherent to talk about goodness and badness without implying that an agent is guided to do anything or respond in any way. Consequentialists will argue that the value of outcomes determines what we ought to do, which in turn guides action, but this is a substantive claim about what we ought to do, rather than a claim about the intrinsic nature of value. I suspect most people would agree that the abolition of slavery was a very good thing, and that its institution in bygone centuries was very bad. These are claims about the value of historical states of affairs. It is possible to know these evaluative facts, assuming they are facts, and to be aware of the features that make them good and bad, without thereby being led to act in any particular way, or respond with any particular attitude.

This echoes what we said about the phenomenology of value in chapter 2: there seem to be things which are or would be good or bad but which do not impose any demands on agents to pursue or avoid them, or favour or disfavour them, like the existence of happy egrets in a universe otherwise devoid of sentient life. That contemporary governments ought not repeat the atrocities of slavery, or that reparations to the descendents of the victims of the practice ought to be paid, are further claims. Facts about value may play a role in determining ought facts like these, but only when coupled with some consequentialist thesis such that we ought to avoid the bad and promote the good. It is this ought category that is intrinsically action-guiding. The evaluative facts alone are not necessarily the sorts of things that guide our responses.

Of course, this is not to deny that the good *can* be action-guiding, and it evidently often is. We regularly seem to do things *because* it would be good: why did I order the pizza? Because it tastes good. Why do I not join in the anti-immigration protests? Because it would be a terrible thing to do. This also reflects the intuition behind the 'guise of the good' thesis, which says that whenever an agent acts intentionally, or for a reason, she does so in the belief that what she is doing is good in some way (Orsi, 2015). Something similar seems applicable to attitudes like approving: I can approve of pizza for dinner in light of the fact that it is good, and perhaps we only ever approve of things while thinking or believing that it is good. However, even if it is true that we often do x or approve of y in light of the fact that x and y are good, or that we always do x or approve of y thinking or believing that what we are doing or approving of is good, this will not secure the claim that the good is intrinsically response-guiding in the way that ought is.

Recall that the way the argument from availability works in the case of ought and fit is that these two categories are essentially response-guiding, being essentially response-guiding entails that they must be determined by considerations available to the agent, and availability entails an epistemic constraint. If the good is not essentially response-guiding, then this argument will not get off the ground, even if the good is sometimes response-guiding. The existence of goods that cannot possibly or need not guide our responses in any way entails that value is not essentially response-guiding. If this is right, then a major motivation for the argument from availability has no purchase in the evaluative case.

For these reasons, as well as the others discussed in this section, the kinds of considerations that the epistemic perspectivist about ought and fit can appeal to in support of their views do not support epistemic perspectivism about value to the same extent. Goodness simpliciter is therefore quite different from ought and fit with regard to the relevance of a particular agent's epistemic position. At this point, the reader might allow that I have given some reason to believe claims 1 and 2 stated at the beginning of this chapter; that is, 1. Epistemic perspectivism about fit is supported by the same considerations that support epistemic perspectivism about ought, and 2. These considerations do not support epistemic perspectivism about good. But what about my third claim, which says that this result poses a problem for Value-First theories of normative categories? Well, a Value-First theory is a view which holds that all normative categories are grounded by, or explainable in terms of, value or the good.

Suppose you want to argue that goodness is this basic category. One component of this theory is the claim that ought can be explained in terms of good. This is essentially G.E. Moore's view in *Principia Ethica.* For Moore, to say that some action is right is just to say that this action will bring about the greatest amount of good (Moore, 1993, pp. 75-76 and 197). Problems with this view are well-known (see Hurka 2014, pp. 50-52 for discussion). The fact that ought is plausibly constrained by agents' epistemic position, while value is not, does not generally feature among these problems. However, I think the claim that the right action in a given situation just is the action which brings about the most good is incompatible with epistemic perspectivism about the deontic categories, provided objectivism about the good is true. This is because, as we have seen, the right action in the Twin Towers case, according to epistemic perspectivism, is *not* the action which produces the most good: Tom ought to direct the people to take the stairs, even though this leads to everyone dying. This may be the action with the highest 'expected utility', but expected value is not really value at all, it is rather a function of the objective value of the possible outcomes of the set of available actions along with the probabilities of each outcome obtaining when particular actions are performed. Moore did not attempt to define ought in terms of expected utility, and even if he did, this would not be a Value-First theory.

Here is another, more general reason to think that the results of this chapter spell trouble for Value-First accounts of normativity. In their (2021), Francesco Orsi and Andrés G. Garcia level an ‘explanatory objection’ against the fitting-attitude analysis of value. According to the fitting-attitude analysis, x is good if and only if it is fitting to respond to x with a pro-attitude. In a nutshell, Orsi and Garcia’s objection is that if A is reducible to B - if value is reducible to fitting attitudes - then any fact that explains the instantiation of B must also explain the instantiation of A. However, it is not the case that any fact which explains why x is valuable also explains why a pro-attitude is fitting in response to x (Orsi and Garcia, 2021, p. 1209). The reason for this, according to Orsi and Garcia, is that in order to explain the fittingness of an attitude we need to refer to both properties of the attitude itself and the attitudes object, so that fitting admiration is to be explained, inter alia, by the nature of admiring and the objects of admiration. Value, on the other hand, can be explained by the properties of the valuable object without reference to attitudes of any kind (Orsi and Garcia, 2021, p. 1210). If the structure of this argument is sound, then it also works against the exact opposite of the fitting-attitude analysis of value; namely the view that fittingness is reducible to the evaluative.

The problem for any Value-First account of fittingness is this: if something like epistemic perspectivism about fit is true, while objectivism is true of value, then there will be facts which explain the instantiation of evaluative properties which do not also explain the instantiation of the fittingness relation. These will be facts about the way the world is. As we have seen, the value of the various outcomes in the Twin Towers case is determined by and dependent on objective facts alone. For instance, the fact that everyone dies explains why such an outcome is bad. However, this fact cannot explain why some attitude is fitting in response to this outcome because which attitude, if any, is fitting, depends on an agent’s epistemic perspective on the outcome. Distinctively epistemic facts are necessary for explaining the instantiation of the fittingness relation. Not so for value. If we can reject reductive views of normative categories on the explanatory grounds articulated by Orsi and Garcia, then we should reject any reduction of fit to the good.

5.5. Conclusion

I have argued that epistemic perspectivism about fittingness is a viable view. A debate between epistemic perspectivists and objectivists about fit can be motivated in a similar way to the more familiar debate about ought. If two standard arguments for the perspectivist position about ought - namely the argument from rationality and the argument from availability - do in fact lend support to this view, then they also provide the same degree of support to epistemic perspectivism about fittingness. Goodness simpliciter on the other hand is not even the kind of category over which a perspectivist vs. objectivist debate arises, and even if it did, neither the argument from rationality nor the argument from availability can be leveraged by the epistemic perspectivist about good. While this is by no means a knockdown argument against Value-First theories of normative categories, a proponent of such a view will need to explain how it is that the objectivist category of the good grounds the plausibly perspectival ought and fit.

Chapter 6. The Essential Contestability of Affective Concepts

6.1. Introduction

In chapter 5, we saw how the instantiation of fittingness is plausibly constrained by a particular agent's epistemic position. On this sort of view, whether it is fitting for an agent to have a certain attitude in a given situation depends not only on the actual, objective features of her situation, but also on her evidence, beliefs, or knowledge. One feature of the theory of fit that does not sit comfortably with epistemic perspectivism about the category is the seeming truth of the fittingness biconditionals, i.e. admiration is fitting in response to x if and only if x is admirable, and amusement is fitting in response to y if and only if y is amusing. If properties like 'being admirable' and 'being amusing' are determined by purely objective features of the world, unconstrained in any way by the mental states of agent's, then defenders of epistemic perspectivism about fit, which does entail an agential constraint, will have something of an uphill battle in convincing their more objectivist-minded opponents.

In this chapter, I argue that properties like being admirable and being amusing, or more specifically the concepts that represent them, which I will refer to as Affective Concepts (ACs), are Essentially Contestable Concepts (ECCs). If this is right, then a deep indeterminacy animates the concepts which slot into the right-hand side of the fittingness biconditionals. Although the claim that ACs are ECCs does not directly support epistemic perspectivism about fit, by introducing a degree of indeterminacy into our understanding of thick-fittingness categories, an additional kind of agent-relativity can be seen within the aptic family, or so I shall argue.

First discussed in detail by W.B. Gallie in the 1950s, ECCs are thought to populate various areas of social and political discourse. For Gallie, the central, unifying aspect of ECCs is that they are concepts we can and do contest – we dispute their meaning and proper use – where these contests are both irresolvable and "perfectly genuine": there is no simple way of determining which of two or more conceptions is the correct one even while each conception may be sustained by respectable arguments and evidence (Gallie, 1956, p. 169). Drawing on the relevant literature, including Gallie's paper as well as work on political concepts, I develop an account of ECCs which emphasises the above characterisation: ECCs are concepts over which it is possible to engage in genuine, irresolvable disputes, with the potential for conflicting proposals as to the meaning and use of the concept to be equally legitimate, provided they are well-defended with respectable arguments. Here is a question from Alasdair MacIntyre which provides a representative example:

> "Are slavery in Attica in the fifth century B.C., in Ireland in the ninth century A.D., and in Virginia in the nineteenth century A.D. variants of the same institution" (MacIntyre, 1973, p. 1)

This is a question that partly turns on the nature of the concept [slavery][32]. Suppose we think that a component of the concept [slavery] is *involuntary servitude*. And suppose that virtually all instances of slavery in the US in the nineteenth century involved involuntary servitude. Our concept [slavery] can thus be applied to a particular institution once extant in Virginia. Now suppose that in medieval western Europe, a significant proportion of people chose to enter slave-like servitude voluntarily, for various economic and social reasons. Are slavery in the US in the 19th century and in medieval Ireland variants of the same institution? If [slavery] is essentially contestable, then this question does not have a definitive answer. Whether involuntariness is an integral component of the concept [slavery] would be a question open to irresolvable debate that may nonetheless be sustained by respectable arguments and evidence on both sides.

My aim in this chapter is to convince those who find the fittingness biconditionals plausible that if ACs are best understood in terms of the fittingness of their corresponding attitudes, then they are good candidates for ECCs. This entails that the fittingness of an attitude is something over which it is possible to have genuine, irresolvable disputes nonetheless sustainable by respectable arguments and evidence, and a corollary of this is that ACs do not obviously represent objective features of the world. If this is at all convincing, then we will have provided further evidence for the anti-objectivist/subjectivist conclusions about the nature of fittingness expressed in chapter 5.

The chapter proceeds as follows. Section 6.2. clarifies some of the concepts we will use in discussing what ECCs are. Section 6.3. lays out some conditions that a concept must

[32] When I mention concepts in this chapter, I will use square brackets: [...]. When I mention linguistic terms, I will usually use single inverted commas: '...'.

meet in order to be essentially contestable in the way just clarified, while section 6.4. aims to show that the kinds of affective concepts which feature in the fittingness biconditionals meet these criteria. Section 6.5. then concludes.

6.2. Some Clarifications

Before continuing, I want to highlight two aspects of the kind of contestation at play in disputes over ECCs. First, a distinction can be made between what David Plunkett and Tim Sundell call "canonical disputes" on the one hand, and "metalinguistic disputes" on the other. Canonical disputes are arguments in which "speakers disagree over the literally expressed content of what they are saying. Insofar as one views content in terms of propositions, this will amount to thinking that the speakers disagree about the truth of the propositions that they literally express." (Plunkett and Sundell, 2013, p. 6).

For example, if I say 'Donald Trump is a member of the Republican party', and you say 'no, he is not - I'm sure he quit the party when he lost the 2020 election', then our disagreement is of this canonical sort. There is a proposition - 'Donald Trump is a member of the Republican party' - whose content is affirmed by my utterance, and denied by yours. The resolution of the dispute rests on the truth or falsity of the proposition. I say true, you say false, and at most one of us is right, assuming there is no ambiguity about when someone counts as being a member of the party. In such a disagreement, we are not contesting the meaning or proper use of the concept [being a member of the Republican Party]. We agree upon the nature of this concept and use it univocally in our dispute over whether or not Trump is in fact a member of the GOP.

In contrast, suppose we are debating the relative strengths and weaknesses of political ideologies in the 21st century, and I say 'Donald Trump is a Republican', to which you reply 'no he isn't'. Now it might be that the term 'Republican' is also used here to pick out the concept [being a member of the Republican Party], in which case the dispute would again appear to be canonical. However, the context of this conversation suggests that we may not be using the term 'Republican' in the same way, and so we may be employing different concepts. As Plunkett and Sundell point out,

> "regardless of one's precise views in semantics, it should be uncontroversial that at least one crucial type of data for figuring out what a speaker means by a term T are

> facts about the speaker's usage of T — patterns of usage that reflect her disposition to apply that term one way or another, more generally" (Plunkett and Sundell, 2013, p. 16).

Assuming as I will throughout this chapter that the *meaning* of a term is the *concept* one is referring to, the quote above indicates an alternative way to think about our dispute as to whether or not Donald Trump is a Republican.

While I may be disposed to apply the term 'Republican' only to those things which are members of the United States Republican Party, the context dictates that the broader sense of being a disciple of the ideology of Republicanism is more apt. Yet even if we both use it in this second sense, I might be disposed to use the term to describe people who believe in the tenets of Republicanism and self-identify with the label, while you reserve the appellation for people who actively pursue Republican ideals in practice. Or again, even if we are both inclined to use it in the latter way, I may be disposed to call someone who pursues policies akin to p and q a Republican, while you would do no such thing, only referring to someone as a Republican if they pursue policies like r and s. Replacing Plunkett and Sundell's concept [athlete] with our [Republican], they suggest that the dispute arising from your denial of my claim that 'Trump is a Republican'

> "reflects a disagreement about which of two competing concepts, C1 or C2, is more appropriate to the conversation. What is at issue is how the term 'Republican' should be used in this context. In other words, the dispute is about the character of the expression 'Republican'" (Plunkett and Sundell, 2013, p. 17).

On this view, we can both mean different things by the same term, which is to say we can be using different concepts, while still being engaged in a genuine disagreement. When I say 'Trump is a Republican' and you reply that he is not, we may be employing one and the same word with two distinct characters where we thereby proffer different proposals concerning the best use of the term in a specific context. Following Christine Swanton's (1985) terminology, I will refer to the different possible meanings and uses of a term like 'Republican' as *conceptions*. Swanton discusses three ways of understanding the distinction between a concept and a conception, including a particular schema or structure (concept) which can be filled out differently (conception), a common content (concept) which can be variously interpreted (conception), or an "exemplar" (concept) against which a given particular may be measured with different conclusions (conceptions); (Swanton, 1985, pp. 812-13).

There are doubtless many other ways of cashing out the concept-conception distinction; Maite Ezcurdia (1998) provides an insightful discussion of some of the more technical issues involved. In what follows, I employ a thin notion of a *conception* where it amounts to a proposal within a metalinguistic negotiation. It is akin to what Nat Hansen (2021) has called a "metalinguistic proposal", which is similar in spirit to Plunkett and Sundell's "metalinguistic negotiations", although Hansen stresses that he is talking about language, which does not necessarily entail anything about concepts (Hansen, 2021, p. 15). Hansen begins his paper with the following definition: "metalinguistic proposals are speech acts that involve an intention for an audience to come to have a reason to use or understand the use of a linguistic expression in a particular way" (Hansen, 2021, p. 1). I will take this claim about linguistic expressions to be mirrored in the conceptual case. Substituting "concept" for a linguistic expression, and "conception" for a metalinguistic proposal, we can think of conceptions as kinds of metaconceptual proposals.

To return to our dispute over Donald Trump's political credentials, this can be viewed as each of us making alternative conceptual proposals. The contest over the concept [Republican] arises when I propose that we have reason to understand and use [Republican] in way R1, while you disagree, proposing instead that it be understood and used in way R2. As Hansen points out, metalinguistic proposals are important when it comes to "deciding what expression to use to refer to a particular object." (Hansen, 2021, p. 4). The same can be said for conceptual proposals. Whether or not the concept [Republican] should be applied to Donald Trump is a central element of our dispute. When I propose that it should be applied to him for reasons a, b and c, and you say that it should not for reasons x, y and z, we are proposing alternative *conceptions* of the *concept* [Republican].

This leaves us with the first important aspect of the kind of disagreement that characterises the contest in essential contestability. ECCs involve conceptualisations; proposals for how the concept corresponding to a particular term should be understood and used in a given context. When two people engage in a disagreement of the form p vs. not p, they are not necessarily at odds over the literally expressed propositions, i.e. they are not necessarily engaged in a canonical dispute. Instead, they could be engaged in a meta-conceptual dispute about the meaning and use of particular concepts embedded in the propositions. A conceptual dispute of this kind is a disagreement about what features and components ought to make up the concept and how it should be applied; in other words, which is the best conception of the concept. Essential contestability is basically a property a concept has when there is no definitive right answer as to which of two competing, well-supported conceptions is the best.

A second distinction can be drawn between a *contested* concept and a *contestable* one. To see how this works, consider the following two scenarios:

(1) A concept might in fact be contested without being contestable.
(2) A concept might be contestable but not in fact contested.

Not all concepts are open to conflicting yet equally defensible conceptions. For instance, certain concepts corresponding to natural kind terms plausibly fall into this category. Take the concept Gold, and suppose that gold is a material composed of a chemical element with the symbol Au. Now imagine you and I clash over a piece of jewellery in an antique store. You say 'look at this beautiful gold necklace!', to which I reply 'it is beautiful, but it is not gold.' If our dispute is canonical, which is the most likely scenario given the context, then we need only ask the jeweller, or carry out whatever test might be done to determine the presence of gold. If the element with the symbol Au is not found, then you will be forced to concede to being fooled by the necklace; it is not in fact made of gold.

However, even in the knowledge that the necklace does not contain the element in question, you might persist with a metalinguistic dispute and challenge mine and the jeweler's use of the the term and concept [gold], insisting that any yellowish metal counts as gold, or that the concept is applicable to all objects which are generally worn around the neck. In such an eventuality, your proposals for the meaning and use of the concept [gold] are very unlikely to move me or the jeweler. There is not a series of legitimate conceptions of [gold] in this context. There seems to be just one, and that is the conception employed by myself and the expert on precious metals. No matter how much you may contest our usage, the concept is not obviously contestable. Although this is a rough and ready example, and most concepts have vague edges, the point is that essential contestability is not a property a concept acquires by the mere fact that it is contested. The base concept must be such that it allows for a range of alternative conceptions that can be sustained by respectable arguments and evidence. Proposals that [gold] in the context of describing jewelry can be applied to any yellowish metal, or things that go around the neck, are never going to get off the ground; the concept does not seem amenable to alternative, well-supported conceptions.

In discussing scholarly disputes over concepts such as [power] (Steven Lukes, 1974), [justice] (Swanton, 1985), [the rule of law] (Jeremy Waldron, 2002) and [democracy] (Collier et al, 2006), the fact that the concepts under consideration are at the centre of complicated, seemingly intractable debates in academia as well as in public discourse is taken as a prima facie reason for thinking that they are ECCs. However, I am interested in ACs, which though often central to modern moral theory, have not themselves been subject to a great deal of contentious theorisation, although a primary goal of D'Arms and Jacobson's *Rational Sentimentalism* (2023) is to make up for this deficiency. It is the use of concepts like [shameful], [admirable] and [disgusting] in everyday talk and conversation which I will call essentially contestable. Actual contestation among academics is therefore not a primary motivation for this view, although again it does provide some prima facie plausibility. Furthermore, it might be that there is in fact widespread agreement in everyday talk about ACs. In building the case for ACs as ECCs, I will show how this is not the case; we do in fact contest ACs in the manner of proposing alternative conceptions of them all the time. However, nothing much should be taken to hang on the prevalence or not of such disputes. I will attempt to show that ACs are *contestable,* whether or not they are in fact regularly *contested*.

I have made two qualifications to the notion of a disagreement as it figures in essential contestability. The first is that disputes of the form p vs. not p can amount to more than a straightforward disagreement over whether p is true or false. The parties to the dispute may associate different meanings with one and the same term embedded in p, and at a metalinguistic level, their dispute is about which of their proposed meanings, or conceptions, is best. They are contesting the concept itself by proposing alternative conceptions, rather than the truth value of the proposition in which it appears. The second point to keep in mind is that being in fact contested is not strictly relevant to the question of whether a concept is an ECC in the sense studied here. Being in fact contested may provide some initial reason for thinking that a particular concept is contestable, but it is not sufficient, since a concept, such as [gold], can be contested without being contestable.

As a final proviso before getting down to business, note Jeremy Waldron's warning that the notion of ECCs "is clearly vulnerable to overuse" (Waldron, 2002). By way of shirking difficult problems and puzzles which give rise to incompatible schools of thought, it might be tempting to say something like, 'oh well, we are dealing with an ECC after all'. However, even if a particular concept is regularly contested in the metaconceptual sense, and even if such disputes can be sustained with reasonable arguments on opposing sides, we should not automatically conclude that the concept is an ECC, since there may be other explanations. We therefore ought to say more about the specific features that make a concept essentially contestable.

6.3. The Criteria of Essential Contestability

As originally set out by Gallie in his influential paper, an ECC is a concept which conforms to a number of conditions. There are four which are primarily concerned with the contestability of concepts.[33] I borrow Pekka Väyrynen's formulation of the list:

> "(E1) ECCs are 'appraisive', in that they ascribe some kind of 'valued achievement'.
> (E2) ECCs have an 'internally complex character'.
> (E3) This internally complex character admits of a wide range of descriptions depending on the relative significance attributed to the contributions of the component features.

[33] Three more complete Gallie's original list of seven. These include (E5) ECCs are concepts which are used aggressively and defensively against other possible conceptions, (E6) conceptions of ECCs are ultimately derived from "an original exemplar whose authority is acknowledged by all the contestant users of the concept", and (E7) "the probability or plausibility, in appropriate senses of these terms, of the claim that the continuous competition for acknowledgement as between the contestant users of the concept, enables the original exemplar's achievement to be sustained and/or developed in optimum fashion" (Gallie, 1956, p. 180). Given that Gallie himself specified that (E1)-(E4) are "the most important necessary conditions to which any essentially contested concept must comply" (Gallie, 1956, p. 172), and the fact that (E5)-(E7) speak more to actual contestedness, rather than contestability, I restrict my discussion to the first four.

(E4) Such descriptions are 'open' in character, in that they admit of considerable modification in different circumstances in a way that cannot be prescribed or predicted in advance." (Väyrynen, 2014, pp. 473-474)

Väyrynen's paper is dedicated to doing away with (E1), and I am inclined to agree with him. (E1) basically says that in order for a concept to be essentially contestable, that concept must be in some sense *evaluative*. This looks plausible enough when we consider the sorts of broadly political and abstract theoretical concepts that tend to serve as the paradigmatic ECCs. For instance, the example we began with, [republican], may either commend or denigrate the person or practice it is applied to. Shifting to a European context, if we believe in the merits of the *res publica*, to describe someone as a republican is to say something positive about them, whereas if we think that monarchism is the best form of government, calling someone a republican would be to appraise them negatively.

Gallie himself devoted a paper to [art] as an essentially contested concept (Gallie, 1956). [Art] also wears its appraisiveness on its sleeve; when we say of some object that it qualifies as a work of art, we imply a positive evaluation. If you think that a child's drawing does not really count as art, you probably think it is not *good enough* to be art. Väyrynen argues at some length for the claim that being evaluative is not in fact a necessary condition for being an ECC. There are plenty of concepts which do not seem to ascribe any valued achievement in the way that concepts like [republican] and [art] plausibly do, but which make good candidates for ECCs nonetheless. A clear cut case Väyrynen marshals in support of this point are concepts of biological species.

Suppose you and I disagree about whether two birds belong to the same species. There appear to be various components and features of [biological species], such as descent from a common ancestor, the ability of members of the species to reproduce with one another, and similarity of phenotypic and morphological structures etc. If you claim that bird x belongs to species A, but bird y does not, and I disagree, claiming that both birds belong to species A, we may be engaged in a metaconceptual dispute about the concept [species]. Perhaps I am foregrounding morphology and reproductive abilities, which the two birds share, while you think descent from a common ancestor is key, a feature which these two birds lack. We are proposing alternative conceptions of [biological species], and our dispute is about which one is best, proper or correct. According to Väyrynen, disputes over membership to a biological species satisfy (E2)-(E4) – which we will flesh out in a little more detail below – insofar as they "often invoke multiple criteria that can be legitimately weighted in different ways, and thanks to wonders of evolution these matters cannot be prescribed in advance" (Väyrynen, 2014, p. 477). However, although Gallie's notion of a "valued achievement" may be quite broad, encompassing concepts as diverse as [republican] and [art], it would be stretched beyond breaking point if we tried to shoehorn [biological species] into the set of evaluative concepts. There is no positive or negative appraisal associated with claims about birds x and y belonging to species A.

Another example is the concept of [language]. The Yiddish scholar Max Weinreich is credited with giving life to the famous adage "a language is a dialect with an army and navy". In the classification of languages, while it is unlikely that military prowess is given much weight by serious linguists, the phrase does point to the arbitrary nature of the distinction between a language and a dialect. We can imagine a heated debate over the following question: is Scots a dialect of English, or a language unto itself? In wrestling with the issue, parties to the dispute may argue for their favoured conception of [language], with one insisting that because of genetic factor x and morphological features y and z, Scots is not a language, but a dialect of English, while another denies this, claiming that x, y and z do not exhaust the concept [language], but socio-cultural factors a and b, along with lexical features c, are integral to the concept, and because Scots instantiates a, b and c it is in fact a distinct language, not a mere dialect. If this is a dispute over an ECC, which is not implausible, then (E1) is dispensable. Although there may be a certain prestige attached to being a language as opposed to a mere dialect, it would be a push to call the concept [language] in itself evaluative.

So, if a concept can be essentially contestable without being explicitly evaluative, I will bracket (E1). What of (E2)-(E4)? Quoting Gallie, Collier et al point to the close connection between (E2) and (E3):

> "'the internal complexity of a concept makes it plausible that different users may view, or describe, its meaning in different ways'". Together, these conditions tell us that an ECC is a concept which "includes a variety of possible components or

> features. Further, the concept is 'initially variously describable', with the consequence that, a priori, 'there is nothing absurd or contradictory' in the existence of alternative meanings." (Collier et al, 2006, p.217).

This all looks true of the concept [biological species] discussed above: it has a variety of components making it complex, and these components can be weighted in different ways, making it variously describable. To take another example, from the range of paradigmatic political ECCs, Collier et al focus on [democracy]. In defining the term used to name the concept, it is the standard practice of anglophones to turn to the Greek etymology: *dēmos* means people, *krátos* means something like strength, power, or rule, and so democracy means *rule by the people*. If this is all there is to [democracy], then it is far from "internally complex"; it is a perfectly simple concept. However, this would be to fail to distinguish the definition of a word from conceptions of the concept. A term may be simple in a definitional sense, yet still allow for conceptions made up of a variety of possible components and features. Collier et al point to work in comparative politics, quoting O'Donnell and Schmitter (1986), who provide a list of components for their conceptualisation of [democracy] as it might be used to describe particular political entities. Among these components can be found:

> "(1) civil liberties; (2) democratisation, in the specific sense of meaningful electoral competition with universal suffrage, and fair and open elections; (3) democratisation of social institutions and economic processes; and (4) democratisation in the sense of the extension to citizens of substantive benefits and entitlements" (Collier et al, 2006, p. 223).

Whether or not this constitutes a good conception of [democracy] is not important. The point is that a conception of [democracy] including these components is not the same thing as the definition of the term, nor are the two incompatible. O'Donnell and Schmitter can both accept our rudimentary etymological definition of the term 'democracy' as it might appear in a dictionary, while simultaneously promoting their complex conception of [democracy].

To the extent that [democracy] meets (E2) and (E3), it is on track to being a candidate for essential contestability, and there is good reason to believe that it does: the above conception indicates that [democracy] is both internally complex and variously describable. Suppose you and I are discussing whether some state, call it P, is a democracy or not. Our dispute takes the following form: I say 'P is a democracy', to which you reply 'no, P is not a democracy'. Our dispute can be understood in terms of each of us proposing alternative conceptions of [democracy]. Each conception can be defended. We can imagine the dispute continuing as follows: I give decisive weight to component (2), claiming that P's free and fair elections make it a fully fledged democracy, while in your denial that P is a democracy you employ a different conception from mine, one which puts much greater emphasis on the extension to the citizens of substantive benefits and entitlements (4), such that according to your conception of [democracy], P's underperformance with respect to benefits and entitlements means that it is not really democratic, even if it routinely holds free and fair elections. We have attached different weights to the different components of [democracy], giving us internally complex, variously describable conceptions, showing democracy to meet (E2) and (E3).

That leaves (E4): the 'open character' of alternative conceptions. ECCs are concepts which evolve and change over time. Referring again to studies in comparative politics, Collier et al discuss how particular political scientists have recognised the open, changeable character of the concept [democracy] from the 19th century to the 20th:

> "rather than apply the same standard for democracy in both periods, these authors focus on whether a country is democratic *by the norms of that historical period.* Recognizing that empirically democracy has a different meaning in the two eras, they adapt their comparison accordingly" (Collier et al, 2006, p. 224).

(E4) says that there is no way of predicting or prescribing what counts as a legitimate conception of an ECC. This is because such concepts are subject to the vagaries of historical processes including social, cultural and political change. The person who denied that P is a democracy in the above dispute might facetiously concede, 'well, P would indeed be a thriving democracy, if this were the 19th century'. What is more, there is no way for either parties to the dispute to say with any certainty what components may be relevant to the concept [democracy] by the 22nd century. If a concept satisfies (E2)-(E4), as [democracy] plausibly does, then it is an ECC in the way we have distinguished. To these three conditions however, it is important to add a further criterion. According to Gallie,

> "it is quite impossible to find a general principle for deciding which of two contestant uses of an essentially contested concept really 'uses it best'. (Gallie, 1956, p. 189).

That there can be no general principle for determining which conception of an ECC is best does not entail that we cannot have good reason to favour a particular conception in a particular context. For instance, if we are discussing the nature of democracy in the 21st century, a conception of [democracy] which construes the electoral component as extending the vote only to landed males is clearly much *worse* - that is far less reasonable - than one which extends the vote to all adults. Nevertheless, that no such general principle exists to determine which conception is *the best* does lend some support to a stronger principle which I also want to endorse. This is Christine Swanton's Sceptical Thesis. To paraphrase:

> **Sceptical Thesis (ST)**: there is at least one concept C of x which admits of a variety of conceptions of x and which is such that for any conception C* of C, there is no warrant for the belief that C* is the *best* conception of x. (Swanton, 1985, p. 814).

If we accept that there are concepts which satisfy (E2)-(E4), we should accept ST as a corollary, for if a concept is internally complex, variously describable and open in character, it is not clear what sort of grounds there could be for the claim that C* is the best conception of a given ECC: since the components that make up an essentially contestable concept can be weighted in various different ways, producing a set of alternative conceptions, and because these possible weightings cannot be prescribed or predicted in advance, there will be no standard against which we can determine which conception is best, provided the alternative conceptions are equally well-supported by reasons and arguments. As we are concerned in this chapter with ACs, I aim to show that conceptions of concepts like [shameful], [admirable], and [disgusting] satisfy (E2)-(E4), as well as ST. I do this in the next section.

6.4. On the Essential Contestability of Affective Concepts

Christine Tappolet enumerates the following and calls them Affective Concepts (ACs): [admirable], [pride-worthy], [disgusting], [shameful], and [fearsome] (Tappolet, 2016, p. 80). At first blush, what distinguishes these concepts from others is that they are descriptors, designated by adjectives, which are linguistically connected to verbs and nouns which refer to emotional experiences. For Tappolet's selection above, these would be admiration, pride, disgust, shame and fear respectively. This linguistic connection between 'admire' and 'admirable' and 'disgust' and 'disgusting' is one of the cornerstones that underlie the plausibility of the biconditionals that provide definitions of ACs: x is admirable if and only if it is fitting to admire x, and x is disgusting if and only if it is fitting to be disgusted by x. At the heart of these definitions of ACs is the relation of fittingness.

As I hope will be evident by now, there is a strong intuitive pull to the thought that if I admire a brutal despot, or feel ashamed at winning an olympic gold medal (assuming I do so without the aid of performance enhancing drugs), then I have got something *wrong*; my emotional responses are unfitting. If ACs are definable in terms of this relation of fittingness, we need a working understanding of fit in order to address the question of whether ACs are ECCs? Tappolet's own position is that emotions are perceptions of evaluative properties, they present the bearers of such properties as being a certain way, and fittingness is a matter of accurate representation (Tappolet, 2011). On this sort of view, to say that disgust is a fitting response to x is to say that when you feel disgust in response to x you accurately perceive or represent the property 'being disgusting' which x has.

I have not devoted much space to addressing representational accounts of the nature of fittingness in this thesis, mainly because I have been concerned with distinguishing fittingness from other normative categories, and this view of fittingness as accurate emotional perception is not obviously a normative reductivist view. Nevertheless, I will quickly mention two reasons to reject views like Tappolet's which can be found in Howard's (2018). The first is that such a view would seem to tie the nature of fittingness to quite substantial and controversial commitments in other areas. In Tappolet's case specifically, we need to take on her perceptualist account of the emotions. Although there may be good reasons to do so, it is very much a contested thesis in the philosophy of emotions. Dokic and Lemaire (2013) offer a series of objections against the view that emotions are perceptual experiences. Furthermore, Howard quotes a good point made by McHugh and Way in reference to accurate representation accounts of fit in general, which is that it would be surprising to find that our understanding of the fittingness of attitudes was tied to our acceptance of controversial claims about the representational qualities of attitudinal responses (McHugh and Way, 2016, p. 597).

A second reason to reject the accurate representation account of fit is due to Sigrun Svavarsdóttir (2014). The thought here is that when an agent responds to an object in a manner that is unfitting, and misascribes a corresponding property to the object, they are criticisable in a distinctive way. For instance, when I admire some government's genocidal policies and ascribe to them the property of being admirable, I have misplaced my emotions; my emotional response is distinctly 'wrong' in a way that warrants blame or disapprobation. For Svavarsdóttir, this would be difficult to account for if the fittingness of affective attitudes were simply a matter of (in)accurate representation (Howard, 2018, p. 6). The case of perceptual representation is especially clear: I am not generally criticisable for misperceiving a mirage as a nearby oasis.

Another way to understand the nature of fittingness at work in definitions of ACs is to explain it in terms of other normative categories. This is the sort of view I have been arguing against throughout this thesis. In chapter 3, we argued that fittingness is not determined by the balance of reasons, which plausibly does determine deontic status. Roughly, when we are deciding what we *ought to* or *may* do, we weigh up the relevant reasons for the available options. We ought to do what we have most reason to do. There are two other possible verdicts: if the balance of reasons counts against some option, that option is forbidden, and if our reasons neither count in favour of nor against an option, that option is permissible. I argued that permissibility in the case of normative reasons for responses really *matters,* and so the balance of reasons issues in three distinct normatively significant verdicts. Taking disgust and its corresponding AC [disgusting], Peter Goldie has an example of reasons as justifiers of affective responses which shows how fit as most reason might work:

> "the fact that the meat is maggot infested is a reason that justifies your perceptual judgement that the meat is disgusting, and the fact that it is maggot infested will also justify your feeling of disgust (Goldie, 2004, p. 169).

In determining whether disgust is fitting, we weigh the (right-kind) reasons in favour of and against the response. Plausibly, being maggot infested counts quite strongly in favour of disgust, while if the meat is well-cooked and juicy, these sorts of considerations may count against disgust. If the considerations that favour disgust outweigh those against it to some sufficient degree, then disgust is a fitting response, or so an account of fittingness in terms of normative reasons might go. While I think this picture is basically on the right track, the main point of chapter 3 was that it is misdescribed. Features like maggot infestation and juiciness are not in fact normative reasons, they are what I called 'fit-making considerations' which are distinct from reasons insofar as weighing them only issues in two normatively significant verdicts - fitting or unfitting - rather than three. I also argued that explaining fittingness in terms of value is unlikely to succeed. This is because like ought, but unlike value, that the fittingness of attitudes is partly to be explained in terms of the agent's epistemic perspective is a tenable view. The fact that epistemic considerations potentially play a role in determining fit, but not in determining value simpliciter, counts against explanations of fit in terms of value simpliciter.

So, if a representationalist account of fittingness fails, and it is not possible to explain fit in terms of other normative categories such as reasons or the good, and yet we suppose that it is in fact a normative category - as opposed to a matter of non-normative constitutive correctness - then we are left with *fittingness fundamentalism.* As it stands, fittingness fundamentalism is most readily understood in terms of what it is not. To say that an attitude fits a particular object is not to say that it correctly represents the object, nor that there is sufficient reason to have or endorse the attitude, nor that the attitude is good or best. Instead, fittingness is a "basic normative property" (McHugh and Way, 2016, p. 576). Although McHugh and Way make this claim in the context of going on to argue that fittingness is *the* basic normative property insofar as it can explain the nature of values, reasons, and oughts, we can take fit to be basic without committing ourselves to Fittingness-First, since X-Fundamentalism does not entail X-First, as we saw in chapter 2.

The fittingness of an attitude to an object is not reducible to some other normative property or relation between the two, but a first order normative judgement in its own right; when a shoe fits a foot, it just fits. In the rest of this chapter, I will assume that fittingness fundamentalism is true. Of this basic normative category we can ask the following question: what features of a particular object ground the fittingness of an emotional response and the corresponding ascription of an AC? In other words, what are the fit-making considerations of particular affective attitudes?

I now want to suggest that answering this question reveals the way in which ACs are ECCs. To see how, let's stick with the example of a piece of meat and whether or not it is disgusting. In this chapter, we are assuming that ACs like [disgusting] are analysable in terms of fittingness biconditionals, as in x is disgusting if and only if it is fitting to respond to x with disgust. So, what makes disgust fitting? It is reasonable to assume that most if not every human's conception of [disgusting] will count being maggot-infested among the features which partly determine the fittingness of disgust. Whether this is down to the fact that being maggot-infested and rotten reliably trigger disgust, or because being disgusting stands in a supervenience relation to being maggot-infested (Teroni, 2007, p. 410), things get a little murkier when we look at more contentious disgustingness judgments. What about a fresh piece of meat, sprinkled with salt and pepper, sizzling on a frying pan? Consider the following dispute:

(1) That piece of meat sizzling on the pan is disgusting.
(2) No it's not, it's delicious.

I think the model of ECCs outlined in the previous section does a good job of capturing what's going on in this dispute, and the way in which this indicates the essential contestability of [disgusting]. Here's how. The first thing to notice is that the disagreement between the speakers of (1) and (2) - who I will call Jack and Jill respectively - need not be about the truth values of the propositions expressed. That is, it need not be a 'canonical dispute'. It may be that they are engaged in a 'metalinguistic dispute' instead, endorsing different meanings and applications of the concept [disgusting]. Given our analysis of ACs, these distinct meanings and applications can be understood in terms of the fittingness of disgust. Suppose the conversation continues:

Jill: ‘How is it disgusting?’

Jack: ‘It’s a murdered animal being prepared for consumption’.

One way of interpreting this exchange is as Jack proffering and defending a particular conception of [disgusting], where a conception is the set of features he takes to ground the fittingness of disgust, in the same way that a conception of [democracy] includes a set of features that are taken to determine whether some social organisation is democratic. In the case of [disgust], one of the considerations Jack takes to support the fittingness of disgust is the fact that the meat is a slaughtered animal. When Jill denies Jack’s assertion, she can be understood as rejecting this conception, rather than challenging the truth value of what he literally asserts. As the dispute continues, certain moral connotations of ‘murder’ and ‘consumption’ might sway Jill to alter her conception of [disgusting], admitting the features of the meat Jack points to as grounds for the fittingness of disgust. It might even be that the feature Jack refers to is always a pro tanto reason to feel disgust: insofar as the piece of meat on the pan represents the murder and consumption of a sentient lifeform, one has to that extent a reason to feel disgust toward it, whether or not it is fitting to be disgusted by the meat. In light of Jill’s disagreement with Jack, we can suppose that there are other features of the sizzling meat which Jill takes to speak against feeling disgust, and in favour of a kind of gustatory pleasure. Countering Jack’s claim about the grounds for disgust, she might say something like: ‘but it smells so good, and I’m pretty sure it's going to taste even better, and what about all that nutritious protein.’

Again Jack can accept that insofar as the meat smells and tastes good - which he is unlikely to actually think, given his claim that it is disgusting - then one would have pro-tanto reasons to respond with a feeling of gustatory pleasure, rather than disgust. Nevertheless, the dispute remains live, with Jack insisting that the meat is disgusting while Jill continues to deny the same, because each weighs the features in question differently; they endorse different conceptions of [disgusting]: Jack maintains that even if the meat does smell and taste good, being the product of a slaughtered animal outweighs this fact, resulting in an overwhelming reason to conclude that disgust is fitting and to apply the concept [disgusting], while Jill holds that even though the meat is the product of a slaughtered animal, smelling and tasting good outweighs this fact, and so the concept [disgusting] is not necessarily applicable in this instance.

This is where our three key conditions of essential contestability can shed some light. First, ACs like [disgusting] are 'internally complex' (E2) in more ways than one. [Disgusting] is not a simple concept like [gold] or [square] since as we have been stressing throughout, it is definable in terms of at least two components, the attitude of disgust and the normative category fittingness. Furthermore, the fittingness of disgust is itself potentially grounded by a plethora of fit-making considerations, as the dispute between Jack and Jill indicates. These may include things like being maggot infested, having an unpleasant smell, a horrible taste, involving immoral acts in some way, involving death etc. If these various potential components of the fittingness of disgust suggest that [disgusting] meets (E2), then the concept also plausibly meets (E3): "this internally complex character admits of a wide range of descriptions depending on the relative significance attributed to the contributions of the component features" (Väyrynen, 2014, p. 5). The various potential features that contribute to the fittingness of disgust can be afforded different degrees of significance by our two disputants.

For Jack, the fact that the meat is a chunk of a dead animal killed by humans takes centre stage. His conception of [disgusting] will then make most instances of butchery all-in disgusting. Jill on the other hand can acknowledge that butchery counts in favour of feeling disgust, but she can nevertheless attribute little significance to this feature in her conception of [disgusting], taking considerations of smell and taste to outweigh it when it comes to the question of whether disgust is a fitting response to things like sizzling steaks. [Disgusting] then is variously describable, with one possible conception having a distinctively moral flavour – Jack's conception, where butchery is a decisive factor – while another makes [disgusting] more a matter of sensation – Jill's conception, where taste and smell are the weightiest considerations, butchery being of secondary importance. [Disgusting] is thus both internally complex and describable in various different ways, depending on the weight assigned to different fit-making features. According to (E4), "such descriptions are 'open' in character, in that they admit of considerable modification in different circumstances in a way that cannot be prescribed or predicted in advance." (Väyrynen, 2014, p. 5). As we saw in the political literature, this openness is to be understood in terms of the unpredictability of conceptual change over time. If it is plausible that the ancient Greeks used and endorsed a conception of [democracy] which is very different from conceptions used and endorsed today, and if these uses across historical contexts cannot be prescribed or predicted in advance, might the same hold for ACs?

To carry forward Jack and Jill's dispute, and without worrying about actual statistics, let's assume it is currently the case that a significant proportion of the world's population consumes animal meat, and are thus unlikely to feature being an instance of butchery prominently in their conception of [disgusting], even while they assume the pro tanto force of the reason. On a global scale, we can imagine a situation in which changing socio-cultural tides and ecological conditions lead to a wholesale adoption of conceptions of [disgusting] which take butchery as a decisive reason to respond with disgust, such that Jill's conception above might one day appear as wrong-headed as ancient Greek conceptions of [democracy] which saw the concept applied to states that restricted the franchise to an extremely narrow set of property owning men.

The point is that thinking of [disgusting] as essentially contestable means that the concept admits of various different conceptions such that they are open to modification, be it an individual convinced by a friend to assign greater weight to particular grounds for emotions, or conceptual development over centuries of social and linguistic change. (E4) says that there are no hard and fast rules for how such conceptual adjustment is going to go, nor prescriptions regarding how a concept must be conceived. It might be that by 2050, conceptions of [disgusting] which do not recognise butchery as an overwhelming reason to feel disgust will be universally rejected and condemned as barbaric, but then again, maybe not. Finally, according to the Special Thesis, for either of Jack or Jill's conceptions, there is no warrant for the belief that that conception of [disgusting] is the best one. When adjudicating the metaconceptual disagreement between Jack and Jill, where Jack proposes and defends a conception of [disgusting] that emphasises moral considerations, and Jill proposes and defends a conception that emphasises sensory considerations, there is no view from outside, as it were, from which we can judge which conception is best, and this is *because* [disgusting] is variously describable and open in character.

If our discussion of [disgusting] in this section is right, and the concept does indeed satisfy (E2)-(E4), as well as the Special Thesis, then we have good reason to think that it is essentially contestable. This concludes the task of showing that ACs, of which [disgusting] is a representative example, are ECCs. What are the implications of this conclusion for the theory of fittingness? I do not think that the argument of this chapter gives us reason to think that the normative category of fittingness itself is in some sense subjective. This was one upshot of the argument of chapter 5, where we raised the plausibility of the view that fittingness is relativised to an agent's epistemic position. The thin category of fittingness could be perfectly objective even while the 'thick-fittingness categories' analysable in terms of fit, which I have been referring to as Affective Concepts, are indeterminate in the sense entailed by their status as ECCs. However, if we combine the conclusions of chapters 5 and 6, with the former implying that the thin fittingness category is 'subjective' to the extent that it is epistemically constrained, and the latter implying that thick-fittingness categories are indeterminate, then we have some reason to think that an *aptic* family of normative categories which includes both thin and thick-fittingness categories has a number of agent-relative constituents.

To see how this works, suppose we follow Selim Berker (2022 and 2024) and the implications of the anti-reductivist arguments in this thesis and assume that there is an independent aptic family of normative categories which are not explainable in terms of reasons or values. Suppose also that within this family we find (1) a dyadic aptic relation of fit between a response and its object: my disgust *fits* the rotten food, and a converse relation between the object and the response: the rotten food *merits*, *calls for* or *is worthy of* my disgust, (2) a thin monadic aptic property that a response bears when its object merits that response: my disgust is *fitting*; or *merited, called for, suitable, appropriate* etc, and (3) a thick aptic property that an object bears when a certain response is fitting: the rotten food is *disgusting* (Berker, 2024, p. 1053). Now, it might be thought that these aptic categories are objective in the sense that they do not depend for their instantiation on any agent-relative factors. For instance, McHugh and Way maintain:

> "that your epistemic position is normally irrelevant to whether a person is admirable, a joke amusing, or a landscape awesome. As we have seen, these properties at least correspond to what it is fitting to admire, be amused by, and be in

> awe of. Thus what is fitting cannot generally depend on your epistemic position" (McHugh and Way, 2022, p. 78).

If both what is fitting and the thick-fittingness properties like 'being admirable' do not depend on any agent's epistemic position, we might also think that the thick-fittingness properties are more like 'being square' than 'being democratic' in that they are not essentially contestable. Chapters 5 and 6 challenged both of these objectivist assumptions about the aptic categories: we have reason to believe that the dyadic relation of fit does depend on one's epistemic position - at least insofar as we have reason to believe that a similar view about the deliberative ought is true - and we have reason to believe that the thick-fittingness properties are essentially contestable insofar as disgustingness, a representative example, plausibly meets a series of conditions for essential contestability. To the extent that we have reason to believe these anti-objectivist claims, we have reason to believe that some aptic categories are either subjective or indeterminate in nature, despite what is often asserted to the contrary.

6.5. Conclusion

In this chapter, I have endeavoured to show that ACs are ECCs. Toward this end, I developed an account of essential contestability drawing on the theory of 'metalinguistic disputes' as well as the criteria of essential contestability discussed in the literature inspired by Gallie. A concept is essentially contestable if it is internally complex, insofar as it is composed of multiple components; variously describable, insofar as there are a number of different conceptions of the concept depending on the weight given to the various components; and open in character, insofar as these conceptions admit of considerable modification in different circumstances in a way that cannot be prescribed or predicted in advance. On top of these conditions, it was stressed that several conceptions of an ECC can in a sense be equally legitimate in that there are no set criteria for determining which one is 'best'.

I then applied this account to concepts which correspond to the fittingness of particular affective responses; the concept [disgusting] being my chosen example. Disgustingness is the fitting object of disgust; the concept [disgusting] can be defined in terms of fittingness: something is disgusting if and only if it is fitting to be disgusted by it.

The concept [disgusting] admits of various different conceptions insofar as myriad features of a particular object potentially ground the fittingness of the emotion. A dispute of the form x is disgusting versus x is not disgusting may be understood as a metaconceptual dispute, where both parties are proposing different meanings and uses for the concept [disgusting]. The speaker of the first statement proposes a use whereby the fittingness of disgust is grounded by the features of x, while the second speaker denies this, holding an alternative conception of [disgusting] such that the features of x do not ground disgust. This scenario makes [disgusting] a concept which is complex, variously describable and open in character. If this is a plausible interpretation of disagreements about whether some x really is disgusting, admirable, or fearsome; then Affective Concepts are Essentially Contestable, and by extension, the properties picked out by ACs, such as disgustingness, are less than fully objective.

Chapter 7. The Deontic and the Evaluative: A Family Affair

7.1. Introduction

To briefly summarise the progress made so far, this thesis has advanced a number of arguments designed to describe the nature of fittingness while showing that grounding it in some other normative category is fraught with difficulty. We have seen that X-First views have trouble accounting for the phenomenologies of the range of normative categories they purport to explain, including fittingness. We have also seen that fittingness is determined by fit-making considerations, not normative reasons, and that it is importantly distinct from the evaluative categories, given that epistemic perspectivism about fit is a well-supported position, while a similar view about the good is not. Along the way we have made the case for ascribing a number of features to the fittingness relation. For instance, in order to be fit-evaluable a response must meet certain conditions such as having an intentional object and constitutive standards of correctness, and the concepts corresponding to thick-fittingness categories are essentially contestable. This chapter takes a closer look at the methodology we have been employing - that of explaining the nature of normative categories in terms of the features that unite and divide them - this time focusing on the relation between deontic and evaluative categories.

It is widely held, and often tacitly assumed, that the deontic and the evaluative constitute two quite separate and distinct families of normative categories. As I have been doing throughout this thesis, I use the term 'normative' in a rough and ready sense, relying on an intuitive distinction between the normative and the non-normative: terms such as 'good', 'bad', 'ought', 'forbidden' and 'appropriate' are all normative terms, while 'desk', 'heavy' and 'rectangular' are not[34], and by normative *categories* I continue to follow Selim Berker's (2022) usage, such that a normative category is whatever the 'worldly analogs' of normative terms are; things like the evaluative *property* of goodness. By *families* I mean groupings of related normative categories.

The thought that the deontic and the evaluative constitute independent families of normative categories is well attested. Christine Tappolet (2013) provides a roundup of reasons for the claim that good and ought are essentially different kinds of normative concepts. These include the fact that there is more variety in the evaluative family:

[34] See Philip Stratton-Lake (2018).

goodness comes in attributive and predicative forms, as well as sub-dividing into good *for,* good *as,* and good *at* etc (Tappolet, 2013, p. 3). Ought on the other hand does not display this level of versatility, notwithstanding some possible distinctions along these lines, such as the one between an *agent-relative* ought and an *agent-neutral* ought. Another difference Tappolet highlights is their 'domains of application': while evaluative categories apply to states of affairs, actions, persons and objects – we can say 'being at the beach is a good thing', 'going to the beach is good', 'the children are as good as gold' and 'the weather is good today' – deontic categories seem to apply to actions, persons, and perhaps states of affairs, but not to objects: we can say 'we ought to go to the beach', 'I have an obligation to go to the beach', or even 'it ought to be the case that the weather is fine today', but we obviously cannot say that we ought to weather. This indicates some difference between the domains of application of deontic and evaluative categories.

So, that the deontic and evaluative constitute quite separate, independent families of normative categories has a good deal of intuitive appeal. With that said, the distinction is rarely treated in a systematic way, which is why Berker's (2022) recent attempt to do just that provides some valuable insights. The member categories of Berker's proposed deontic family are the following:

required/obligatory/mandatory	*ought to/should/must/have a duty to*
permitted/permissible/allowed	*may/can (on a deontic reading)*
forbidden/impermissible/prohibited	
optional/merely permitted (i.e. permitted but not required)	

The categories constituting the evaluative family with which Berker contrasts the deontic are given in a table that looks like this:

good	*better than*	*strongly best (i.e. better than all the rest)*
bad	*worse than*	*weakly best (i.e. at least as good as all the rest)*
neutral/indifferent	*equal in value to*	*strongly worst*
	on a par with (?)	*weakly worst*
	at least as good as	*second best*
	no worse than	*good enough*
	at least as bad as	*better than most*
	no better than	

- (Berker, 2022, p. 22)

Now, Berker wants to establish in a systematic way that the deontic and the evaluative are indeed independent, separate families of normative categories. This is done in the context of arguing for the claim that along with the deontic and the evaluative, the aptic or fitting constitutes a third family of independent normative categories, although he does not think this entails that none of these can be explanatorily prior to another. For instance, an analysis of the good in terms of fitting attitudes is compatible with the evaluative and the fitting being distinct kinds of categories (Berker, 2022, p. 47). Naturally, I am very sympathetic with Berker's general conclusion that there are important differences between different kinds of normative categories. However, I am not convinced by how he gets there, and I think the relations between normative categories within the putative families are more complex than he would have us believe. I argue for this position by interrogating Berker's arguments for the claim that the deontic and the evaluative constitute entirely distinct families. His method involves proposing five distinguishing features; features that deontic categories have but evaluative categories do not, or vice versa. The goal is to show that deontic categories and evaluative categories have distinctive logical and inter-relational structures, and it is in virtue of these different structures that they comprise two distinct families.

The aim of this chapter is to show how Berker's proposed features are in fact shared by categories across the two families. If this is right, then placing the deontic and evaluative categories listed above within two completely separate families solely on the basis of their logical structure is a mistake. I do not want to deny that there are relevant distinctions to draw between the deontic and the evaluative. In fact, I argue that the evidence brought to bear on the similarities between their underlying logical structures supports a particular approach to the relations between certain deontic and evaluative categories. This is the 'reason-guiding role' approach. My argument here is inspired by the work of Ralph Wedgwood, especially his (2001 and 2018). Wedgwood thinks that the semantic value of moral terms is given by the role they play in addressing practical reasoning problems[35].

Reasoning involves the transition from premise attitudes to conclusion attitudes, as in the movement from a desire for a coffee plus the belief that the café around the corner does a good Americano, to an intention to go to the café around the corner. The notion of a particular concept's *role* in reasoning can be taken quite broadly, encompassing various aspects of how a given concept guides the reasoner. This can include rules about which moves are legitimate or illegitimate when reasoning from or to beliefs or intentions containing those concepts, and what kinds of situations we use certain concepts to reason about.

While Wedgwood employs this framework to establish a semantics for moral terms, I will use it as a tool for understanding the relationship between the deontic and evaluative families. The thought is that the similarities in logical structure that emerge from consideration of Berker's features can be explained in terms of the fact that both deontic and evaluative categories guide reasoning in particular ways, while the differences that

[35] This account is couched in the language of 'concepts' and 'meaning', similar to the previous chapter, whereas in the other chapters we mostly spoke in terms of normative categories (properties and relations). However, Wedgwood writes that "it is precisely this action-guiding conceptual role of normative concepts that determines what property or relation these normative concepts represent or refer to" (Wedgwood, 2018, p.35). I will not necessarily endorse this view, and I certainly won't defend it. However, as in the previous chapter, I will use the notion of a 'conceptual role' to talk about the nature of normative properties and relations without worrying too much about any discrepancies between the metaphysical and the conceptual.

remain can be explained in terms of the more fine-grained reason-guiding that specific categories are involved in. If I am right, the significance is this: we should not think of the deontic and evaluative as two separate families distinguished by deep logical and inter-relational differences, but rather as two sides of the same coin distinguished by the specific reasoning problems addressed by different categories within the two families. The chapter proceeds as follows. Sections 7.2-7.5 consider each of Berker's proposed distinguishing features in turn. It is argued that upon closer inspection all five features can be found among both deontic and evaluative categories. Section 7.6 then offers an account of these results in terms of the roles that the categories play in reasoning. Section 7.7 briefly concludes.

7.2. Duality

The first proposed distinguishing feature can be presented as follows:

1. **Duality***:* certain pairs of deontic categories are interdefinable; this is not the case for certain pairs of evaluative categories.

Here is Berker's leading example:

> "D1. ϕ-ing is required =df not-ϕ-ing is not permitted.
> D2. ϕ-ing is permitted =df not-ϕ-ing is not required." (Berker, 2022, p. 26)

The meanings of these two deontic categories - required and permitted - are clearly connected, for to say that an action is required is the very same as saying that not doing it is not permitted. Berker goes on to present the interdefinability relation between a further selection of his deontic categories:

> "D3. ϕ-ing is forbidden =df not-ϕ-ing is required.
> D4. ϕ-ing is required =df not-ϕ-ing is forbidden.
> D5. ϕ-ing is forbidden =df ϕ-ing is not permitted.
> D6. ϕ-ing is permitted =df ϕ-ing is not forbidden" (Berker, 2022, p. 27)

For Berker, this interdefinability he calls 'duality' is one feature that marks the deontic categories as a normative family distinct from the evaluative because good, bad, and their kin are not interdefinable in an analogous way. To take the example of the pair good/bad, the following does not hold:

> E1. x is good =df not-x is bad.

Berker explains why E1 does not hold:

> "for one thing, we need not hold that the sorts of things that can be good are also the sorts of things that can be negated, so 'not-x' might not make sense for some relevant values of x. And for another, even when 'not-x' does make sense (as it arguably does when x is an action or a state of affairs), that x is good is compatible with not-x being good or bad or neutral: sometimes it is good both to do something and not to do it, and other times it is good to do something without its being actively bad or actively good to refrain from doing it" (Berker, 2022, p. 26).

We can see this play out if we consider certain kinds of education: knowing that Oliver Cromwell sacked the town of Drogheda and massacred its inhabitants in 1649 might be good in the sense that knowledge is good in itself, or insofar as it will help me pass my history exam, but 'not knowing' this fact is not therefore bad by definition. On the contrary, not knowing about the sacking of Drogheda might be *good* in the sense that ignorance is bliss; I need not suffer the sadness and regret that comes with knowledge of the episode, or it might be bad in the sense that a failure to recognise historical crimes increases the likelihood of repeating them. Or again it might be neither good nor bad, but evaluatively neutral; in the way that knowing the exact number of blades of grass in your back garden seems evaluatively neutral. This is just to say that 'x is good' does not necessarily equate to 'not-x is bad'; good and bad do not appear to be definitional duals.

However, Berker does not offer any examples of his own for the claim that both x and not-x can be good, and it might be noticed that the examples I have given appealed to different senses of good. The evaluative 'good' and 'bad' are multivocal terms. We can distinguish, at the very least, between the categories of intrinsically good/bad, instrumentally good/bad, good/bad-for and good/bad-as.[36] I would suggest that Berker's claim that good and bad are not duals of one another trades on an ambiguity between these different senses; while it may be the case that x can be good in one sense and not-x be good in another – entailing the falsity of E1 – the following alteration may not be so easily dismissed:

E1*. x is y-good =df not-x is y-bad.

For example, if we continue with the case of knowing a historical fact, and we stipulate that what we mean when we say that this knowledge is good is that it is *instrumentally good* insofar as it will help me pass my history exam, then it seems we can pick out a dual, *instrumentally bad,* which is interdefinable with instrumentally good in the way that Berker denies: if it is the case that knowing that Cromwell sacked Drogheda in 1649 is instrumentally good qua my exam prospects, then it must be the case that not knowing this fact is instrumentally bad in relation to the same end; lacking this knowledge will hurt my chances of passing the exam. In general, if some action or state of affairs is good in the sense that it contributes to a desired end, then the negation of that action or state of affairs cannot be good in the same sense, it is in fact bad in that particular sense by definition. Perhaps unqualified categories of good and bad do not display duality in the way that unqualified categories of forbidden and required do, but unlike forbidden and required, good and bad, which come in a variety of different flavours, demand qualification, and when we look at biconditionals containing these qualified evaluative categories – x is instrumentally good iff not-x is instrumentally bad – they do appear to be duals of one another, and so duality does not help to distinguish the deontic family of categories from the evaluative.

At this point, one might object to my rendering of the duality of instrumental goodness and instrumental badness. Suppose I am hungry and want to sate my appetite. Eating a burger is instrumentally good in this situation, but, it might be suggested, not eating a burger is not instrumentally bad by definition, because I could have a pizza, or a nice salad. However, I do not think this belies the duality of instrumental goodness and badness. 'Not eating a burger' remains instrumentally bad insofar as my goal is to sate my

[36] A.C. Ewing lists nine different senses of good (Ewing ,1948, pp. 112-116).

appetite, and not eating a burger will not sate my appetite. This is true regardless of the availability of other culinary options. When there are a number of means which are equally good for achieving some end, it is still the case that not ϕ-ing, where ϕ is any one of those means, is bad insofar as not ϕ-ing does not contribute to the achievement of the end.

Now, the defender of the deontic-evaluative split along the fault line of *Duality* might want to point out that in arguing for the duality of evaluative categories, I have unwittingly uncovered another significant difference between the evaluative and the deontic: the deontic categories do not require qualification to produce logical duals, whereas the evaluative ones do. This does not strike me as a promising route for Berker to take. Duality simply says that certain deontic pairs are interdefinable, while "no such relation exists among the evaluative categories" (Berker, 2022, p.25). The fact that we have to look to the more fine-grained members of the evaluative family to find duals hardly shows that duality is not a feature of any evaluative categories. Rather, it is a natural consequence of the fact that the evaluative family has a far greater number of member categories, some more complex than others, when compared with the deontic family, which contains a relatively small number of simple categories. This last point is taken on by Tappolet (2013) as an important difference between the evaluative and the deontic: the rich diversity of fine-grained evaluative categories contrasts with the limited set of basic deontic ones. Again I am not sure this helps Berker's case. Recall that we are interested in the interrelations and logical connections between the categories themselves. What has the size of the putative families got to do with this?

To stretch the 'family' metaphor somewhat, suppose I have ten children and my sister has two. Both sets of children belong to the same extended family; they are cousins. The quantity of individuals on either side has no bearing on their family connection, which is determined by the fact that they are the children of siblings. I would suggest that the same is true of deontic and evaluative categories. The fact that there are a greater number of evaluative than deontic categories tells us very little about the logical relations between the categories themselves. For this we look to features like duality. Size does not matter, but duality does, at least as far as Berker's project of distinguishing between the deontic and the evaluative on purely logical grounds goes. Berker's claim is a big one. He says that the deontic family contains definitional duals, while the evaluative does not. The point of comparison here is a logical relation. If I am right, and there *are* evaluative categories which are duals of one another, then the same logical relation can be found in both families. The fact that in the evaluative case, the duals are quite specific, fine-grained categories, while the deontic duals are basic, seems to me irrelevant to the question of whether duality is a feature of the evaluative and the deontic families taken as wholes. Furthermore, brute facts about the quantity and variety (or lack thereof) of categories in the two groupings is a superficial difference which does not tell us anything especially deep about the logical nature of the categories.

Finally, even if we concede that their size and diversity represent significant differences between the deontic and the evaluative families, this would not be a problem for the present argument. As noted in the introduction, it may well be the case that there are important distinctions to draw between the two families, distinctions which will be addressed in section 6 below. My point here is just that duality is not exclusively a feature of the deontic family, and so it cannot serve to distinguish the deontic from the evaluative.

7.3. Opposition

The second distinguishing feature for Berker is:

2. **Opposition***:* the evaluative categories good and bad are *polar opposites*, but no deontic categories are related in this way.

Comparing the pairs forbidden/permitted and good/bad, this is what Berker has in mind:

> "to be forbidden is to be non-permitted, and to be permitted is to be non-forbidden. By contrast, goodness and badness are what I call polar opposites: each is the inversely charged flipside of the other, not its mere lack or absence. To be bad is to be anti-good, not to be non-good, and to be good is to be anti-bad, not to be non-bad" (Berker, 2022, p. 27).

Another way of putting this is to say that forbidden and permitted are *contradictories*, whereas bad and good are non-contradictory opposites, or *contraries*. Forbidden and permitted are contradictories insofar as the following is true:

F1. ϕ-ing is forbidden if and only if ϕ-ing is non-permitted.

This is to say that the one deontic category is equivalent to the negation of the other; forbiddenness is the negation of permittedness, and vice versa. Furthermore, one of the predicates 'is forbidden' and 'is permitted' must be true and the other false. For any action ϕ, ϕ-ing is either forbidden or it is permitted, it cannot be both and it cannot be neither. To this last point it might be replied that ϕ-ing can be neither forbidden nor permitted, as it can fall under a third category, required. But a requirement entails permission. If you are required to love thy neighbour, then it is also permitted to love thy neighbour, noticing of course that permission here is distinct from *merely permissible*, or optional, which is not entailed by a requirement. Good and bad are not contradictories in this sense because G1 is not true:

G1. x is good if and only if x is non-bad.

It is not the case that x is good if and only if x is non-bad because for some x which is non-bad, it may be neutral, having no particular value one way or the other, and so be neither good nor bad. An example might be the fact that it is 2pm. Assuming I am not running late for any appointments or deadlines, the fact that it is 2 o'clock is not a bad thing, but neither is it a good thing, it is evaluatively neutral. So, while it is true that good and bad are opposites, they are non-contradictory opposites. What is more, for Berker, the opposition that evaluative categories do exhibit is of a distinctive kind he calls *polar opposition*: goodness is anti-badness, and badness is anti-goodness. We might think of this non-contradictory, polar opposition by analogy with the pair happy/sad. Happiness and sadness are not contradictories. Being not-sad does not imply that I am happy. Perhaps I'm feeling angry, envious, or just apathetic. The relation between happiness and sadness is better understood like this: they sit at opposite poles of an emotional continuum, with something like 'being content' on the side closer to happiness, and 'being a little upset' closer to the pole of sadness. Happiness and sadness are the 'inversely charged flipsides' of one another, to use Berker's useful locution. The same is true of good and bad. Along the axis of evaluation, goodness occupies one pole, badness the other.

Berker's claim is then that no deontic categories are related in this way. But what about forbidden and required? The first thing to notice is that they are not contradictories, as the following does not hold:

F2. ϕ-ing is forbidden if and only if ϕ-ing is non-required.

It is not the case that ϕ-ing is forbidden if and only if ϕ-ing is non-required because for some ϕ which is non-required, it may be merely permitted, which is to say optional, and therefore not forbidden. There are countless examples. There is no requirement for me to eat an apple right now, but it certainly isn't forbidden either. I can if I want, but there is no obligation either way. This entails that for any ϕ, it is not the case that ϕ-ing must be either required or forbidden. In ordinary circumstances, eating an apple is neither required nor forbidden; it is entirely optional. Forbidden and required thus exhibit contrariety in much the same way that good/bad and happy/sad do. They are polar opposites, with required being at one pole, forbidden at the other, and optional in between. Requiredness is anti-forbiddenness, not non-forbiddenness and forbiddenness is anti-requiredness, not non-requiredness. Berker anticipates this move, writing that "forbiddenness is neither anti-requiredness nor anti-permittedness; rather, it is related to requiredness via D3." (Berker, 2022, p. 27). I do not think this is an adequate response, primarily because two categories can be logically related in more than one way. D3 is compatible with forbiddenness being anti-requiredness. To see this, recall that D3 was Berker's illustration of the duality of forbidden and required:

D3: ϕ-ing is forbidden =df not-ϕ-ing is required.

With Duality, we modify the subject of the proposition, as in 'ϕ-ing' to 'not-ϕ-ing', whereas the relation of Opposition involves modifying the predicate, as in 'forbidden' to 'anti-forbidden' or 'non-forbidden'. I see no reason to think that forbidden and required cannot be both duals and contraries of each other: ϕ-ing is forbidden is equivalent to not-ϕ-ing is required *and* ϕ-ing is forbidden is equivalent to ϕ-ing is anti-required. To illustrate, it can be simultaneously true that stealing is forbidden means that refraining from stealing is required, and that stealing is forbidden means that stealing is anti-required. I conclude that polar opposition, or contrariety, can be found among both the evaluative and the deontic categories.

7.4. Gradability and Neutrality

I now take Berker's third and fourth proposed distinguishing features together:

3. **Gradability***:* deontic categories are not gradable, while evaluative categories are.
4. **Neutrality***:* pairs of deontic categories do not have a third, neutral state in between them, while pairs of evaluative categories do.

To begin with 4, the falsity of Neutrality follows quite straightforwardly from our rejection of Opposition. First of all, it is clear that there is a neutral state between good and bad. Facts about the time of day, as we noticed above, are ordinarily neither good nor bad, but neutral. Indeed, there is a wide vocabulary of commonplace terms for describing things that fall within this evaluatively neutral space:

> A: How are you?
> B: I'm fine
> A: How was your journey?
> B: It was ok.
> A: What was the film like?
> B: It was alright. Decent, but not particularly good.

In none of these conversations does B apply the category bad, but nor has anything been described as good. The person's emotional state, the journey, and the film are neither good nor bad, but somewhere in between. This much Berker is happy with. His contention is that neutrality cannot be found among the deontic categories. However, from the discussion in section 3, it should be equally clear that there is a neutral state between required and forbidden. These two categories are contraries, not contradictories. Some actions are neither required nor forbidden, but merely permitted, which is to say they are optional. This seemed quite obvious when we considered the action 'eating an apple'. In ordinary circumstances, if I were to ask whether I am required to eat an apple, the answer would be no, and if I were to ask whether I am forbidden from eating it, the answer would also be no.

The implication is that eating the apple is optional, or merely permitted, a deontic category lying between required and forbidden. Berker does have a response to this line of thought. He writes:

> "requiredness and forbiddenness do not fall on a spectrum with mere-permittedness lying in between. If I can save someone's life by giving them either antidote A or antidote B, we would never say, 'Giving that person antidote A is neither required nor forbidden but in between,' in the way in which we might say of a different situation, 'that outcome is neither good nor bad but in between.'" (Berker, 2022, p. 28).

I think Berker's intuition here can be dealt with. If it seems strange to say 'giving that person antidote A is neither required nor forbidden but in between', this is only because giving that person *either* antidote A *or* B is *required*. It is tempting to see the requirement to give A or B as carrying over to both options i.e. there is a requirement to give antidote A *and* a requirement to give antidote B. But in the case as stipulated, this simply is not true. The requirement is to give just one. It may even be forbidden to give both A and B, insofar as this could do more harm than giving neither. Therefore, if *either* A *or* B will save the patient, giving A is neither required nor forbidden, and giving B is neither required nor forbidden, but giving A *or B* is required. The existence of this latter requirement explains away Berker's intuition in the quote above. We should thus see no problem in accepting that an action's being neither forbidden nor required represents a kind of deontic neutrality.

Another line of response Berker might pursue is to say that even if optional is the neutral category between required and forbidden, there is no neutral category between forbidden and permitted. After all, we have accepted that they are contradictories, which is to say that for any conceivable action, it is either forbidden or permitted, never neither. Again I would highlight a dialectical point here. Berker's overall claim is that the evaluative *family* and the deontic *family* have distinctive logics, not merely that some specific categories within the families do. This implies that all categories are up for grabs when it comes to the question of whether some feature can be found among the categories of a particular family. I therefore do not think the fact that there is a neutral category between good and bad but there is no neutral category between forbidden and permitted undermines my argument, because we do find a neutral point between another pair of deontic categories, namely forbidden and required.

Turning to Gradability, it is quite obvious that the evaluative categories are gradable. We often say things like ‘my journey to work today was considerably worse than yesterday’s’, or ‘the new Scorsese film is much better than his previous one’. Better and worse are graded evaluative categories. Berker says he has trouble seeing how deontic categories could be gradable in this way, but it seems quite clear to me that committing mass murder is *more forbidden* than posting a mildly offensive joke on the platform formerly known as Twitter. I also know that I am required to do my best to save a drowning child should I happen across such a situation, but surely I am far *less required* to pick up the small amount of litter I left on the beach. ‘More forbidden’ and ‘less required’ are graded deontic categories. This is echoed and buttressed by Thomas Hurka’s remarks on the gradability of moral wrongness - another deontic category: “that one act is more seriously wrong than another is often intuitively compelling in itself; thus it seems self-evident that murder is morally worse than breaking a promise” (Hurka, 2019, p. 42).

Christine Tappolet claims to have a response to the intuitive sense in which deontic categories can come in degrees. We are asked to consider a situation in which we must make a choice between two forbidden acts, lying and killing. If we accept that we ought to lie, rather than kill, it seems reasonable to suppose that this is true because lying is ‘less forbidden than’ killing, which is to say that not killing is ‘more required than’ not lying. For Tappolet, this apparent gradability can be explained in terms of *priority*, rather than degree, where priority is glossed as follows:

> “when we say that we should lie rather than kill, we do not mean that killing is ‘more forbidden’ than lying; what we mean is that in the case of a conflict the requirement not to kill *overrides* the requirement not to lie” (Tappolet, 2013, p. 5, my emphasis).

I agree that there is a difference between priority and degree. The fact that my name is prior to yours on a roll call is clearly not a matter of degree in any ordinary sense. Suppose we are allocating time slots in a seminar series by letting the participants choose in alphabetical order, so that Aaron Aaronson gets first choice and so on. If my surname begins with M while yours begins with T, then I have priority over you. But this is simply in virtue of our names and the convention of allocating slots in alphabetical order, it has nothing to do with any disparities of degree between us. However, for x to *override* y strikes me more as a matter of degree than this non-graded priority. What does 'override' mean if not something like 'outweigh'? The requirement not to kill someone overrides the requirement not to lie because the former is a stronger, weightier requirement. This is what explains the fact that we should tell a lie rather than kill, were we compelled to do one or the other. If the sense of overriding at issue in a conflict between two requirements involves gradable notions such as weight and strength, then appealing to non-gradable priority, as in alphabetical order, is specious at best. Conflict between two requirements confirms the claim that some deontic categories are gradable, rather than undercutting it. To conclude this section, Berker has suggested that gradability and neutrality can be found among the evaluative categories but not the deontic. However, I think we have good reason to allow that there are deontic categories which exhibit both features.

7.5. Alternatives Dependence

Here is Berker's fifth and final distinguishing feature:

5. **Alternatives Dependence***:* "whether ϕing is permitted or required depends, in part, on the normatively relevant properties of all the alternatives to ϕ-ing" (Berker, 2022, p. 29). Whether x is good or bad on the other hand does not depend on alternatives to x.

Expanding on the premise, Berker points out that "an action that is required in one situation might have the same (non-comparative) properties and yet be forbidden in another, because certain additional courses of action are available" (Berker, 2022, p. 29). To see how this works, let's take a classic trolley case. Suppose there is a runaway trolley careering toward five people tied to the track, while I stand next to a lever. If I pull the lever, the trolley will be diverted to a second track where one person is sadly tied. Granting for the sake of argument that some form of consequentialism is true, if my only two options are to do nothing and allow the train to kill the five, or pull the leaver, killing the one to save the five, then I am required to pull the lever. However, if there is a switch beyond the lever which will stop the trolley dead in its tracks, this third option is the one that is required, while merely pulling the lever becomes forbidden. This is to say that what I am required or forbidden to do is alternatives dependent.

Berker insists that evaluative categories do not display alternatives dependence in this way. I fail to see why not, for it seems that exactly the same reasoning applies to the 'moral goodness' of pulling the lever. If it is indeed morally good to pull the lever and kill one to save the five, then in a scenario where the third option of flicking a switch to stop the trolley is available, pulling the lever is no longer a morally good action, but a bad one, because I will kill one when I could have saved all six. The goodness or badness of my actions at the trolley track thus depends on the possible actions I can take. Alternatives dependence is therefore a feature of both deontic and evaluative categories.

It might be objected that we have merely shown that the evaluation of *actions* is alternatives dependent, and that this tells us nothing about the evaluation of *things*. Consider a selection of different food options, an Italian buffet let's say. There may be a number of good options to choose from; lasagna, risotto, gnocchi etc. Would we say that the quality of the lasagna is dependent on the alternatives? It seems abundantly clear that we would not. Plausibly, the goodness or badness of the lasagna is determined solely by the properties of the lasagna; the methods of preparation and cooking, the quality and freshness of the ingredients and so on. If the risotto were to run out, this would have no effect on the value of the lasagna, or conversely, if a caterer were to add a delicious bowl of spaghetti alle vongole, this would have no effect on the quality of the lasagna, although it will likely have a substantial effect on my preferences. The lasagna is good (or bad) regardless of whatever else happens to be on the table. In a case like this, it looks as though good and bad are not alternatives dependent.

I do not think this case hurts the argument I am advancing for two reasons. Firstly, it might be that the categories good and bad are subject to different conditions when applied to actions versus things, this being another consequence of the great diversity and depth of the evaluative family which we noticed in our discussion of Duality. There I conceded that good and bad *simpliciter* may not be duals of one another, but that the more fine-grained *instrumentally* good and bad are. If Berker's claim is that Duality is a feature of the deontic family but not of the evaluative because it cannot be found among *any* evaluative categories, then showing that it can be found among instrumental value categories, if not unqualified ones, is enough to reject the claim. I think the same reasoning is valid with respect to Alternatives Dependence. If one wants to say that Alternatives Dependence is a feature of the deontic family but not the evaluative family, this can be rejected by pointing out that good and bad as applied to actions *are* alternatives dependent, whether or not the same is true of good and bad as applied to objects.

Secondly, I think we can in fact find alternatives dependent evaluative categories which apply to objects if we root around in the family a bit more, namely comparatives and superlatives. Is the pasta better than the lasagna? Does the excessive amount of salt spoil the risotto, making it worse than the otherwise bland gnocchi? And which dish is the best? Answers to questions such as these clearly depend on the alternatives available on the buffet table. Even if the categories good and bad as applied to objects such as food are not alternatives dependent, better and worse, and the best and the worst, which also belong within the evaluative family, seem to be. To take another example, suppose we are evaluating the players on a football team. It seems perfectly possible to determine the quality of an individual player in isolation from the rest of the squad: we can put Roy through his paces and assess how good he is at dribbling, tackling, and shooting etc, and come to a conclusion about his goodness as a football player. However, there are lots of comparative evaluative properties of Roy which can only be determined by assessing how he performs relative to the other players on the team. Who is he better than? Who is he worse than? And where does Roy fall in an ordering of the players from best to worst? These evaluative properties depend on alternatives to Roy. If this is true, we have another reason to think that Alternatives Dependence can be found among both the deontic and evaluative categories.

7.6. The Deontic, the Evaluative and Reason-Guiding Roles

Thus far, I have been arguing that a series of logical and inter-relational features which Selim Berker presents as grounds for a sharp distinction between deontic and evaluative categories are in fact shared across the two proposed families. This suggests that we cannot neatly separate the deontic and the evaluative purely on the basis of logical structure. This conclusion raises at least two questions. First, what explains the deep similarities between certain deontic and evaluative categories? And second, what are the real differences that survive the arguments of the previous sections, and how are these to be explained. I now want to answer these two questions with the following two-part proposal:

(1) The fact that the deontic and evaluative families share certain logical and inter-relational features can be explained in terms of the general reason-guiding roles that both kinds of category play.
(2) There are further non-logical or inter-relational differences between the two families, and this fact, along with the fact that there remain certain logical and inter-relational differences among specific members of the two families, can be explained in terms of the categories' more specific reason-guiding roles.

What do I mean by a reason-guiding role? Well, when deciding what to do, or when evaluating something, we engage in reasoning. This involves moving from beliefs and/or intentions, the premises, to further beliefs and/or intentions, the conclusion (or perhaps discarding, revising, or suspending judgement concerning the premises)[37]. For example, suppose I want to go to a football match, and in order to get there I need to take the number 37 bus. The process of practical reasoning in this case leads me from two premises - my intention to go to the match plus the belief that the number 37 bus will take me there - to a conclusion: an intention to leave now in order to catch the 37 bus (McHugh and Way, 2018, p. 153). It is simply this movement from premise attitudes to conclusion attitudes that I will be talking about when I talk about reasoning. According to Ralph Wedgwood:

> "The meaning of a term is given by the *basic rules of rationality* governing its use. That is, the term's meaning is given by certain rules that specify ways in which it would be rational or irrational to use the term in question" (Wedgwood, 2001, p. 6, emphasis in original).

[37] See McHugh and Way (2018) for a discussion of the reasoning process as I am conceiving it. They offer an account of 'good reasoning' which consists in the preservation of fittingness; reasoning is good when it moves from fitting premises to fitting conclusions. Although it is beyond the scope of this chapter to weigh in on the debate between different theories of good reasoning, when I talk about the reason-guiding roles of normative categories I will assume that these are features of good, successful reasoning, however good reasoning itself is to be defined.

Although Wedgwood is talking about linguistic terms, I am going to suggest that we can understand the *nature* of normative categories in this way. That is, part of what it is to be a normative category is to be subject to certain rules of reasoning, and that the best way to explain the similarities and differences between different normative categories is to articulate the reason-guiding roles they play. This gives us the following view:

> **The Reason-Guiding Roles View**: the logical similarities and differences between any two deontic and evaluative categories x and y can be explained in terms of the roles they play in guiding reasoning.

To substantiate (1) above, what are some features of the generic reason-guiding role that both deontic and evaluative categories play? First of all, reasoning about values and oughts is governed by particular rules of inference. Opposition is a good example of this. In section 3 above, we noticed that for particular pairs of deontic and evaluative categories, it is not the case that they are contradictories. This is true of good and bad as well as required and forbidden. What this means is that the following inferences are not licensed: from 'x is not bad' to 'x is good', or from 'ϕ-ing is non-required' to 'ϕ-ing is forbidden'. This rule of inference is a feature of the role that these categories play in practical reasoning.

To see how this works in practice, suppose I want to determine whether or not it is morally forbidden to eat meat. It is clearly not the case that if I believe that eating meat is not morally required, then I can infer that it is forbidden, because being non-required is compatible with it being merely permitted. Or suppose I want to determine whether or not Kendrick Lamar's latest record is any good. It is clearly not the case that if I believe it is not bad, then I can infer that it is good, because being not bad is compatible with it being just average. The logical feature of non-contradictory opposition shared by certain evaluative and deontic categories is a rule that governs reasoning from beliefs with that particular deontic or evaluative content to further beliefs and intentions. This is part of the reason-guiding role such categories play. The fact that both good and bad, and required and forbidden, are non-contradictory opposites can be understood in terms of the fact that both pairs guide reasoning in a way that blocks inferences from 'x is ¬F' to 'x is G', where F and G are contrary normative categories

Another similarity between the two families is Alternatives Dependence, discussed in section 7.5, which can also be explained in terms of a rule which guides reasoning with specific categories. Recall the alternatives dependence of ought: assuming I ought to pull the lever in a standard trolley problem, killing one to save five, if we introduce a further option, a button which will halt the trolley, saving all six, then it is no longer the case that I ought to pull the lever. This is what it means for ought to be alternatives dependent. Pace Berker, we argued above that the very same is true of evaluative categories, most obviously good as applied to actions: pulling the lever can either be a good thing or a bad thing to do, depending on the alternative courses of action available to the agent. Plausibly, alternatives dependence is a procedural rule governing deliberation about what ought to be done and what is good to do. As we have seen, what ought to be done and what is good to do are both sensitive to the relevant alternative courses of action. If there is a button which will stop the trolley in our runaway trolley case, and this information is accessible to me, then my reasoning from the belief that pulling the lever will save five while killing one to the conclusion that I ought to pull the lever is faulty insofar as I have broken the alternatives dependence rule which applies to reasoning about oughts: I have failed to take an alternative course of action into account, namely the option to push a button which will save all 6, a fact that leads to the conclusion that I *ought not* pull the lever (assuming as we did in section 7.5 that some form of consequentialism is true).

This suggests that it is a feature of successful reasoning to ought conclusions that all relevant alternatives be taken into account. Reasoning with oughts is guided by alternatives dependence. Exactly the same is true about reasoning to conclusions about what is good to do; whether pulling the lever is good or bad also depends on the available options. Framing alternatives dependence as a rule which governs reasoning about what ought to be done and what is good to do sheds light on this similarity between the two categories: the fact that both ought and good are alternatives dependent can be understood in terms of the fact that both categories guide reasoning in the manner just described.

At this point, the reader might be wondering, 'but what about the difference that survived our discussion of Alternatives Dependence?' After all, section 7.5 did acknowledge that while all paradigmatic deontic categories are alternatives dependent, in certain contexts, good and bad clearly are not. While good and bad as applied to actions look straightforwardly alternatives dependent, this is not the case for good and bad as applied to things. This is evident from the conclusions drawn above about the quality of dishes at a buffet: the fact that some dish x is good is unaffected by the alternative dishes available. However, whether x is the best dish, another evaluative category, does seem to be alternatives dependent. X may be the best dish on the buffet table at time t-1, but if a caterer then introduces a better dish, x is no longer best at t-2. What we have is a situation where some evaluative categories are alternatives dependent (good as applied to actions and comparatives and superlatives) while others are not (good as applied to things considered in themselves).

By paying attention to the different kinds of reasoning that these categories participate in, we can easily make sense of this. Ought, good as applied to actions, comparatives, and superlatives all have something in common that good as applied to things lacks: they are typically used in reasoning which terminates in a *decision*. This sort of reasoning always involves a choice between competing alternatives. Good as applied to things on the other hand is not primarily used in this way. To continue with our buffet example, reasoning with oughts, good as applied to actions, and comparatives and superlatives all tend toward conclusions about *what to eat*. Indeed, it is plausible that if x is better than y, then, other things equal, you ought to choose x over y (Wedgwood, 2001, p. 24), and if x is the best dish on the table then you ought to choose it, again other things equal. From this perspective, it looks as though the categories better than and the best have as much in common with ought as they do with good, their supposed kin within the evaluative family, and this is because of the kind of reasoning problem they address: they play a role in deliberations about what to do. Each of these categories shares the feature of alternatives dependence, and this can be understood in terms of the fact that they share this deliberative role.

Good as applied to things on the other hand does not typically play this role. We often say that the risotto is good, and that the fish is also good, without being engaged in any kind of deliberation about what to eat. It is for this reason that good as applied to things is not alternatives dependent. This chimes with a difference between the deontic and the evaluative mentioned in the introduction; Tappolet's 'domains of application'. Particular normative categories are typically used in particular situations for particular purposes.

However, these purposes cut across the deontic-evaluative divide, such that the deontic 'ought' and the evaluative 'better than' and 'the best' are typically used for deciding between different possible options, while good as applied to things is not generally used for this purpose. Again this distinction between particular deontic and evaluative categories can be explained in terms of the particular functions or roles they play in reasoning; determining quality vs. determining what to do. Articulating the different roles that different normative categories play gives us the tools to account for the similarities and differences between them.

The view this section has been describing says that the various deontic and evaluative categories can be grouped together on the basis of the roles they play in reasoning, where some roles cut across the two traditional groupings. It is not the case that we can place all deontic categories within one family, and all evaluative categories within another, purely on the basis of logical markers that distinguish the two families. What reasons might there be to reject this picture? Here is one: the view I have proposed collapses the familiar and highly intuitive distinction between the deontic and the evaluative. If the reason-guiding roles view entails that deontic and evaluative categories are just the same sort of normative category, then this would be problematic insofar as we likely have lots of other reasons to retain the deontic/evaluative distinction. There is no problem here however, because the reason-guiding role view does not entail that the two traditional families collapse into one another. The aim of this chapter has *not* been to argue for the quite radical conclusion that we should discard the notion of distinct deontic and evaluative families all together. What I have tried to show is that Selim Berker's proposal that we can draw the distinction on the basis of logical features that characterise one family but not the other is flawed, and that these logical features in fact indicate interesting interrelations between the two kinds of normative categories, where this is compatible with the view that the two kinds remain distinct.

7.7. Conclusion

This chapter attempted to do two things. The first was to cast doubt on Berker's argument for the claim that the deontic and evaluative constitute two entirely separate and distinct families of normative categories. Berker's five proposed distinguishing features can in fact be found among both deontic and evaluative categories. This conclusion is significant insofar as we now know how *not* to conceive of the relationship between the deontic and the evaluative: it is too simplistic to think of them as entirely separate families distinguished by certain logical features. However, this does not entail that there are not important differences between the deontic and evaluative categories, and nothing I have said implies that either is reducible to the other.

My second aim was to suggest a better account of the relationship between deontic and evaluative categories. In light of the fact that some deontic and evaluative categories share features such as non-contradictory opposition and alternatives dependence, along with the fact that some differences persist - good simpliciter as applied to things is not alternatives dependent and categories across the two families have different domains of application - I have argued that the similarities and differences between deontic and evaluative categories are best explained in terms of the roles they play in guiding reasoning.

Chapter 8. Overall Conclusion

This thesis has been an exercise in the description and elucidation of normative categories and the relations between them. If I have managed to provide some reason for accepting two overarching claims, then the arguments presented will have succeeded in their aims. These two claims are:

1. Fittingness is fundamental; it is irreducible to other normative categories.
2. The general project of explaining one normative category in terms of another is subject to objections. We should therefore look to alternative means of explanation, such as understanding the range of normative categories in terms of the features that unite and divide them, which supports the view that they constitute an interrelated web, rather than a hierarchical structure.

Each chapter developed both of these points in tandem. Chapter 2 argued for No-X-First, in the process of which we uncovered a difficulty with the reductive X-First project and learned about the nature of different normative categories by exploring some of their phenomenological features. Chapter 3 taught us a good deal about the nature of fittingness by laying out a series of conditions on fit-evaluability and arguing that both attitudes and actions meet these conditions, while bodily episodes do not. Here too the association of fit with rationality was noted.

Chapter 4 argued that fittingness cannot be explained in terms of normative reasons. This argument turned on a feature of reasons not shared by fit-making considerations, namely a tripartite structure of normatively significant verdicts. In articulating the structure of normative verdicts we have learned something about the nature of reasons and fit. Chapter 5 defended the irreducibility of fittingness by distinguishing it from the evaluative. This is based on a feature which ought and fit plausibly share, but which the good lacks. The former two categories depend in part on an agent's epistemic position, while the latter does not.

This entails that fittingness may be more subjective than is often assumed, a view further advanced in chapter 6, where we argued that concepts representing the thick-fittingness categories in biconditionals such as 'x is disgusting if and only if it is fitting to be disgusted by x' are essentially contestable. Thus, our understanding of the nature of aptic categories becomes richer, without reducing fit to some other normative category. Finally, chapter 7 argued that placing deontic and evaluative categories in two completely separate families based on logical and relational features that supposedly distinguish them is a mistake. This is because categories across the two proposed families share many of these features. However, this does not force us to reduce one to the other, or collapse them into a single family. Instead, we can understand the similarities they share, and the differences that remain, in terms of the roles that particular categories play in reasoning.

It might seem as though my goals have been largely negative, or even destructive: I have tried to show that there is *no* X First, and that fittingness is *not* reducible to any other normative category. I have also criticised some of the ways that normative categories might be distinguished from one another. However, the thesis can also be read as a positive exposition of pluralism about normative categories and an advancement of our knowledge of what fittingness is. To say that there is no X First is to say that there are at least two distinct kinds of normative categories. Traditionally, these are the deontic and the evaluative, and in defending the view that fittingness is independent of reasons and values, I side with Selim Berker in thinking that the deontic, the evaluative and the fitting - or aptic - constitute three distinct kinds of normative categories. However, this does not mean that normative categories are wholly unrelated. On the contrary, I think the exact opposite is true: particular normative categories can remain distinct while being deeply interrelated in virtue of their shared features.

References

Achs, Rachel and Na'aman, Oded (2023). "The subtleties of fit: reassessing the fit-value biconditionals", *Philosophical Studies*, Vol. 180, pp.2523–2546.

Baker, Alan (2022). "Simplicity", in *Stanford Encyclopedia of Philosophy*, (ed.) Edward Zalta.

Berker, Selim (2017). "The Unity of Grounding", *Mind,* Vol. 127, pp. 729–777.

—— (2022). "The Deontic, the Evaluative, and the Fitting", in Christopher Howard and Rach Cosker-Rowland (eds.), *Fittingness: Essays in the Philosophy of Normativity*, Oxford: Oxford University Press, pp. 23–57.

—— (2024). "Is There Anti-Fittingness?", *Ergo: An Open Access Journal of Philosophy*, Vol. 11, No. 39, pp. 1051-1082.

Bliss, Ricki (2023). "Metaphysical Overdetermination", *The Philosophical Quarterly,* Vol. 73, No. 1, pp. 1-23.

Bramble, Ben (2016). "The Experience Machine", *Philosophy Compass*, Vol, 11, No. 3, pp. 136–145.

Broad, C.D. (1930). *Five Types of Ethical Theory*, Routledge.

Broome, John (2007). "Does Rationality Consist of Responding Correctly to Reasons?", *Journal of Moral Philosophy,* Vol. 4, No. 3, pp. 349-374.

—— (2013). *Rationality Through Reasoning*, Wiley-Blackwell.

—— (2015). "Reasons versus Ought", *Philosophical Issues*, Vol. 12, pp.80-97.

Bykvist, Krister (2009). "No Good Fit: Why the Fitting Attitude Analysis of Value Fails", *Mind*, Vol. 118, No. 469, pp. 1-30.

Collier et al (2006). "Essentially Contested Concepts: Debates and Applications", *Journal of Political Ideologies,* vol. 11, no. 3, pp. 211–246.

Cosker-Rowland, Rach (2017). "Reasons or Fittingness First?", *Ethics*, Vol. 128, pp.212–29.

—— (2019). *The Normative and the Evaluative: The Buck-Passing Account of Value*, Oxford University Press.

—— (2022a). "Value-First Accounts of Normativity", in *Fittingness: Essays in the Philosophy of Normativity*, (eds.) Christopher Howard and Rach Cosker-Rowland, Oxford University Press.

—— (2022b). "The Authoritative Normativity of Fitting Attitudes", *Oxford Studies in Metaethics 17,* pp. 108–137.

Cullity, Garrett (2022). "Reasons and Fit", in Chrispher Howard and Rach Cosker-Rowland (eds.), *Fittingness: Essays in the Philosophy of Normativity*, Oxford: Oxford University Press, pp. 23–57.

Dancy, Jonathan (2000). *Practical Reality,* Oxford: Oxford University Press.

—— (2004). *Ethics Without Principles,* Oxford University Press.

D'Arms, Justin and Jacobson, Daniel (2000a). "Sentiment and Value", *Ethics*, Vol. 110, No. 4, pages 722-748.

—— (2000b). "The Moralistic Fallacy: On the 'Appropriateness' of Emotions", *Philosophy and Phenomenological Research*, Vol. 61, No. 1, pp. 65-90.

—— (2023). *Rational Sentimentalism*, Oxford University Press.

Dasgupta, Shamik (2017). "Constitutive Explanation", *Philosophical Issues*, Vol. 27, No.1, pp. 74-97.

de Gaynesford, Maximilian (2004). *John McDowell,* Polity Press.

Di Marco, Marina and Khalifa, Kareem (2022). "Sins of inquiry: How to criticize scientific pursuits", *Studies in the History and Philosophy of Science*, Vol. 92, pp. 86-96.

Dokic, Jérôme and Lemaire, Stéphane (2013). "Are Emotions Perceptions of Value", *Canadian Journal of Philosophy*, Vol. 43, No. 2, pages. 227-247.

Döring, Sabine (2007). "Seeing What to Do: Affective Perception and Rational Motivation." *dialectica*, Vol. 61, No. 3, pp. 363–394.

Ebels-Duggan, Kyla (2023). "Buck-Passing and the Value of a Person", *Oxford Studies in Normative Ethics 13*, pp. 77–98.

Enoch, David (2006). "Agency, Shmagency: Why Normativity Won't Come from What Is Constitutive of Action", *The Philosophical Review*, Vol. 115, No. 2, pp. 169-198.

Ewing, A.C. (1939). "A Suggested Non-Naturalistic Analysis of Good", *Mind*, Vol. 48, No. 189, pp. 1-22.

—— (1948). *The Definition of Good*, Routledge & Kegan Paul Ltd.

Ezcurdia, Maite (1998). “The Concept-Conception Distinction”, *Philosophical Issues,* Vol. 9, pp. 187-192.

Faraci, David (2020). “We Have No Reason to Think There Are No Reasons for Affective Attitudes”, *Mind*, Vol. 129, pp. 225–234.

Fassio, Davide (2022). “What the Doctor Should Do: Perspectivist Duties for Objectivists about Ought”, *Philosophical Studies*, Vol. 179, pp. 1523–1544.

Fleisher, Will (2022). “Pursuit and inquisitive reasons”, *Studies in the History and Philosophy of Science,* vol. 94, pp. 17-30.

Foot, Philippa (1985). “Utilitarianism and the Virtues”, *Mind*, Vol. 94, No. 374, pp. 196-209.

Fritz, James (2023). “Why Fittingness Is Only Sometimes Demand-Like”, *Philosophical Studies*, Vol. 180, pp. 2597–2616.

Gallie, W.B. (1955-1956). “Essentially Contested Concepts”, *Proceedings of the Aristotelian Society,* vol. 56, pp. 167-198.

—— (1956). “Art as an Essentially Contested Concept”, *The Philosophical Quarterly*, Vol. 6, No. 23, pp. 97-114.

Gaut, Berys (2007). *Art, Emotion and Ethics,* Oxford: Oxford University Press.

Gertken, Jan and Kiesewetter, Benjamin (2017). “The Right and the Wrong Kind of Reasons”, *Philosophy Compass*.

Gibbons, John (2013). *The Norm of Belief,* Oxford University Press.

Goldie, Peter (2004). “Emotion, Feeling, and Knowledge of the World” in *Thinking about Feeling: Contemporary Philosophers on Emotions,* ed. Robert Solomon, Oxford University Press.

Graham, Peter A. (2010). “In Defense of Objectivism about Moral Obligation”, *Ethics,* Vol. 121, No. 1, pp. 88-115.

Hansen, Nat (2021). “Metalinguistic Proposals”, *Inquiry,* vol. 64, pp. 1-19.

Heape, Alexander (2020). “How There Could Be Reasons for Affective Attitudes”, *Ethical Theory and Moral Practice,* Vol. 23, pp. 667–680.

Hieronymi, Pamela (2006). "Controlling Attitudes", *Pacific Philosophical Quarterly*, Vol. 87, pp. 45–74.

Hooker, Brad (2024). "Fittingness and Well-Being", *The Journal of Ethics,* Vol. 29, pp. 175–193.

Horgan, Terry and Timmons, Mark (2005). "Moral Phenomenology and Moral Theory", *Philosophical Issues*, Vol. 15, pp. 56-77.

Howard, Christopher (2016). "In Defense of the Wrong Kind of Reason", *Thought: A Journal of Philosophy,* Vol. 5, No. 1, pp. 53-62.

—— (2018). "Fittingness", *Philosophy Compass,* Vol 13, No. 11, pp. 1-14.

—— (2019). "The Fundamentality of Fit", *Oxford Studies in Metaethics 14,* pp. 216–36.

—— (2023a). "Forever Fitting Feelings", *Philosophy and Phenomenological Research,* Vol 107, pp. 80–98.

—— (2023b). "Fitting-Attitude Theories of Value" in Edward Zalta (ed.), *Stanford Encyclopedia of Philosophy.*

Howard, Christopher and Leary, Stephanie (2022). "In Defence of the Right Kind of Reason", in Chrisopher Howard and Rach Cosker-Rowland (eds.), *Fittingness: Essays in the Philosophy of Normativity*, Oxford: Oxford University Press, pp. 221-241.

Hurka, Thomas (2014). *British Ethical Theorists from Sidgwick to Ewing*, Oxford: Oxford University Press.

—— (2019). "More Seriously Wrong, More Importantly Right", *Journal of the American Philosophical Association,* Vol. 5, No. 1, pp. 41-58.

—— (2022). "Against the Fundamentality of Fit", in Christopher Howard and Rach-Cosker Rowland (eds.), *Fittingness: Essays in the Philosophy of Normativity*, Oxford: Oxford University Press, pp. 58-79.

Jackson, Frank (1991). "Decision-theoretic Consequentialism and the Nearest and Dearest Objection", *Ethics*, Vol. 101, No. 3, pp. 461-482.

Kagan, Shelly (1991). *The Limits of Morality,* Oxford: Oxford University Press.

Kant, Immanuel (2018). *Groundwork for the Metaphysics of Morals,* ed. and trans. Allen W. Wood, Yale University Press.

Katsafanas, Paul (2018). "Constitutivism about Practical Reasons", in *The Oxford Handbook of Reasons and Normativity,* (ed.) Daniel Starr, pp. 367-391.

Kenny, Anthony (1963). *Action, Emotion and Will*, Routledge and Kegan Paul.

Kiesewetter, Benjamin (2011). "'Ought' and the Perspective of the Agent", *Journal of Ethics and Social Philosophy,* Vol. 5, No. 3.

—— (2017). *The Normativity of Rationality,* Oxford: Oxford University Press.

Kirchin, Simon (2025). "Introduction: Normativity's Present and Future", in *The Future of Normativity,* ed. Simon Kirchin, Oxford University Press, pp. 1-34.

Korsgaard, Christine (2009). *Self-Constitution: Agency, Identity, and Integrity*, Oxford University Press.

Kriegal, Uriah (2023). "A Fitting-Attitude Approach to Aesthetic Value?", *British Journal of Aesthetics,* Vol. 63, No. 1, pp. 57-73.

Lehrer, Keith (1990). *Theory of Knowledge,* Routledge, London.

Lord, Errol (2015). "Acting for the Right Reasons, Abilities, and Obligation", *Oxford Studies in Metaethics 10,* pp. 26–52.

—— (2018). *The Importance of Being Rational,* Oxford: Oxford University Press.

—— (2023). "Everything First", *Aristotelian Society Supplementary Volume XCVII.*

Lord, Errol and Maguire, Barry (2016). "An Opinionated Guide to the Weight of Reasons", in*,* Errol Lord and Barry Maguire (eds.), *Weighing Reasons,* Oxford University Press.

MacIntyre, Alasdair (1973). "The Essential Contestability of Some Social Concepts", *Ethics*, Vol. 84, No. 1, pp. 1-9.

Mackie, J.L. (1977). *Ethics: Inventing Right and Wrong*, Penguin.

Maguire, Barry (2016). "The Value-Based Theory of Reasons", *Ergo,* Vol. 3, No. 09.

—— (2018). "There Are No Reasons for Affective Attitudes", *Mind*, Vol. 127, pp. 779-805.

Mason, Elinor (2013). "Objectivism and Prospectivism about Rightness", *Journal of Ethics and Social Philosophy,* Vol. 7, No. 2, pp. 1-22.

McConnell, Terence (1988). "Ross on Duty and Ignorance", *History of Philosophy Quarterly* , Vol. 5, No. 1, pp. 79- 95.

McDowell, John (1998). "Value and Secondary Qualities" in *Mind, Value, and Reality,* Harvard University Press.

McHugh, Conor (2023). “Attitudes and the Normativity of Fittingness”, *Aristotelian Society Supplementary Volume XCVII.*

McHugh, Conor and Way, Jonathan (2016). “Fittingness First”, *Ethics*, Vol. 126, pp. 575–606.

—– (2018). ‘What Is Good Reasoning’, *Philosophy and Phenomenological Research,* Vol. 96, No. 1, pp. 153-174.

—– (2022a). *Getting Things Right: Fittingness, Reasons and Value,* Oxford: Oxford University Press.

—– (2022b). “All Reasons Are Fundamentally for Attitudes”, *Journal of Ethics and Social Philosophy*, Vol. 21, No. 2, pp. 151-174.

McElwee, Brian (2017). “Supererogation Across Normative Domains”, *Australasian Journal of Philosophy*, Vol. 95, pp. 505-516.

Mandelbaum, Maurice (1955). *The Phenomenology of Moral Experience,* The Free Press.

Milona, Michael (2016). “Taking the Perceptual Analogy Seriously”, *Ethical Theory and Moral Practice,* Vol. 19, No. 4, pp. 897–915.

Moore, G.E. (1993). *Principia Ethica: Revised Edition*, (ed.) Thomas Baldwin, Cambridge University Press.

Naar, Hichem (2022). “Skepticism about Reasons for Emotions”, *Philosophical Explorations,* Vol. 25, pp. 108-123.

Na’aman, Oded (2021). “The Rationality of Emotional Change: Toward a Process View”, *Nous*, Vol. 55, pp. 245–269.

—– (2022). “What Is Evaluable for Fit?” in Christopher Howard and Rach-Cosker Rowland (eds.), *Fittingness: Essays in the Philosophy of Normativity*, Oxford: Oxford University.

Nozick, Robert (1974). *Anarchy, State, and Utopia,* Basic Books.

O’Donnell, Guillermo and Schmitter, Philippe (1986). *Transitions from Authoritarian Rule: Tentative Conclusions about Uncertain Democracies,* Baltimore: Johns Hopkins University Press.

Olsen, Kristian (2017). “A Defense of the Subjective/Objective Moral Ought Distinction”, *The Journal of Ethics,* Vol. 21, No. 4, pp. 351-373.

Orsi, Francesco (2015). “The Guise of the Good”, *Philosophy Compass*, Vol. 10, No. 10, pp. 714–724.

Orsi, Francesco and Garcia, Andrés G. (2021). “The Explanatory Objection to the Fitting Attitude Analysis of Value”, *Philosophical Studies,* Vol 178, pp. 1207–1221.

Parfit, Derek (2011). *On What Matters*, Oxford: Oxford University Press.

Plunkett, David and Sundell, Tim (2013). “Disagreement and the Semantics of Normative and Evaluative Terms”, *Philosophers’ Imprint*, vol. 13, no. 23, pp. 1-37.

Prichard, H.A. (2002). “Duty and Ignorance of Fact”, in *Moral Writings,* (ed.) Jim MacAdam, Oxford University Press.

Rabinowicz, Wlodek and Rønnow-Rasmussen, Toni (2000). “The Strike of the Demon: On Fitting Attitudes and Value, *Ethics,* Vol. 114, No.3, pp. 391–423.

Raz, Joseph (2002). *Engaging Reason: On the Theory of Value and Action,* Oxford University Press.

Ross, W. D. (1939). *The Foundations of Ethics,* Oxford University Press.

Scanlon, T. M. (1998). *What We Owe to Each Other*, Cambridge: Harvard University Press.

Scarantino, Andrea and Ronald de Sousa (Summer 2021 Edition). "Emotion", *The Stanford Encyclopedia of Philosophy* (ed) Edward N. Zalta, URL: <https://plato.stanford.edu/archives/sum2021/entries/emotion/>.

Schmidt, Thomas (2024). “The Balancing View of Ought”, *Ethics*, Vol. 134, No. 2, pp. 246-267.

Schroeder, Mark (2007). *Slaves of the Passions,* Oxford: Oxford University Press.

––– (2010). “Value and the Right Kind of Reason”, *Oxford Studies in Metaethics 5,* pp. 25-55.

––– (2021a). “The Fundamental Reason for Reason Fundamentalism”, *Philosophical Studies*, Vol. 178: pp. 3107–3127.

––– (2021b). *Reasons First,* Oxford University Press.

Schultz, Conner (2025). “Deliberative Control and Eliminativism about Reasons for Emotions”, *Australasian Journal of Philosophy*, Vol. 103, pp. 72-87.

Sepielli, Andrew (2018). “How Moral Uncertaintism Can Be Both True and Interesting”, *Oxford Studies in Normative Ethics 7,* pp. 98–116.

Shafer-Landau, Russ (2003). *Moral Realism: A Defense*, Oxford University Press.

Sharadin, Nathaniel (2016). “Reasons Right and Wrong”, *Pacific Philosophical Quarterly*, Vol. 97, No. 3, pp. 371-399.

Skorupski, John (2010). *The Domain of Reasons*, Oxford: Oxford University Press.

Smith, Holly (2010). “Subjective Rightness”, *Social Philosophy and Policy,* Vol. 27, No. 2, pp. 64-110.

Snedegar, Justin (2016). “Reasons, Oughts, and Requirements”, *Oxford Studies in Metaethics 11,* pp.155-181.

Solomon, R. C. (1973). “Emotions and Choice”, *Review of Metaphysics*, Vol. 27, No. 1, pp. 20–41.

Stratton-Lake, Philip (2018). ‘Reasons Fundamentalism and Value’ in *The Oxford Handbook of Reasons and Normativity,* ed. Daniel Starr, pp. 275–296.

––– (2022). “Intuitions of Fittingness”, in Christipher Howard and Rach-Cosker Rowland (eds.), *Fittingness: Essays in the Philosophy of Normativity*, Oxford: Oxford University, pp. 130–148.

––– (2025). “Ross and the Problem of Permissibility” in *The Moral Philosophy of W. D. Ross: Metaethics, Normative Ethics, Virtue, and Value*, (eds.) Robert Audi and David Phillips, Oxford: Oxford University Press.

Svavarsdóttir, Sigrun (2014). Having Value and Being Worth Valuing, *Journal of Philosophy*, No. 3, pp. 84–109.

Swanton, Christine (1985). “On the ‘Essential Contestedness’ of Political Concepts”, *Ethics,* Vol. 95, No. 4, pp. 811-827.

Tappolet, Christine (2011). “Values and Emotions: Neo-Sentimentalism’s Prospects”, in *Morality and the Emotions,* (ed.) Carla Bagnoli, Oxford University Press.

––– (2013). ‘Evaluative vs. Deontic Concepts’ in *The International Encyclopedia of Ethics*, ed. Hugh LaFollette, print pages 1791–1799.

––– (2016). *Emotions, Values, and Agency*, Oxford: Oxford University.

Teroni, Fabrice (2007). “Emotions and Formal Objects”, *Dialectica*, Vol. 61, No. 3, pp. 395-415.

Väyrynen, Pekka (2014). “Essential Contestability and Evaluation”, *Australasian Journal of Philosophy,* vol. 93, no. 3, pp. 471-488.

Velleman, J David (2000). *The Possibility of Practical Reason,* Oxford University Press.

—— (2004). “Précis of ‘The Possibility of Practical Reason’, *Philosophical Studies*, Vol. 121, No. 3, pp. 225-238.

Waldron, Jeremy (2002). “Is the Rule of Law an Essentially Contested Concept (In Florida)?”, *Law and Philosophy*, Vol. 21, No. 2, pp. 137-164.

Ward, Shane (2024). “We Have Reason To Think There Are Reasons For Affective Attitudes”, *Inquiry,* Vol. 67, pp. 3969–3987.

Way, Jonathan and Whiting, Daniel (2016). “Perspectivism and the Argument from Guidance”, *Ethical Theory and Moral Practice,* Vol. 19, No. 5, pp. 361–374.

Wedgwood, Ralph (2001). ‘Conceptual Role Semantics For Moral Terms’, *The Philosophical Review*, Vol. 110, No. 1, pp .1-30.

—— (2018). ‘The Unity of Normativity’ in *The Oxford Handbook of Reasons and Normativity,* (ed.) Daniel Starr, pp. 23–45.

Wodak, Daniel (2020). “Who’s on First?”, *Oxford Studies in Metaethics 15*, pp. 49–71.

Wu, Yan et al. (2019). “How Do Amusement, Anger and Fear Influence Heart Rate and Heart Rate Variability?”, *Frontiers in Neuroscience,* vol 13, pp. 1-6.

Zimmerman, Michael (2008). *Living with Uncertainty: The Moral Significance of Ignorance,* Cambridge University Press.

—— (2014). *Ignorance and Moral Obligation,* Oxford University Press.