



MORAL RESPONSIBILITY AND MERIT

BY MATT KING

JOURNAL OF ETHICS & SOCIAL PHILOSOPHY

VOL. 6, No. 2 | AUGUST 2012

URL: WWW.JESP.ORG

COPYRIGHT © MATT KING 2012

Moral Responsibility and Merit

Matt King

IN THE CONTEMPORARY MORAL RESPONSIBILITY DEBATE, most theorists seem to be giving accounts of responsibility in the “desert-entailing sense.” This is meant to distinguish it from causal or legal responsibility and to draw it closer to our other moral concepts. Moral responsibility and desert are natural partners: Morally responsible agents can be blameworthy and praiseworthy – they can deserve blame and praise. This convergence on responsibility in the desert-entailing sense is a welcome development, for it helps secure competing accounts as rival accounts, a status that appears to require having the same target notion in mind. Yet, despite the convergence, it is striking that so little has been said about the notion of desert that is supposedly entailed. One potential worry is that without saying more about desert, we risk merely replacing one difficult concept (moral responsibility) with one just as difficult (desert).

This paper seeks to address this lacuna in the moral responsibility literature. I propose an understanding of desert sufficient to help explain why the blameworthy and praiseworthy deserve blame and praise, respectively. I do so by drawing upon what might seem an unusual resource. I appeal to so-called Fitting-Attitude accounts of value to help inform a conception of desert or merit, one that can be usefully applied to discussions of moral responsibility. I’m less concerned with defending the view than with explicating it as a conjecture and examining what work it might be able to do. As such, this paper is both speculative and overtly noncommittal. I do not seek to argue for a conception of desert so much as to investigate a potential line of thinking about it. I do, however, argue that the candidate view, which I will call Desert as Fittingness (or DAF),¹ merits additional attention. I do so by defending two claims: First, that it does better than extant Fitting Attitude accounts of blameworthiness and praiseworthiness, and, second, that it has an initial plausibility with respect to informing a *general* account of desert. Again, these reasons are not intended to support the view as true, but merely to make the case for taking the view seriously.

1. The Basic View

The outline of DAF connects four thoughts.

First, theories of moral responsibility, if they are to account for responsibility “in the desert-entailing sense,” need some positive account of the no-

¹ See Zaibert (2006) for a separate (and mostly unrelated) discussion of fittingness and desert. First, his attention is focused on theories of punishment, and he employs a different notion of “fittingness” concerned only with aesthetic normativity. Second, he wants to distinguish desert and fittingness, in sharp contrast to my aim here. And third, he conceives of fittingness as a substantive normative relation, rather than as a generic placeholder, unlike the Fitting Attitude accounts, which I discuss below.

tion of desert being used. We need to know what is entailed by the responsibility relation in order to properly evaluate those theories. Given that the very plausibility of treating moral responsibility as responsibility in the desert-entailing sense rests on there being a natural connection between moral responsibility and desert, it seems at the very least prudent to seek a better understanding of the latter.

The second thought concerns two observations. One, that moral responsibility involves blameworthiness and praiseworthiness. When we talk about moral responsibility we are talking about a way of being related to things so as to make individuals blameworthy or praiseworthy for them.² The other observation concerns Fitting Attitude (FA) accounts of value. These accounts provide general explanations of (at least certain) normative properties in terms of the attitudes it is “fitting” to adopt in response to those properties.³ Part of the motivation for such accounts comes from the fact that there seems to be a semantic connection between the properties and relevant attitudes. So, for example, the “desirable” is not what we are able to desire, but what is fitting of desire. This observation appears to extend to a wide array of similar terms. Thus, the admirable is what is fitting of admiration, the enviable is what is fitting of envy and the contemptible is what is fitting of contempt. Similar but differently constructed terms also seem to call for FA treatments. So, the fearsome is fitting of fear and the awesome is fitting of awe, but also the amusing is what is fitting to find amusing. The terms are wide ranging but all naturally call for an attitude that seems fitting or appropriate to hold toward the object that possesses that property.

What is especially relevant to DAF, however, is that the blameworthy and praiseworthy are also attractive candidates for FA treatments. Just as “-able” terms seem to appeal naturally to specific attitudes, so do “-worthy” terms. Thus, the blameworthy is what is fitting to blame and the praiseworthy is what is fitting to praise. If FA accounts are a promising general strategy for explaining this wide range of normative properties – including blameworthiness and praiseworthiness – and if these terms are intimately tied to moral responsibility, then it is plausible to suppose an account of moral responsibility can meaningfully incorporate FA strategies in some way.

The third thought concerns the structure of FA accounts, which are unified by a general schema: An object possesses a normative property, “phi-ability,” just in case it is *fitting* to phi that object. But that fittingness relation is meant to be generic, and different FA accounts will differ in precisely how they characterize what fittingness amounts to. So, one version might take fitting attitudes to be those it is correct to hold, or that one ought to hold, or that are “fill in the blank” to hold. All such views will be FA accounts so long

² At least in principle. One could be morally responsible for an act and yet fail to be either blameworthy or praiseworthy for it. Presumably this occurs when one is responsible for a morally neutral act.

³ For some general discussion, see D’Arms & Jacobsen (2000); Hieronymi (2005); Parfit (2001); Rabinowicz & Rønnow-Rasmussen (2004); Schroeder (2010); Way (2012).

as the fill-in-the-blank relation is a normative relation that attitudes can bear to objects.⁴ Thus, FA accounts rely on there being some normative relation holding between the properties being explained and the relevant attitudes. We can individuate FA accounts in part, therefore, by the normative relation that takes the place of “fittingness” in the general schema.

The fourth and final thought is that the desert relation is a normative relation. Reasons bear on what things deserve what and that which is deserved by whom under the circumstances. It follows that the desert relation is the right *sort* of relation to assume the role held by the generic “fitting” relation.

I hope it is fairly apparent how we get DAF out of these four thoughts. The basic idea behind DAF is that the desert relation *is* the normative relation operative in FA accounts. The admirable deserve admiration; the contemptible deserve contempt. Thus, we get a positive view of desert for use in theories of moral responsibility: Blameworthy (or praiseworthy) agents deserve or merit blame (or praise) *in the same way* that the admirable deserve admiration or the fearsome merit fear. This view connects moral responsibility with a notion of desert, but remains neutral on the conditions on moral responsibility and the conditions on what things deserve what. Moral responsibility remains the relation in virtue of which agents deserve blame or praise, and thus the concept remains significantly unrevised.

I think there is something intuitively compelling about DAF. There seems to be something right in saying that the admirable merit admiration or that the fearsome merit fear. The blameworthy certainly merit blame and the praiseworthy certainly merit praise. So, as a matter of mere usage the view initially fits. However, I do not want to make too much out of this initial intuitiveness (especially because some readers may find it less compelling). Rather, I want to make clear that the view is characterized by employing what may amount to a stipulative notion of desert or merit, one that need not be independently motivated for all possible FA analyses.⁵

As I noted at the outset, I am more interested in examining the prospects of DAF than arguing for its adoption outright. So, I will proceed as an investigator, seeking to determine whether this view warrants special attention. Obviously, I believe that it does. And while I will not claim that this is the true view about desert and responsibility, I will argue that DAF has some attractions in its favor. Chiefly, I argue that it does better than extant FA accounts of blameworthiness (and praiseworthiness) and that it is consistent with a plausible general picture of desert. While I do not suppose these rea-

⁴ See Schroeder (2010) for commentary on general structure. Compare Enoch (2004) on the normative nature of the relation.

⁵ Pojman (1997) holds that desert is a species of merit, wherein merit grounds appropriate treatment of the subject in virtue of some property, whether the subject is responsible for that property or not, while desert requires responsibility for the property in question. This seems unwarranted, both because we ordinarily use “desert” and “merit” interchangeably, and because we ordinarily make desert claims regarding properties we are not responsible for (e.g., “all persons deserve respect in virtue of their personhood”).

sons to settle the case for DAF, I do think they show it to be worth our consideration and justify DAF as providing a worthwhile *strategy* for further inquiry.

2. FA Accounts of Blameworthiness

DAF gives an FA analysis of blameworthiness and praiseworthiness. An agent is blameworthy just in case she deserves blame, and an agent is praiseworthy just in case she merits praise. The view treats these claims on a par with all other FA analyses. Thus, S is admirable just in case S deserves admiration, and S is fearsome just in case S merits fear. Generalized, we get the following statement:

S is phi-able just in case S deserves/merits phi.

It is a genuine FA account in that it explains the relevant properties in terms of the attitudes it is fitting to hold. In the case of DAF, that normative relation is desert/merit.

But this view is not the only game in town. There are other FA accounts of blameworthiness; it is worth looking at the competition. In this section, I will argue that DAF does better than competing views.

Alternative FA accounts of blameworthiness are best characterized by the normative notion used for the generic “fittingness.” Most prominent among these are accounts that construe the appropriateness of blame in terms of fairness.⁶ An agent is blameworthy just in case it is fair to blame him. An early proponent of this view, R. Jay Wallace, holds that since blame is connected to sanction-like behavior, considerations of fairness naturally arise. For Wallace, explaining blameworthiness in terms of fairness also allows us to connect the attitudes and practices of holding each other responsible more closely to the notion of moral obligation, providing a mutually supporting and illuminating account of all these terms.

Likewise, a more recent proposal by Darwall construes the appropriateness of blame in terms of a certain interpersonal authority to make moral demands of each other.⁷ An agent is blameworthy just in case one has violated a demand those in the moral community have the authority to make of him.⁸ While more intricate, perhaps, than Wallace’s, it is clear that Darwall’s view, too, conceives of fittingness in, roughly, moral terms. His account of blameworthiness figures in a more complicated story about the nature of morality itself, but retains Wallace’s focus on appropriate blame being held to moral standards. For Wallace, these standards are norms of fairness. For

⁶ See Wallace (1994).

⁷ See Darwall (2006).

⁸ First, Darwall (2006) does not put his formulation in precisely these terms, though I think he is committed to them. Second, he adds a caveat that the agent does not have a valid excuse. I leave the matter of excuses aside.

Darwall, these standards are norms that lie within the reciprocal respect and authority granted agents as equal moral persons. While the details may differ, each gives an FA account of blameworthiness deeply rooted in moral terms.

This moral concern, while perhaps understandable in explaining a moral term like blameworthiness, burdens these FA accounts with particular problems that DAF can avoid. This may strike some as surprising. It may seem natural to give the desert/merit relation involved in DAF a moral reading too. Or one might be inclined to think that all desert or merit has a distinctively moral flavor. It will become clear as we proceed that DAF does not employ a moralized notion of desert. Indeed, this is an implication of the structure of the view itself. This is a welcome result, for moralizing the “fitting” relation in FA accounts raises special problems.

First, neither Wallace’s nor Darwall’s FA account is extendable to other normative properties. While fairness (or the legitimacy of interpersonal demands) may look attractive for the appropriateness of blame, it seems particularly ill suited to capturing the appropriateness of other notions typically given FA treatments. There’s nothing unfair about failing to fear a coiled copperhead at one’s feet or about admiring a despicable desperado. Proponents of these fairness-based views may think that there is something unique about blame that requires understanding appropriateness in moral terms, but I think it should initially be puzzling why blameworthiness should be given an FA account different from a general account of normative properties. And this is exactly what DAF does not do. Not only does it give an FA account for all the relevant normative properties, it does so in terms of the *same* normative relation: desert/merit. This follows from the structure of DAF. It understands all the relevant properties in terms of the desert/merit relation, including moral properties like blameworthiness, but also less moral terms, like admirableness,⁹ and nonmoral terms, like fearsomeness. This generality seems a mark in favor of DAF’s account of blameworthiness over extant accounts.

One might oppose over-generalizing FA accounts, and think that non-moral properties ought to get separate treatment from moralized terms.¹⁰ Perhaps while the blameworthy morally deserve fear (when fair or authorized), the fearsome only prudentially merit fear – and the awesome, say, are only aesthetically worthy of awe. In other words, we might think that there is something special about blameworthiness because it is a moral notion, while other normative properties, though they may be suited to an FA analysis, nevertheless get a different FA treatment (and perhaps different nonmoral terms get different treatments themselves).

One reason to doubt this is the case, however, is that much of the attraction of FA analyses involves their generality. The large set of normative

⁹ I call admirableness “less” moral but not nonmoral, since one might be admired for one’s good deeds or character.

¹⁰ As an anonymous referee helpfully suggested.

properties bearing a fitting relation to related attitudes calls for a single explanation. If the generic fittingness relation is filled out by a number of *different* relations, we no longer have a single analysis of these terms, but rather a set of related analyses. I cannot argue here that such an approach is incorrect, but I think that if the FA program is promising, as it seems to be, it is partly because of its generality.¹¹

But even were one inclined, for whatever reason, to hold a distinction between moral and nonmoral properties for FA analyses, this distinction is harder to sustain for concepts intimately connected to blameworthiness. In particular, whatever the rationale for explaining blameworthiness in terms of the fairness of blame, it does not seem naturally extendable to praiseworthiness. If A praises B for something despicable, this does not seem to be unfair in any pressing sense. For one thing, it certainly is not unfair to B (and B would seemingly have little complaint). And while it might be unfair to others who are not getting praised for their good works, this would make praise unfair precisely to those one is not praising. This would be an odd account of unfair praising, if praise of some individual was rendered inappropriate for its effects on other, non-praised individuals. When undeserved blame is unfair, it is unfair to the one being blamed, not to other persons. Thus, it seems difficult to see how an account based on fairness could give parallel analyses for blameworthiness and praiseworthiness.

This difficulty seems no less pressing for Darwall's version of the view, which is grounded in an authority to make legitimate demands of each other. On this view, an agent is blameworthy when she has failed a legitimate moral demand we have the authority to make of her. This might suggest praiseworthiness being understood in terms of demands too. But praise does not seem to involve demands at all. While praise may be called for in cases where demands are exceeded, this is not to demand anything of the other person. So, while praise might *reference* a moral demand (or be prompted by one), it nevertheless does not involve demanding anything. Indeed, praise is often most called for when others do things that far outstrip the content of our legitimate moral demands. Finally, there may be cases in which there is no moral demand to be made, but wherein praise is all the more appropriate as a result. For instance, it may turn out that there are no positive duties to aid others, and therefore no blameworthiness for failing to aid them. If there were no demand to aid others, however, this would not show that we are not praiseworthy when we do aid them.¹² Darwall's version of the view, in which legitimate moral demands grounded in interpersonal authority form the basis of

¹¹ See Schroeder (2010) for a fuller examination of the promise of the FA account of value and its generality.

¹² And notice that it cannot be the case that praise is called for when we do things we are not demanded to do. For surely there is no legitimate demand that I watch a particular movie tonight (under ordinary circumstances). But, equally true, I surely am not praiseworthy for doing so.

moral responsibility, seems unable to construct parallel analyses of blameworthiness and praiseworthiness.¹³

Yet it seems as though blameworthiness and praiseworthiness ought to be given parallel FA analyses. Pretheoretically, blameworthiness and praiseworthiness lie on opposite ends of a spectrum, with the blameworthy being worse than the praiseworthy – the blameworthy deserve blame as the praiseworthy deserve praise. If ever there were two terms, both seemingly suited for an FA account, which also were independently conceptually connected, blameworthiness and praiseworthiness seem paradigmatic instances. Thus, a weakness of standard FA accounts of blame is that they do not extend to praiseworthiness, even if we limit our analyses to moral terms.

As a final contrast between DAF and extant FA accounts of blameworthiness, FA accounts of value terms are notoriously subject to the so-called Wrong Kind of Reasons Problem.¹⁴ Recall that FA accounts work by analyzing a normative property in terms of the attitude it is fitting to adopt toward an object with that property. With the Wrong Kind of Reasons Problem, we can seemingly generate cases in which there are reasons for the relevant attitude that do not seem to bear at all on whether the object in question has the relevant property. So, for instance, an evil demon can threaten to harm your loved ones if you do not admire the despicable, or an eccentric billionaire can offer you a million dollars to fear a box turtle. While the strength of these reasons may even make it the case that you *ought* to admire the despicable and fear the turtle, they certainly do not bear on the admirableness or fearsomeness of the objects in question. Thus, they are reasons of the wrong kind, reasons that an FA account of those properties must somehow rule out.

It is natural to suppose that, whatever the normative relation one adopts in one's FA account, that relation will be affected by reasons; wherever there is normativity, there are reasons. This is no less true for DAF. If the admirable is that which deserves admiration, then there will be reasons that bear on it deserving admiration. This means that, in principle, DAF is subject to the Wrong Kind of Reasons Problem, for it remains open, in principle, that we could generate a case that provides a consideration that bears on desert while obviously not bearing on, say, admirableness (or fearsomeness or blameworthiness, etc.).

Nevertheless, DAF is subject to this problem only insofar as *any* FA account is subject to it. And we have reason to be confident that the problem can be solved.¹⁵ In any event, when contrasting FA accounts, given that they

¹³ This should come as no surprise to readers familiar with Darwall. In his recent work (2006) he never so much as mentions praiseworthiness. Relatedly, Wallace explicitly rejects praiseworthiness as central to his target concept of moral responsibility (1994: 61).

¹⁴ See Schroeder (2010) for a good overview.

¹⁵ See Schroeder (2010: 3). He argues convincingly that, given how compelling FA accounts are for properties like “admirable” and “enviable” (etc.), we should have confidence that there must be *some* notion of fittingness that will do the requisite work while remaining immune to reasons of the wrong kind.

all must face the problem,¹⁶ the fact that each faces the problem leaves them on equal footing. However, even if a general solution were to be developed, fairness-based views like Wallace's would face a separate Wrong Kind of Reasons Problem. Relying on an independent notion like fairness can all too easily lead to Wrong Kind of Reasons cases once we introduce a consideration that clearly bears on fairness but not on blameworthiness.¹⁷ It may be unfair of me to blame Mark for stealing my car because I'm an unrepentant car thief. But this no more affects Mark's blameworthiness than the evil demon's threat affects an object's admirability. Or it may be unfair for me to blame Mary for neglecting her duty to go to the movies, for if I had been in the same circumstances, I would have done the very same thing. In both cases, the blame seems hypocritical, and this may be enough to show that such blaming would be unfair. But this does not bear on either Mark's or Mary's blameworthiness. (These cases seem to pose the same problem for Darwall's view, for I might plausibly lack the requisite authority to make the relevant demands of Mark or Mary given my hypocritical position.)

These instances of the Wrong Kind of Reasons Problem would plausibly remain even were a solution to the more general version to be found. Indeed, extant FA accounts of blameworthiness constrain the answers we can provide for the Wrong Kind of Reasons Problem, for they already specify what the right kind of reasons are supposed to be (e.g., fairness-based considerations). In contrast, DAF relies only on the same normative notion present in all FA analyses, and thus is (putatively) solved by whatever solution works generally. And since the present thesis is only that DAF is more promising than extant FA accounts, skepticism about a general solution to the problem is irrelevant. If it turns out that FA accounts must be rejected wholesale, then it follows that DAF must be false. But such a concern is orthogonal to our purposes here.

Moreover, it is worth noting that it is much harder to conceive of considerations that would bear on the desert of an attitude but would not be relevant to the associated normative property. As DAF analyzes blameworthiness, the blameworthy are those who deserve/merit blame. A Wrong Kind of Reasons case would involve a reason that makes the blame appear deserved, and yet that reason clearly does not bear on the agent's *blameworthiness*. Such a case seems extraordinarily difficult to conceive. Mark's and Mary's cases above will not do the trick. Moreover, since the matter of someone's deserving something is often taken to be grounded in facts about the person or thing in question (rather than, say, consequentialist considerations), circumstantial features like the billionaire's offer or the demon's threat similarly will not affect deservingness. Thus, in this respect, DAF may be able to better withstand Wrong Kind of Reasons objections than other competitors (though this is an admittedly speculative suggestion).

¹⁶ Or, perhaps, explain it away. See Way (2012).

¹⁷ See Smith (2007).

I take the above considerations to show DAF to be well motivated against extant FA accounts of blameworthiness, whatever the other merits of such views may be.¹⁸

3. Objection 1: The “Depth” of Desert

I think DAF for blameworthiness and praiseworthiness is attractive, especially when compared against extant accounts. But the basic equivalence at the heart of the view, that the blameworthy deserve blame *in the same way* as the admirable deserve admiration or the fearsome merit fear, is likely to strike some as implausible. Specifically, I suspect that the most typical reaction is that the blameworthy must deserve blame in a deeper sense than in the case of nonmoral properties. While the fearsome may merit fear, this is not a deep evaluation of the fearsome; it is not due to the fearsome’s will or character, and it does not warrant the sorts of evaluations that seem to be distinctive of moral responsibility or desert. While there is an obvious difference between blameworthiness (and praiseworthiness) and the less-moralized normative properties, a natural thought is that this must be an important difference, especially when it comes to the matter of desert. Indeed, one might be inclined to think it a definitional implication of blameworthiness that the blameworthy deserve blame in a different sense from however the fearsome might deserve fear.

This mimics a well-established disagreement in the moral responsibility literature, one that begins with a distinction between moral and nonmoral evaluations.¹⁹ Ascriptions of blameworthiness or praiseworthiness possess a depth that other evaluations lack. While we may “grade” others on their physical appearance, intelligence or athletic aptitude, these judgments lack the robustness of *moral* assessments. Being judged ugly or stupid or clumsy constitutes a negative evaluation to be sure, but none of these superficial assessments imply fault on the agent’s part or license reproach or condemnation. Parallel observations could be made between superficial positive judgments (being beautiful, smart, lithe) and “deep” ascriptions of praiseworthiness or moral merit.

So, what can DAF say about the “depth” of desert? I think the view might address such depth in two ways. Consider how DAF treats the blameworthy, admirable and fearsome. All three deserve the respective attitudes of blame, admiration and fear. But they need not merit these attitudes for the

¹⁸ The extant views I have been considering share a commitment to understanding the fittingness of blame in moral terms. (Wallace speaks in terms of fairness; Darwall appeals to authority, which I consider a variation on the same theme.) This is characteristic of the only competing FA approaches of which I am aware. Certainly, different alternative FA accounts may be possible. I have tried to show why any appeal to an independent moral construal of “fittingness” is problematic, but I do not have the space here to consider every possible candidate.

¹⁹ See Levy (2005) and Smith (2008).

same reasons. The blameworthy deserve blame in virtue of a desert-entailing relation – namely, moral responsibility. The blameworthy are morally responsible for some action, say, and it is because of that relation that they deserve blame. On the opposite end, the fearsome might deserve fear because they are dangerous, or some other descriptive feature. Similarly, a separate story would be told of what makes the admirable deserve admiration.²⁰

Here, the “depth” of desert is captured not in the desert relation itself, but in what grounds that relation. For blameworthiness, this is the fact of the agent’s being morally responsible (plus, perhaps, that conduct being morally objectionable). But the fearsome obviously need not be morally responsible for their fearsomeness, nor for what grounds that fearsomeness. So, it is the moral responsibility relation that possesses the depth that judgments of fearsomeness lack. I hasten to note that it seems the admirable lie somewhere in between. It seems one can be admirable in a “shallower” sense than the blameworthy are assessable,²¹ and yet admiration surely goes deeper than a judgment of fearsomeness. If this is right, then it would be to DAF’s advantage to capture the depth of desert elsewhere, for it will be easier to accommodate the range of assessments, and their varying depths, in terms of the right reasons for each attitude.

A different but related strategy for capturing desert’s “depth” is to focus more specifically on the attitudes involved. Part of the depth of deserving blame, of being blameworthy, involves the target of blame. Blame not only targets the agent; more particularly, blame targets the agent’s moral qualities or attitudes or will. It is, after all, a moral assessment, and part of its depth, therefore, may be explained by reference to the importance that our moral practices have. The moral qualities of others have a particular significance that admirableness and fearsomeness lack.²² This observation implies a weightiness to our moral assessments of others, one that will not be found in nonmoral assessments, normative though they might be. I do not have suggestions for what admiration and fear target. Perhaps a similar line to the above could be used, so that admiration targets those qualities it is favorable to have (suitably restricted) and fear targets the qualities that pose a threat. Whether or not such a line of reply is ultimately successful, it does suggest, straightforwardly enough, how DAF can accommodate the thought that deserving blame has a special significance that meriting fear does not.²³

²⁰ Such a view looks similar to D’Arms & Jacobsen (2000).

²¹ One who follows Strawson (1962) in thinking that blame is constituted by a range of negative attitudes (e.g., guilt, resentment, indignation) might also suggest positive analogues constitutive of praise, and these might include admiration. If so, then the admirable will share the depth of blame, since to be admirable one will also be praiseworthy.

²² Or that admirableness possesses only to the degree that it is sensitive to moral considerations.

²³ Here I am following much of the literature on blame, which distinguishes it from a mere judgment (e.g., “that someone is blameworthy”) and from simple outward treatment (e.g., yelling at someone). Blame is best understood as a complex attitude involving both cognitive

Both of these strategies involve passing the buck a bit. It is something of a promissory note to say that the right kind of reasons for all the relevant attitudes will do double duty by also explaining (or justifying) the depth of certain assessments. If I were arguing for the outright adoption of DAF, I would consider these responses incomplete. But I think they succeed in suggesting defensible lines of reply, and thus support my actual aim, which is simply to investigate the promise of a conjectural view of desert. This task requires a lower standard for reasonable replies.²⁴

4. Objection 2: Denials of Desert

The second objection is an extension of the first. Suppose DAF were true. Then the blameworthy would deserve blame just as the fearsome merit fear. And denying someone what they deserve is at least *prima facie* wrong; indeed, it constitutes a wrong to that person. If you deserve the trophy, or to be promoted, then it is wrong for me to deny you the trophy or the promotion. I do you an injustice. But denying fear to the fearsome is not wrong. One might be mistaken, or foolish, but one has not done anything wrong. So DAF must be false if it implies that denying the fearsome fear is to do them wrong.

I do not find the objection problematic, as I do not find the view on which it is premised to be particularly plausible. Not blaming the blameworthy is not clearly a wrong to them, even though they may deserve it. Indeed, being able to “turn the other cheek” or replacing blame with understanding, while perhaps (at least at times) mistaken or foolish, can also be kind and virtuous.²⁵ If something like this is right, then the fact that it is not a wrong to the fearsome to deny them fear is a not-problematic implication of DAF.

5. Objection 3: The Threat of Vacuity

A final objection targets the basic schema of DAF: S is phi-able just in case S deserves/merits phi. DAF's schema was built by taking the general schema

and conative elements. For good discussions of the attitude of blame, see King & van Roojen (forthcoming); Scanlon (2008); Sher (2007).

²⁴ A related worry might be that DAF rules out a view in which (at least some) attitudes should be typed in part by their being deserved. For instance, Jones is merely irritated when his dog destroys his manuscript but is resentful or indignant when his colleague destroys it. One might be tempted to think that what makes the negative attitude in question one of resentment rather than mere irritation is just that it is deserved by the roommate but not by the dog. But even independent of FA accounts, this view should strike us as implausible. If “being deserved” was part of the very attitude of, say, indignation, then one could never be subject to undeserved indignation. A more likely story is that Jones feels resentment (and not mere irritation) because his attitude targets his roommate's moral qualities or attitudes or will. His dog plausibly lacks these qualities or attitudes, and so only irritation arises.

²⁵ For some discussion of the relation of blameworthiness and blame to forgiveness, see Wallace (1994: 72-3); Nelkin (2011: 42-50).

for FA accounts and replacing the generic “fitting” relation with the desert relation. A worry remains, however, that this replacement has not gained us much. The intuitive appeal of FA accounts when constructed with the generic “fitting” relation attends to all constructions using similar generics, like “appropriate,” “apt,” “suitable” and, the objection contends, “deserves.” And just as an FA account must specify how to understand the fitting relation, so, too, must DAF specify how to understand the desert relation so as to avoid the view being rendered as uninformative as the general schema. Thus, though DAF will analyze the admirable in terms of deserving admiration, this does no better than talk of “fitting” admiration or “appropriate” admiration. Granted, it is a normative relation, but all are equally vacuous in terms of helping to specify when admiration is deserved (or fitting, or appropriate, etc.). So the objection goes.²⁶

There is something to this worry, *if* one insists that DAF provide a comprehensive account of desert. But this is not DAF’s aim. Rather, the idea is to fill in a strategy for providing an account of desert that will make progress on illuminating the relationship between moral responsibility and desert. The very nature of FA accounts means that whatever notion takes the place of the generic relation will retain something of the generality of the basic schema. But as my main goal here is only to motivate a strategy for pursuing a positive account of desert, I do not take it as a shortcoming that DAF cannot provide a full accounting of the notion.

Nevertheless, the objection largely misses the mark. DAF’s use of desert is not as an empty placeholder; it is not the equivalent of fitting. Rather, we are to understand this large set of normative properties (those calling for FA analyses like “admirable,” “enviable,” “fearsome”) as deserving their respective attitudes, in the way that the blameworthy deserve blame (and vice versa). While this may generate some counterintuitive results (e.g., that the fearsome merit fear), these are not decisive objections, as I have argued. Moreover, the very fact that DAF can generate potentially counterintuitive results shows that the desert relation being employed is not a generic normative relation. If it were, there would be nothing at all objectionable about DAF, any more than there can be about the very general, generic FA schema that uses “fittingness.”

6. Desert and Fittingness

Thus far we have been considering a conjectural candidate view about desert. Accounts of moral responsibility ought to say something about desert. Moral responsibility also concerns blameworthiness and praiseworthiness, terms that independently suggest an FA analysis. FA accounts, in turn, require a normative relation to relate the analyzed properties and relevant attitudes, and desert is a normative relation of the right kind to play that role. Putting

²⁶ My thanks to Steven Sverdlik for raising this concern.

these thoughts together generates DAF: The desert relation relevant to moral responsibility just is the normative relation for all FA accounts of value, thus the blameworthy deserve blame in the same way the admirable deserve admiration and the fearsome merit fear. I have also considered a potential worry that it is too implausible to suppose that the desert relation employed throughout is the same, and the initial support for this objection is that blaming others involves a critique that is deeper than, say, fearing them. In response, I have suggested two strategies for accounting for the “depth” or special significance of moral evaluation, particularly blame and praise, consistent with DAF’s commitments. I have also suggested that DAF has plausible implications regarding the denials of deserved attitudes, and that it does give us an adequate *initial* characterization of desert, which makes DAF more substantive than the generic schema all FA accounts must employ.

I want to close with (i) a final observation, and (ii) a brief sketch of why giving an account of the desert relation is important. The heart of DAF lies in adopting desert as the basis for the fittingness relation all FA accounts of value require. I noted at the outset that desert, as a normative relation, is of the right kind to serve as the relation required by FA accounts. In this section, I want to say a bit more about why desert looks like a good fit. In the closing section, I will explore just one implication of DAF for extant debates in moral responsibility.

Desert independently suggests a notion of fittingness. Consider retributive theories of punishment, wherein desert plays a defining role. The punishment deserved is the one that “fits” the crime.²⁷ Moreover, we know that this notion of fit is meant to capture something different than repaying the criminal in kind. There is a sense of restoring a balance, of making the universe right again. Whether or not one adopts such a metaphorical position as criterial to an account of punishment, there is something about retribution that invites a sense of making punishment appropriate or fitting. The phrase “getting one’s just deserts” also suggests such fittingness. There is a natural inclination, I think, to suppose that part of the core of desert, in its widespread applications, involves achieving a certain state of affairs that is most appropriate to the circumstances. Punishment is just one possible instantiation of such a state of affairs.²⁸

Similarly, when we talk about competitions, we can readily distinguish winners from those who deserved to win. And when those who most deserve to win do in fact win, there often seems to be a special fittingness to

²⁷ As enshrined in the notion of *lex talionis*. I do not mean here to defend retributivism, only to note the connection between desert and fittingness that retributivism often invokes.

²⁸ I want to use “state of affairs” talk pretty loosely here. A state of affairs might limit an FA analysis to some persons blaming others as being fitting, say, rather than the attitude of blame (itself) fitting the blameworthy. Obviously, I want the account here to be neutral on this score, and so the notion of a state of affairs ought to be interpreted to service both kinds of cases.

that state of affairs. The underdog who worked hardest, overcame the most, played most fairly and still triumphed is a most fitting champion.²⁹

We also talk of people deserving each other. Sometimes we mean this as a tribute to a relationship, other times as a derisive slight. But in both cases we are appealing at least in part to a certain sort of fit. The individuals involved fit each other, and their being so related is a most fitting state of affairs.

None of these observations is confirming evidence for DAF. But I do take them to suggest that DAF may be on the right track. It at least seems to get some things right about desert, it is not obviously revisionary and it supplies a more attractive FA account of blameworthiness and praiseworthiness than its extant competitors. These reasons are sufficient for taking DAF seriously, and it seems worthwhile to investigate where developing such a view might lead.

7. DAF and Conceptions of Moral Responsibility

As I noted at the outset, virtually all theories of moral responsibility seem to be trying to give accounts of the same thing: responsibility in the “desert-entailing sense.” If an agent is morally responsible for some *x*, it follows that she can deserve praise or blame in virtue of being responsible for *x*. One reason to get clearer on this desert relation is simply because it is vital to a full understanding of moral responsibility and so little has thus far been said about it. But a second reason to seek a positive account of the desert relation is because such an account will plausibly bear on the suitability of extant accounts of moral responsibility. To gesture toward just one possible implication, I want to consider two different conceptions of moral responsibility from the literature.

Gary Watson has famously argued for distinguishing between two conceptions (or “faces”) of moral responsibility.³⁰ The first he calls responsibility as attributability, or its “aretaic face.”³¹ When we speak of actions that an agent is responsible for, our concern is sometimes that those actions express the agent’s commitments or values or that they are reflective of her character. Responsible action is attributable to the agent as a self-reflective and evaluative creature.

The second notion Watson calls responsibility as accountability. Here what matters is holding each other responsible, which requires more than some relation of an individual to his conduct: “[I]t also involves a social setting in which we demand (require) certain conduct from one another and respond adversely to one another’s failures to comply with these demands.”³²

²⁹ As may be the champion who simply performed perfectly.

³⁰ Watson (1996).

³¹ Ibid.: 229.

³² Ibid. .

These two notions, Watson tells us, come apart. One can be accountable for some conduct without being responsible for it in the attributability sense, as in cases of strict liability (or, perhaps, negligence). Similarly, one could be responsible for some bit of behavior in the attributability sense, and yet fail to be accountable to us, because we lack the standing to hold her responsible for it.³³ Similarly, they differ with respect to blame. To blame Jones in the aretaic sense is simply to view Jones to have acted badly, in a way that can reflect on him as a moral agent. Accountability blame, on the other hand, goes further: It implies sanction-like behavior. This makes additional considerations relevant. For instance, since sanctions are sensitive to issues of, say, fairness, avoidability of the action appears relevant to accountability blame in a way it is not for aretaic blame.³⁴

If both faces are truly conceptions of moral responsibility, we should expect them to share certain features, especially when it comes to entailing desert. But examining this expectation suggests two observations. First, the two “faces” of responsibility are not related in the same way to desert. If moral responsibility is responsibility in the desert-entailing sense, we should expect any reasonable conception of moral responsibility to support the following inference:

If S is morally responsible for x, then S deserves something (in principle) in virtue of being so related to x.

But as attributability and accountability each involve patently *different* relations between agents and their actions, it is implausible to suppose that both could sustain the same conception of desert. For instance, since considerations of fairness matter asymmetrically to accountability, fairness matters to what agents deserve in terms of their accountability but not necessarily to their attributability. If this is right, then it seems we have grounds for thinking two corresponding conceptions of desert exist, one for each conception of responsibility. Such a result might call into question the very distinction at hand, causing us to reevaluate whether attributability and accountability represent two conceptions of the same concept – moral responsibility – or represent two closely related concepts, one or the other of which might correspond to the target concept in debates about free will and moral responsibility.

The second observation is that it may be that one face cannot plausibly entail desert at all. Derk Pereboom, for instance, claims that a theory of moral responsibility is a theory about what he calls “basic desert.” He has argued

³³ Watson’s example is an agent who betrays her values (231). We can legitimately appraise her in light of this conduct, rightly thinking her cowardly and dishonorable, while also rightly thinking that, in this matter, it is nobody’s business what she does with her life, and so we lack any grounds for holding her to account.

³⁴ Recall the special place of fairness and similar considerations for Wallace’s and Darwall’s accounts. They both seem preoccupied with accountability rather than attributability.

that no one is responsible for anything, because to be the true source of one's actions would require agential powers that are empirically implausible. But in denying that anyone is responsible for anything, Pereboom only means to deny that anyone deserves blame and praise for what they do. He is perfectly willing to accede that we could be responsible in the sense of being merely accountable to each other for our conduct. But this would be responsibility of a weaker sort than that required to justify ascriptions of deserved blame and praise.³⁵ So, Pereboom wants to draw a line: He thinks accountability is insufficient for basic desert. If he's right that basic desert is the sort of desert to be entailed by the true account of moral responsibility, accountability will not, after all, be a proper conception.

We need not settle these matters here to appreciate their significance. While the ability to make sense of desert entailment may be independently assessed for each conception Watson proposes, DAF and its competitors, once developed, provide a context within which to evaluate the implications for desert of extant views of moral responsibility and their commitments. A particular virtue of DAF is that it draws from a perfectly general account of analyzing a large set of related normative properties, of which blameworthiness and praiseworthiness are but a tiny subset. Thus, if DAF is defensible in its own right, it proposes to provide a theoretically advantageous tool for use in current debates.³⁶

Matt King
University of California, Los Angeles
UCLA School of Law
mail.mattking@gmail.com

³⁵ See Pereboom (2007: 86).

³⁶ I would like to thank Michael McKenna, Mark Schroeder and two anonymous referees for comments on a previous draft. A version of this paper was presented at St. Bonaventure University, the University of Minnesota-Duluth and the third annual Rocky Mountain Ethics Congress. I would like to thank those audiences for their comments and questions, as well as Steven Sverdlik for his extremely helpful commentary.

References

- D'Arms, J. and Jacobsen, D. (2000) "The Moralistic Fallacy: On the 'Appropriateness' of Emotions," *Philosophy and Phenomenological Research* 61: 65-90.
- Darwall, S. (2006) *The Second-Person Standpoint: Respect, Morality, and Accountability*, Cambridge, Mass.: Harvard University Press.
- Enoch, D. (2004) "Why Idealize?" *Ethics* 114: 759-87.
- Hieronymi, P. (2005) "The Wrong Kind of Reasons," *Journal of Philosophy* 102: 437-57.
- King, M. and van Roojen, M. (forthcoming) "Praising and Blaming," in M. Sbisà and K. Turner, eds., *Pragmatics of Speech Actions*, Berlin: De Gruyter Mouton.
- Levy, N. (2005) "The Good, the Bad and the Blameworthy," *Journal of Ethics & Social Philosophy* 1(2): 1-16.
- Nelkin, D. (2011) *Making Sense of Freedom and Responsibility*, Oxford: Oxford University Press.
- Parfit, D. (2001). "Rationality and Reasons," in D. Egonsson, J. Josefsson, B. Petersson, T. Rønnow-Rasmussen and I. Persson, eds., *Exploring Practical Philosophy: From Action to Values*, Burlington, Vt.: Ashgate, pp. 17-39.
- Pereboom, D. (2007) "Hard Incompatibilism," in J. M. Fischer, R. Kane, D. Pereboom and M. Vargas, *Four Views on Free Will*, Malden, Mass.: Blackwell, pp. 85-126.
- Pojman, L. (1997) "Equality and Desert," *Philosophy* 72: 549-70.
- Rabinowicz, W. and Rønnow-Rasmussen, T. (2004) "The Strike of the Demon: On Fitting Pro-Attitudes and Value," *Ethics* 114: 391-423.
- Scanlon, T. (2008) *Moral Dimensions: Permissibility, Meaning, Blame*, Cambridge, Mass.: Harvard University Press.
- Schroeder, M. (2010) "Value and the Right Kind of Reason," *Oxford Studies in Metaethics* 5: 25-55.
- Sher, G. (2007) *In Praise of Blame*, New York: Oxford University Press.
- Smith, A. (2007) "On Being and Holding Responsible," *Journal of Ethics* 11: 465-84.
- ____ (2008) "Control, Responsibility, and Moral Assessment," *Philosophical Studies* 138: 367-92.
- Strawson, P. F. (1962) "Freedom and Resentment," *Proceedings of the British Academy* 68: 1-25.
- Wallace, R. J. (1994) *Responsibility and the Moral Sentiments*, Cambridge, Mass.: Harvard University Press.
- Watson, G. (1996) "Two Faces of Responsibility," *Philosophical Topics* 24: 227-48.
- Way, J. (2012) "Transmission and the Wrong Kind of Reason," *Ethics* 122: 489-515.
- Zaibert, L. (2006) "The Fitting, the Deserving, and the Beautiful," *Journal of Moral Philosophy* 3: 331-50.