

# Consequentialism and Moral Worth

NATHANIEL SHARADIN

*The College of New Jersey*

Sometimes, agents do the right thing for the right reason. What's the normative significance of this phenomenon? According to proponents of the special status view, when an agent acts for the right reason, her actions enjoy a special normative status, namely, worthiness. Proponents of this view claim that self-effacing forms of consequentialism cannot say this plausible thing, and, worse, are blocked from having a perspicuous view of matters by the self-effacing nature of their consequentialism. In this article, I argue that this claim is based on an illicit assumption. I show that whatever version of the special status view proponents of that view prefer, self-effacing consequentialists can adopt a version of it. Moreover, I show that proponents of extant versions of the special status view have reason to prefer the specific version of that view I articulate on behalf of self-effacing consequentialists.

## I. INTRODUCTION: THE PHENOMENON OF MATCH

It's best to start with the bad cases. Consider:

*Airport:* Marguerite asks her friend, Cyril, for a ride to the airport on Friday. Cyril agrees because, that way, Marguerite will agree to do the same for him next week.

*Theft:* James notices that someone left their laptop in his classroom. He considers stealing the laptop, but decides not to take it because he thinks he'll probably be caught if he does.

*Stocks:* KT is wondering whether to invest her life-savings in an individual stock or in a broad-based, low-cost index fund. She decides to invest in the index fund because she prefers the ticker symbol.

Cyril, James, and KT each do what's right: Cyril takes his friend to the airport, James doesn't steal the laptop, and KT makes the correct investment decision. But intuitively, Cyril, James, and KT's actions are nevertheless normatively problematic. In what way? As a first stab at it, we can say that in each of these cases Cyril, James, and KT do what's right *out of the wrong motivation*.

What, precisely, does it mean to say that an agent does what's right 'out of the wrong motivation'? Let's sharpen up our language. Call an agent's reasons for doing what she does, when she does something on the basis of reasons, her *motivating reasons*. An agent's motivating reasons are the reasons *for which* she acts. A proviso: this is not meant as an analysis of what motivating reasons are. Attempts at giving such analyses are fraught, and anyway are beyond the scope of this

article.<sup>1</sup> For the purposes of this article, we can rely on our intuitive grasp of the notion of reasons for which an agent acts, whatever, in the end, this kind of thing turns out precisely to be.<sup>2</sup> Now we can describe the trouble with Cyril, James, and KT in the following way: their motivating reasons are somehow askew.

Asked from what? From the *right reasons* for their respective actions. (I'll say more about the nature of the right reasons below, in [section II](#).) When an agent's motivating reasons fail to match the right reasons, the resultant action suffers some normative flaw. This is what happens in Cyril, James, and KT's cases. And, symmetrically, when an agent's motivating reasons do match the right reasons, the agent's action is normatively better off than it otherwise would be.

This idea, that there's something normatively important about (failures of) match between an agent's motivating reasons and the

<sup>1</sup> This way of characterizing motivating reasons is relatively uncontroversial. For a very incomplete list of those who characterize motivating reasons in roughly this way, compare: J. Dancy, *Practical Reality* (Oxford, 2003); T. Scanlon, *What We Owe To Each Other* (Harvard, MA, 1998); J. Raz, *From Normativity to Responsibility* (Oxford, 2011); D. Parfit, *On What Matters: Volumes I & II* (Oxford, 2011); T. Scanlon, *Being Realistic About Reasons: The John Locke Lectures* (Oxford, 2014); M. Smith, *The Moral Problem* (Oxford, 1994); J. Broome, 'Reasons', *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*, ed. J. Wallace, et al. (Oxford 2004), pp. 28–55; D. Whiting, 'Against Second-Order Reasons', *Nous* 49.4 (2016), pp. 398–420, and the essays collected in both *Reasons for Action*, ed. D. Sobel and S. Wall (Cambridge, 2011) and *Reasons for Belief*, ed. A. Reisner and A. Steglich-Petersen (Cambridge, 2011). For more detailed accounts of the nature of motivating reasons in both the practical and epistemic contexts, see M. Schroeder, *Slaves of the Passions* (Oxford, 2007); Scanlon, *Owe*; Parfit, *Matters*; J. Turri, 'The Ontology of Epistemic Reasons', *Nous* 43.3 (2009), pp. 490–512; S. Kearns and D. Star, 'Reasons as Evidence', *Oxford Studies in Metaethics Volume 4*, ed. R. Shafer-Landau (Oxford, 2009), pp. 215–42; P. Hieronymi, 'The Wrong Kind of Reason', *Journal of Philosophy* 102.9 (2005), pp. 437–57; J. Dancy, *Ethics without Principles* (Oxford, 2004); M. Smith, 'The Humean Theory of Motivation', *Mind* 96.381 (1987), pp. 36–61; M. Smith, 'A Constitutivist Theory of Reasons: Its Promise and Parts', *Law, Ethics, and Philosophy* 1 (2013), pp. 9–30; Dancy, *Reality*; J. Hyman, 'How Knowledge Works', *Philosophical Quarterly* 49 (1999), pp. 433–51; R. Neta, 'Treating Something as a Reason For Action', *Nous* 43.4 (2009), pp. 684–99; N. Sharadin, 'Reasons Wrong and Right', *Pacific Philosophical Quarterly* 97.2 (2015), pp. 371–99; N. Sharadin 'Nothing But the Evidential Considerations?', *Australasian Journal of Philosophy* 94.2 (2016), pp. 343–61; and the essays collected in *Motivational Internalism*, ed. G. Bjornsson, et al. (Oxford, 2015).

<sup>2</sup> Equivalently, we might rely on our intuitive grasp of the considerations on which the agent's action is based. This is the terminology epistemologists tend to prefer in the symmetrical case of belief. See, for instance, G. Harman, *Thought* (Princeton, 1973); G. Pappas, 'Basing Relations', *Justification and Knowledge* (Dordrecht, 1979); M. Swain, 'Justification and the Basis of Belief', *Justification and Knowledge* (Dordrecht, 1979); J. Tolliver, 'Basing Beliefs on Reasons', *Grazer philosophische Studien* 15 (1982), pp. 149–61; J. Kvanvig, 'Swain on the Basing Relation', *Analysis* 45.3 (1985), pp. 153–8; L. Lemke, 'Kvanvig and Swain on the Basing Relation', *Analysis* 46.3 (1986), pp. 138–44; K. McCain, 'The Interventionist Account of Causation and the Basing Relation', *Philosophical Studies* 159 (2012), pp. 347–82. Here I will stick to the locution of the reasons for which an agent acts.

right reasons, isn't a new one. The sentiment is given voice in the common-sense platitude that agents should do the right thing *for the right reason*.<sup>3</sup> And this common-sense idea, that there's something normatively important about cases of match between agents' motivating reasons and the right reasons, is echoed in a great deal of philosophy.<sup>4</sup> Here, I won't be defending the claim that match matters normatively. Instead, I'm going to assume that common sense,

<sup>3</sup> Despite what the slogan might suggest, the thought that agents' motivating reasons should match the right reasons does not involve the thought that agents should act for the right kind of reasons. So, the question of how the so-called 'right' and the 'wrong' kind of reasons interact is orthogonal to the issue I'm interested in here, which is the match between the right reasons (of any kind) and motivating reasons. See Sharadin, 'Wrong'; Hieronymi, 'Wrong Kind'; J. D'Arms and D. Jacobson, 'Wrong Kinds of Reasons and the Opacity of Normative Force', in *Oxford Studies in Metaethics Volume 9*, ed. R. Shafer-Landau (Oxford, 2014), pp. 215–42; J. Olson, 'The Wrong Kind of Solution to the Wrong Kind of Reason Problem', *Utilitas* 21.2 (2009), pp. 225–32; Schroeder, *Slaves*; A. Reisner, 'The Possibility of Pragmatic Reasons for Belief and the Wrong Kind of Reasons Problem', *Philosophical Studies* 145 (2009), pp. 257–72; J. D'Arms and D. Jacobson, 'The Moralistic Fallacy: On the 'Appropriateness' of Emotions', *Philosophy and Phenomenological Research* 61.1 (2000), pp. 65–90, for discussions of the distinction between the right and the wrong kind of reasons.

<sup>4</sup> In the *Groundwork* Kant famously seems to have held that only actions done for the sake of duty possess moral worth. See especially I. Kant, *Groundwork of the Metaphysics of Morals* (Cambridge, 1785), pp. 397–8. Neo-Kantian ethicists, though they are careful to refine or hedge Kant's claim, often endorse something similar. See R. G. Henson, 'What Kant Might Have Said: Moral Worth and the Overdetermination of Dutiful Action', *The Philosophical Review* 88.1 (1979), pp. 39–54; P. Stratton-Lake, *Kant, Duty, and Moral Worth* (Oxford, 2000); C. Korsgaard, 'From Duty and for the Sake of the Noble: Kant and Aristotle on Morally Good Action', *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty*, ed. S. Engstrom and J. Whiting (Cambridge, 1996); J. Hernandez, 'Impermissibility and Kantian Moral Worth', *Ethical Theory and Moral Practice* 13.4 (2010), pp. 403–19; K. Simmons, 'Kant on Moral Worth', *History of Philosophy Quarterly* 6.1 (1989), pp. 85–100; T. E. Hill, *Human Welfare and Moral Worth: Kantian Perspectives* (Oxford, 2002). Non-Kantians also recognize the import of match. See, for instance, J. Markovits, 'Acting for the Right Reasons', *Philosophical Review* 119.2 (2010), pp. 201–42; N. Arpaly, *Unprincipled Virtue: An Inquiry into Moral Agency* (Oxford, 2002); N. Arpaly, 'Moral Worth', *Journal of Philosophy* 99.5 (2002), pp. 223–45; P. Sliwa, 'Moral Worth and Moral Knowledge', *Philosophy and Phenomenological Research* 92.2 (2015), pp. 393–418; K. Sorensen, 'Effort and Moral Worth', *Ethical Theory and Moral Practice* 13.1 (2010), pp. 89–109; J. Gert, 'Moral Worth, Supererogation, and the Justifying/Requiring Distinction', *Philosophical Review* 121.4 (2012), pp. 611–18; M. Schroeder, 'Knowledge is Belief for Sufficient (Objective and Subjective) Reason', *Oxford Studies in Epistemology* (Oxford, 2013), pp. 226–52. I'll discuss Markovits's and Arpaly's views in more detail, below. The difference shows up in unexpected places too: it's in Rawls's preference for what he calls an 'overlapping consensus', in which agents' political behaviour is based on the right (political) reasons, over a mere 'modus vivendi', in which, although agents' political behaviour is the same, it is based on their idiosyncratic interests. See J. Rawls, 'The Idea of an Overlapping Consensus', *Oxford Journal of Legal Studies* 7.1 (1987), pp. 1–25. On the epistemic side, things are perhaps even more stark. This is because match between the right reasons and motivating reasons is usually assumed to be a necessary condition for epistemic justification, which in turn is thought to be a necessary condition for knowledge. For an especially clear statement of this idea, see P. Bondy, 'Counterfactuals and Epistemic Basing Relations', *Pacific Philosophical Quarterly* 96.4 (2015), pp. 542–69. For scepticism on this point, see P. Silva, 'Does

and the philosophers who echo it, have things roughly correct. There's *something* normatively important about cases where agents do the right thing for the right reason.<sup>5</sup>

Can we say anything more philosophically illuminating about this normative importance? The orthodox view about what, precisely, is normatively important about cases of match is what I'll call the *special status view*. According to the special status view, there's a special normative status, viz. *worth* or *worthiness*, that an action enjoys when an agent's motivating reasons match the right reasons. My interest here isn't in the special status view *per se*. Instead, I'm interested in a claim made by proponents of that view, notably Julia Markovits and Nomy Arpaly, to the effect that certain consequentialists cannot avail themselves of the special status view, and moreover are blocked, by their consequentialism, from giving an alternative, philosophically perspicacious view of matters. These consequentialists, according to this line of thought, are faced with a dilemma: either fail to account for the normative importance of cases of match, or give up their normative theory. In this article I'll explain how consequentialists can avoid this dilemma.

Here is the detailed plan. In [section II](#) I provide the details we'll need regarding the special status view. I'll then ([section III](#)) explain why proponents of the special status view such as Markovits and Arpaly think certain consequentialists are incapable of availing themselves of this view and hence face the dilemma noted above. Next ([section IV](#)), I'll explain how consequentialists should respond to the dilemma. Here, I'll also sketch an argument to the effect that proponents of the special status view themselves have a reason to take the line I suggest on behalf of consequentialism. In the penultimate section ([V](#)) I'll consider and reply to four objections. The conclusion of the article ([section VI](#)) is that, contrary to what proponents of the special status view say, consequentialists don't have any special difficulties explaining the normative importance of cases of match between motivating and right reasons. Let's get on with it.

Doxastic Justification Have a Basing Requirement?, *Australasian Journal of Philosophy* 2 (2014), pp. 1–17.

<sup>5</sup> This phenomenon isn't limited to action. Similar remarks apply, *mutatis mutandis*, to attitudes for which there can be reasons, e.g. belief, resentment, admiration, envy, etc. If we think there can be reasons for these attitudes, then we should be able to distinguish between the right reasons for having these attitudes and the reasons for which an agent has them. And it is possible for these two sets of reasons either to match or to fail to match. To ease exposition in this article, I'll focus my attention on the phenomenon as it occurs in the case of action, though I will, in these notes, sometimes highlight a parallel with the case of belief.

## II. THE SPECIAL STATUS VIEW: A QUICK PRIMER

Recall, the cases we're interested in are ones where an agent's motivating reasons (fail to) match the *right reasons* for their action. According to the special status view, the normative importance of cases of (failure to) match is that such (failure to) match (fails to) imbue the action in question with a special kind of normative status, viz. *worthiness*.<sup>6</sup> When an agent's motivating reasons for an action match the right reasons, the agent's action is thereby in some way worthy.

But what, precisely, are the *right reasons* for an action? According to proponents of the special status view, the right reasons for an action are the right-making reasons for that action, where the right-making reasons for an action are understood as the *normative* reasons, or the reasons that *justify* the action.<sup>7</sup> This idea is intuitively attractive, and it's borne out by thinking about cases. For instance, in Airport: the right reason for Cyril to take Marguerite to the airport is (something like) *that she's his friend*. Equally, the reason that makes taking Marguerite to the airport the right thing for Cyril to do is (something like) *that she's his friend*. Intuitively, then, there's a coincidence between right reasons and right-making reasons. The present idea is that this coincidence isn't just common, it's universal: right reasons just *are* right-making, i.e. normative, reasons. This idea about what the right reasons are is accepted by all proponents of the special status view.<sup>8</sup> To anticipate: it will be problematic later on. But for now, we should simply accept it for the sake of understanding the account.

So, what does the special status view say about cases of match? The basic idea of the special status view is summed up nicely by Julia Markovits; she argues for the:

*Coincident Reasons Thesis*: An action is morally worthy if and only if the agent's motivating reasons coincide with the reasons morally justifying the action.<sup>9</sup>

I've been speaking of 'match' rather than 'coincidence' between reasons, but it's clear that Markovits's idea in the Coincident Reasons Thesis

<sup>6</sup> Going forward, to ease exposition, I'll just focus on cases of match and ignore *failures* to match, with the understanding that cases of failure should be understood in a symmetrical manner.

<sup>7</sup> As above, this is not meant as an analysis of what normative reasons for action are. That issue is fraught, and, again, beyond the scope of this article. See Schroeder, *Slaves*; Scanlon, *Owe*; Parfit, *Matters*; Turri, 'Ontology'; Kearns and Star, 'Reasons'; Dancy, *Principles*; Smith, 'Constitutivist'; and Dancy, *Reality* for some attempts to analyse normative reasons.

<sup>8</sup> Markovits, 'Acting' and Arpaly, 'Worth' and *Unprincipled* are perhaps clearest in their acceptance of this idea.

<sup>9</sup> Markovits, 'Acting', p. 204.

is the same as the one we've been considering. Note too that while Markovits's principle is explicitly targeted at instances of *moral* worth, there's no reason we need to restrict our attention to such cases. After all, given that normative reasons can plausibly come in a range of different kinds (e.g. prudential, aesthetic, moral, epistemic), it's reasonable to suppose agents' actions (and attitudes) can equally enjoy a range of different kinds of worth.<sup>10</sup> In any case, going forward, my focus will mainly be on instances of moral worth; so, the idea is that agents' actions have moral worth when their motivating reasons coincide, i.e. match, the moral reasons that justify (i.e. are moral normative reasons in favour of) the action.

Supposing that something like the Coincident Reasons Thesis describes the conditions under which an action enjoys moral worth, we might wonder what the normative implications of an action's having this special status are meant to be. In other words: Why is having this status in any way *special*? Markovits provides some guidance, saying that: 'Morally worthy actions (the thought is) aren't just right actions – they are actions for which the agent who performs them *merits praise*. . . . Morally worthy actions are ones that reflect well on the moral character of the person who performs them.'<sup>11</sup> Nomy Arpaly, also a proponent of the special status view, says much the same, claiming that morally worthy actions are 'praiseworthy actions' or 'esteem-worthy actions'.<sup>12</sup> Hence the normative significance of actions done for the right reasons is that such actions, according to the current line of thought, reflect on the character of the agent, and hence merit a kind of praise corresponding to the kind of reason for which the agent acted.

And *that's it*; that's all we need to know about the view for what's to come. As I said, my aim here is only to sketch *very briefly* the details of the special status view. In the context of what's to come, I'm interested in arguing that consequentialists can adopt a version of the special status view in reply to the dilemma purportedly facing them. Since I'll end up presenting a recipe for cooking up a version of the special status view given *any* actual ingredient special status view, proponents of the

<sup>10</sup> It's worth noting here that Markovits also comes close to endorsing the epistemic corollary of the Coincident Reasons Thesis, saying that 'Something like the Coincident Reasons Thesis plausibly describes the conditions under which our beliefs are epistemically worthy or justified. Our beliefs have epistemic worth – are epistemically justified (not just justifiable) – when we believe them for the epistemic reasons why we ought to believe them – when, that is, we believe them in response to our evidence' ('Acting', p. 214). Mark Schroeder, 'Knowledge is Belief', also appears to endorse something like CRT and the special status view when it comes to the epistemic worth of belief. Here, as I've said, my focus is on the case of action.

<sup>11</sup> Markovits, 'Acting', p. 203, emphasis in original.

<sup>12</sup> N. Arpaly, 'Moral Worth and Normative Ethics', *Oxford Studies in Normative Ethics*, ed. M. Timmons (Oxford, 2015), pp. 86–105; Arpaly 'Worth'.



special status view who for any reason are not happy with the sketch of their view just given are invited to ignore the details of what I've just said and start paying attention right *now*.

### III. THE CONSEQUENTIALIST'S DILEMMA

According to proponents of the special status view, the special status view is unavailable to certain consequentialist theories, in particular versions of consequentialism that are *self-effacing*. A self-effacing consequentialist view, familiarly, is one where the theory's account of the criterion of right action – net contribution to happiness, say – is divorced from the theory's account of the right decision procedure in action – following the advice of one's elders, say.<sup>13</sup> The theoretical pressure to divorce these two notions is also familiar. It's that, as it happens, things go quite poorly *from the point of view of happiness* when agents systematically act on the basis of their judgements about their actions' net contributions to happiness.<sup>14</sup> Here's Markovits making this point regarding a particular set of behaviours:

Promise keeping, for example, could not survive among agents who lost their motivation to keep their promises as soon as they became convinced that utility would be better served by breaking them. 'I promise I'll do it' would come to be no more reliable an indicator of future intentions than 'I'll do it if I judge it best.'<sup>15</sup>

And, the thought continues, it would be quite a bad thing indeed *from the perspective of maximizing net happiness* were the institution of promising to erode in this way. The worry is not meant to be restricted to the institution of promises. Instead, the worry is the quite general one that directly attempting to do what consequentialism says we ought to do, e.g. maximize happiness, would, on the whole, be *worse* at achieving what this view says it would be best to achieve than if we didn't directly attempt to do this and instead simply did what, say,

<sup>13</sup> As recommended in, for instance, H. Sidgwick, *The Methods of Ethics*, 7th edn. (London, 1907). It is a delicate matter whether we think such self-effacing, or, in Williams's phrase, 'Government House' consequentialist views are acceptable. Here, I won't be addressing this fraught issue directly. Instead, I'm going to argue that a particular charge against such views, viz. that they make a mess out of our judgements concerning the moral worth of actions, doesn't stick. For an excellent overview of the topic, including a defence of self-effacing views quite generally, see B. Eggleston, 'Rejecting the Publicity Condition: The Inevitability of Esoteric Morality', *The Philosophical Quarterly* 63 (2013), pp. 29–57.

<sup>14</sup> Of course, the particular version of this pressure will depend on the theory's choice of axiology. Here I'm assuming a version of hedonism, just to have something with which to work.

<sup>15</sup> Markovits, 'Acting', pp. 231–2.

common-sense morality says we ought to do, e.g. keep our promises, tell the truth, refrain from violence, etc.

For a consequentialist view that takes this worry seriously, one familiar manoeuvre is to divorce the criteria of right *action* from the criteria of right decision or right *motivation*: in other words, to pull apart the right-making reasons from the motivating reasons recommended by the theory. Such self-effacing theories, i.e. ones that *background* their account of the right-making reasons and foreground some non-identical account of the correct motivating reasons, cannot, according to proponents of the special status view, avail themselves of that view. In short, this is because if we are self-effacing consequentialists then it becomes utterly mysterious why we should think match between an agent's motivating reasons and the right-making reasons is supposed to be a good thing. After all, it's not even recommended by the theory itself!

Self-effacing consequentialist views therefore appear to occupy an unenviable, quite awkward position. For, recall, the phenomenon to be explained is the phenomenon whereby an agent's motivating reasons for her action sometimes match the right reasons for the action. And the puzzle was one of explaining the manifest normative significance of this phenomenon. As we saw, the special status view does so by articulating the way in which match between an agent's motivating reasons and the right reasons engenders a positive normative status, viz. *worth*, in agents' actions. But this view appears to be unavailable to self-effacing consequentialists; worse, it appears to be a Bad Thing when agents' motivating reasons match the right reasons!

Here is why. For self-effacing consequentialists, an action's being worthy as a result of match is a sign that *something has gone wrong*. (The self-effacement hasn't worked.) And, symmetrically, an action's failing to be worthy is, in that respect, a sign that *something has gone right*. (Self-effacement achieved!) It would be better were agents to be motivated in action by reasons other than their normative moral reasons. Hence, worth, engendered by match between normative reasons and motivating reasons, signals a normative *problem* to be corrected, rather than (as the special status view has it) a positive normative status to be pursued. Similarly, lack of worth, engendered by a lack of match, signals a positive normative status to be pursued, and not (as the special status view has it) a normative problem to be corrected.

In effect, then, the charge against self-effacing views is that they've made a muddle out of the phenomenon we're interested in here and hence have failed to provide a perspicuous account of it. Worse, self-effacing consequentialist views appear to be blocked, by the structure



of their view, from *any* perspicuous view of matters, since they think that, quite generally, *agents ought not to be motivated by the normative reasons*.

Markovits and Arpaly, two proponents of the special status view, are especially clear about the purported trouble. Here is Markovits:

So if the [special status view] is correct, and if utilitarianism is indeed [self-effacing], then even in a good world – perhaps even the best world accessible to us from our actual world – morally worthy actions will be very rare. This strikes me as a highly implausible conclusion. . . . [I]t is an objection to a normative theory if it entails, as I have argued utilitarianism does, that even in the best circumstances people will usually have to act *unworthily* to act *rightly*.<sup>16</sup>

Arpaly is largely on board with Markovits's criticism, saying that:

[I]f utilitarianism were correct then people who keep promises and are not motivated by [right-making reasons] would act with no moral worth. One way to put it is to say that she would act with no more moral worth than Kant's prudent grocer. . . . But despite what one would expect if utilitarianism were true, it is implausible to think that everyone who keeps promises for [reasons other than consequentialist reasons] is equivalent to the grocer in the moral worth of his or her actions. Yet, if utilitarianism were correct, it would follow that these people, just like the grocer, act for morally irrelevant reasons and not for the right-making features of their actions, and so act without moral worth.<sup>17</sup>

Markovits and Arpaly are both troubled by the same thing: if self-effacing consequentialism is correct, then, most of the time, most of what intuitively seems like (morally) worthy action will not be so. Worse, in Markovits's slogan, self-effacing views are forced to say that people will usually have to act *unworthily* (i.e. for non-right-making reasons) in order to act *rightly* (i.e. in accord with the right-making reasons). But, intuitively at least, that's quite wrong.

What's a self-respecting self-effacing consequentialist to do? One line of response would be to reject the intuitions driving the objection by rejecting the notion of worth altogether. Perhaps we could hold that the idea of 'worth' is a hold-over from Kantian or religious ethics. But as Markovits points out, that's a bit drastic, and anyway:

[I]t simply isn't true that the idea of moral worth is merely a manufactured Kantian artifact – on the contrary, it is an important part of our everyday practice of moral judgment. . . . [T]he concept of a morally worthy act is closely bound up with other centrally important moral concepts, including those of praiseworthiness and blameworthiness, and, of course, that of the good or virtuous person. A utilitarian cannot abandon the notion of moral

<sup>16</sup> Markovits, 'Acting', p. 233–5, emphasis added.

<sup>17</sup> Arpaly, 'Normative Ethics', pp. 4–5.

worth without embracing an error theory about much of our ordinary moral discourse.<sup>18</sup>

The fact that the notion of moral worth is bound up in these ways in our ordinary practice of moral judgement is probably sufficient to show that it is not merely a ‘manufactured Kantian artifact’. But I am not sure it shows that the notion of moral worth is not an objectionable holdover from religious ethics, or more broadly from ethical traditions that emphasize the ‘internal’ or motivational aspects of agents’ behaviour over the outcomes delivered by their actions. In any case, this is not the place to have that argument. Or rather, though this might be as good a place as any to stage that fight, I’ll put it off to another time. Markovits is surely correct that taking this line would involve embracing an error theory about a large portion of our moral discourse. On the assumption that we should avoid doing so wherever possible, we can grant the general point, at least for the sake of argument. The counterintuitive results noted above are the bathwater, and the phenomenon of worth is the baby. What the self-effacing consequentialist needs to do is clear, but how to do so isn’t. In the next section I’ll explain how she should proceed.

#### IV. SOLVING THE PROBLEM

Recall the slogan: agents should do the right thing *for the right reason*. In discussing the special status view, I’ve been assuming, with proponents of that view, that the ‘right reasons’ are, in Markovits’s phrase, *right-making* reasons: they’re reasons that make it the case that the agent’s action is the right thing to do. In other words, the *right* reasons are *normative* reasons. Hence match that engenders worth, as we’ve understood it thus far, is between an agent’s motivating reasons and the normative reasons. But as I’ll now explain, this particular understanding of the phenomenon is not obligatory; indeed, understanding match in this way runs together two issues best kept separate because (I’ll suggest) it assumes that a particular kind of first-order normative theory is correct, an assumption that’s illicit in the present context.

To see the illicit assumption, it’ll be helpful to return to the beginning. Recall, the common-sense phenomenon with which we began was that, holding fixed the normative correctness of an agent’s action, some motivations for that action were normatively superior to others. (Recall KT, James, and Cyril.) So we naturally have

<sup>18</sup> Markovits, ‘Acting’, p. 236.

a distinction between *correct action* and *correct motivation*. We're familiar with the idea that the *right-making*, i.e. normative, reasons are the reasons that make it the case that this or that action is correct. Let's add to this the idea that the *right reasons* (as in: agents should do the right thing *for the right reasons*) are the reasons the acting for which constitutes the correct motivation.

Our puzzle was one of explaining what was required in order for an agent's motivation to make a difference to the normative status of the action, and what sort of difference it made. The special status view comprises two claims. First: *correct motivation engenders worth*. This is a claim about what happens, normatively speaking, when an agent is motivated correctly, i.e. when she's motivated by the right reasons. Second: *right reasons are identical to right-making reasons*. This is a claim about what correct motivation comprises, viz. that being motivated by the right-making (normative) reasons is what's motivationally correct.

But putting things this way should make it clear that, in principle at least, it's possible to separate the first claim from the second. In other words, it's possible to agree with the special status view that correct motivation engenders worth (and that's why it's normatively important) while disagreeing with a *particular version of that view* over what correct motivation actually comprises. Indeed, as I'll now explain, the special status view had probably better not even *include a view* on what correct motivation actually comprises – at least not if it's being used, as Markovits and Arpaly are using it, in the course of an argument against a range of first-order normative theories such as self-effacing consequentialism.

To see why this is, notice that a complete first-order normative theory isn't just in the business of telling us what (it's right) to do, though it had better do some of that. It's also in the business of telling us how (it's right) to be motivated. And just as different first-order normative views will differ in what they claim is right, different first-order normative views will differ in how they think it's right to be motivated to do it. For instance, Kant not only tells us that we must do this or that action (our duty), but also that we must in doing it be motivated by the thought that it is what we must do (our duty). And he argues separately for each claim. Neo-Aristotelians such as Foot not only tell us that we must do what's characteristic of flourishing members of our species, but also that we must in doing so be motivated (it's difficult, but roughly) not because it is so characteristic. Self-effacing consequentialists tell us that we must do what realizes the best possible practical outcome, but that we must in doing so *not* be motivated by any such thoughts. And so on. Complete first-order normative views provide accounts of both matters.

Hence in the present context, which is one of arguing against a particular first-order normative theory, it's illicit to work with a version of the special status view that itself assumes a particular answer to the question 'How ought agents to be motivated?'. But that is exactly what Markovits and Arpaly do in making their case against self-effacing consequentialism. Once we see that Markovits and Arpaly's special status view's ideas regarding the normative importance of match between an agent's motivating reasons and the right reasons, i.e. the *reasons it's correct to be motivated by*, can be divorced from a substantive view about what the contents of the reasons it's correct to be motivated by are (e.g. that they're the same as the content of the *right-making* reasons), self-effacing consequentialists can happily endorse the – properly purified – version of the special status view without any inconsistency. And they can then go on to plug in their *own* view about what the right reasons, the reasons it's correct to be motivated by, are.

So, that is how I propose self-effacing consequentialists evade the dilemma posed to them by proponents of the special status view such as Markovits and Arpaly. Markovits and Arpaly assume, illicitly, that correct motivation is identical to motivation by *right-making* reasons (rather than simply *right* reasons). That is why they think, mistakenly, that the structure of self-effacing consequentialists' views precludes them from having any account of the relevant phenomenon. But once we drive a wedge between these two ideas – between the reasons it's correct to be motivated by in doing what one does and the reasons it's correct to do what one does – an amended special status view – the special status view *Light* as I'll call it – is available to such theories. So, while there's an important debate to be had about whether and to what extent we should embrace self-effacing first-order normative views, that debate cannot be smuggled in as a debate over whether such views can account for moral worth.<sup>19</sup> In the next section, I'll consider some objections to this suggestion.

Before moving to objections, however, let me quickly sketch an argument to the effect that even proponents of the original special status view would be better off with the *Light* version just outlined. Here is how that argument goes.

Proponents of the special status view are forced to admit there are sometimes right actions that *cannot* be performed worthily because of the possibility that the normative reasons are somehow blocked from being part of the acting agent's motivating reasons. Markovits, for one, acknowledges this point, saying that if the special status view

<sup>19</sup> Thanks to an anonymous referee for suggesting this way of putting things.

is true, then ‘there could be some right acts – acts we’d be obligated to perform – which we could not perform worthily’.<sup>20</sup> And, as Markovits is happy to admit, this fact ‘follows from the [special status view] regardless of which normative ethical theory is the right one’.<sup>21</sup> Indeed, as Philip Stratton-Lake points out, it’s possible to imagine a world where, if the special status view were true, it would be *systematically* impossible for agents to act so that their actions have moral worth:

Consider a world in which there is an omnipotent, evil demon whose aim is to stop good people from doing what they should in the light of the normative reasons why they should so act. . . . He achieves this by making it the case that if a good person ever acts from the normative reasons why she should so act, he will make it such that this action is wrong, and he tells them this. Every good person knows, therefore, that she cannot do the right thing from the normative reasons why this is right. For they know that if they are motivated to act in this way, then their actions will be morally wrong.<sup>22</sup>

This possibility looks an awful lot like the possibility facing self-effacing consequentialists, viz. the possibility that agents ought not to act for their normative reasons. Why isn’t this supposed to be just as troubling for proponents of the special status view as it was for self-effacing views? The answer, says Markovits, is that such examples are ‘pretty artificial’ and that ‘there will not be many real-life situations in which we ought to act unworthily’.<sup>23</sup> And it’s ‘no objection to a normative theory if it entailed that we might, in rare circumstances, be obligated to act unworthily’.<sup>24</sup> Fair enough: it’s no objection to a normative theory that it entails this, given that the special status view is correct.

But here’s the thing: it surely *is* a cost of one’s account of how worth is engendered by match – that is, the special status view – if it entails that we might, even if only in rare circumstances, be obligated to act unworthily when there’s an *alternative* account of worth available – the special status view *Light* – that does not entail this. In other words, it’s surely better if our theory of match and worth allows that agents *can* always act both *rightly* and *worthily*, and so, given that they always can, that they always should. But we can only say this, that agents are always obligated to act both *rightly* and *worthily*, if we allow for the right reasons to come apart from normative reasons, as the *Light* view does. We cannot say this if we insist, as the unamended special status

<sup>20</sup> Markovits, ‘Acting’, p. 235.

<sup>21</sup> Markovits, ‘Acting’, p. 235.

<sup>22</sup> Stratton-Lake, *Kant*, p. 18.

<sup>23</sup> Markovits, ‘Acting’, p. 235.

<sup>24</sup> Markovits, ‘Acting’, p. 235.

view does, that the right reasons are always everywhere identical to normative reasons.

Here's another way to put the point. Proponents of the special status view such as Markovits are forced to say that agents sometimes ought to act unworthily. Perhaps that's not a cost of accepting the particular first-order normative theory those proponents accept since, as Stratton-Lake's example shows us, the mere possibility that this will happen is something that follows from the special status view regardless of what first-order normative theory we accept. But, again, it *is* a cost of the special status view *itself*, at least if there's an alternative view available that doesn't have to pay it. And there is such an alternative. If proponents of the special status view such as Markovits were instead to accept the special status view Light, they could say the following: in ordinary circumstances the right reasons are in fact identical in content to normative reasons, and so agents ought to act on the basis of their normative reasons for acting (and so ought to act worthily). However, in exceptional circumstances, the content of the right reasons comes apart from the content of normative reasons – that is, perhaps, what makes such circumstances exceptional. In such circumstances, it's still true that agents ought to act worthily. What's different is that, in these exceptional circumstances, agents ought to act on the basis of reasons – the right reasons – that are *not* identical in content to normative reasons.

So, not only should self-effacing consequentialists adopt the special status view Light to resolve their dilemma, it seems as if proponents of the unamended special status view themselves have reason to move towards the Light version of the view. I'll now consider four objections to this proposal.

## V. OBJECTIONS AND REPLIES

### V.1. *The right reasons are necessarily identical to normative reasons*

*Objection:* The right reasons are necessarily identical to normative reasons. Since the present suggestion denies this is so, it must be mistaken.

*Reply:* The right reasons are the reasons it's correct to be motivated by in doing what it is correct to do. The normative reasons are the reasons it's correct to do what it is correct to do. These are manifestly not the same concept, though depending on the content we give to each they may share an extension. But their sharing an extension depends, as we've seen, on what first-order normative theory we think is most compelling. Moreover, the two uncontroversially *do* come apart in some first-order domains.



For instance, the two regularly come apart in the domain of instrumental normativity. Consider the content of a normative instrumental reason for performing some action  $\varphi$ . A normative instrumental reason for  $\varphi$ -ing might plausibly be that by  $\varphi$ -ing one promotes one's interests, whatever these are. But, familiarly, being motivated by the fact that some action promotes one's interests is often counterproductive, i.e. it can interfere with the successful achievement of those interests. Instrumental rationality therefore sometimes enjoins one to be motivated by concerns other than achieving one's interests, despite achieving one's interests (plausibly) being the aim of instrumentally rational action. So, in acting so as to achieve one's interests, the reasons it's instrumentally correct to do what one does, i.e. that it will promote one's interests, will sometimes come apart from reasons it's correct to be motivated by.

This is sufficient to show that the right reasons for action can not only be divorced *in principle* from normative reasons for action, they are regularly separated *as a matter of fact*. The two are not only not necessarily identical, they are not even actually so.

V.2. *The right moral reasons are necessarily identical to normative moral reasons*

*Objection:* Granted, it's possible for the right reasons to come apart from the normative reasons in certain normative domains, e.g. the domain of instrumental normativity. In those domains, an action might have e.g. instrumental worth, in virtue of match between right (but not normative) reasons and an agent's motivating reasons. But in certain domains, and in particular in the domain of *moral* normativity, the two are always one: the right *moral* reasons are necessarily identical to normative *moral* reasons. Since the present suggestion denies this is so, it must be mistaken.

*Reply:* The objection lacks teeth unless the objector can answer the following question: why should it be that, in the moral domain specifically, the reasons it's morally correct to be motivated by in doing what one does, i.e. the right moral reasons, must be identical to reasons it's morally correct to do what one does, i.e. to normative moral reasons? Unfortunately, I see no way for the objector to answer this question in a way that avoids begging the question.

Recall that this idea, that the right moral reasons are always identical to normative moral reasons, is being used in the context of an argument against a first-order moral theory that holds, as part of its theory, that the two are *not* always identical. Hence here we cannot simply assume an alternative first-order moral theory, e.g. Kantianism, that entails the two *do* come hand in hand, notice this fact, and then

leverage this fact into an argument against the target first-order moral theory.

Perhaps the idea is just that right reasons very strongly intuitively *are* always identical to normative reasons in the moral domain, and that it's more costly to give up that strong intuition – as we would have to do were we to accept both self-effacing consequentialism and my suggestion on behalf of that view about how to solve our problem – than it is to give up the self-effacing consequentialist's first-order moral theory.

I'm not sure how this line of thought is supposed to make any progress. Self-effacing consequentialists will no doubt differ in how strongly intuitive they think it is that the right reasons are necessarily identical to normative reasons in the moral domain. Indeed, there's a long history of self-effacing consequentialists and consequentialists with what we might call *self-effacing sympathies* reporting the contrary intuition.<sup>25</sup> So without some further argument, in particular one that doesn't rely on assuming any particular first-order moral theory, which again is question-begging in the present context, I don't see why we have any reason to accept that, necessarily, in the moral domain the right reasons are always identical to normative reasons. Moreover, as we saw at the end of [section IV](#), there's some positive reason even proponents of the special status view have for allowing that the two can come apart, viz. that doing so allows them to say that agents are *always* obligated to act both rightly and worthily.

### V.3. *We shouldn't be neutral between competing first-order moral theories in solving the right reason problem*

*Objection:* The reply just given to the previous objection assumes we should be neutral between competing first-order normative theories when thinking about the normative significance of the phenomenon of match. In particular, it assumes we should be neutral between competing *moral* theories. But we *shouldn't* be neutral between such theories.

*Reply:* Yes, we *should* be neutral between competing first-order normative theories, and in particular between competing moral theories, in attempting to account for the normative significance of match, at least if we *can*. Importantly, the normative significance of match, though it's obviously a philosophers' problem, is grounded in a commonplace, everyday phenomenon, viz. the phenomenon whereby agents sometimes do the right thing *for the right reason* and the fact

<sup>25</sup> Most famously, perhaps, J. S. Mill, *Utilitarianism*, 2nd edn. (Indianapolis, 2002), ch. 2.

that, moreover, when this happens it seems to bear some normative significance for us. The phenomenon is inarguably widespread in our ordinary normative thought and talk.<sup>26</sup> Moreover, the phenomenon is acknowledged by ordinary people who hold a *wide range* of competing first-order moral theories. If we failed to remain neutral between competing first-order moral theories, as the objection urges, then, to borrow a point from Markovits, we would have to embrace an error theory about much of our ordinary moral discourse.<sup>27</sup>

This is because, as we saw above, it's certainly possible to hold a first-order moral theory according to which the right reasons are not always identical to normative reasons. Self-effacing consequentialists are a case in point. But there are other examples too. If we fail to remain neutral between such first-order moral theories and theories such as Kantianism that hold that the right reasons are always identical to normative reasons in our theorizing about why match matters, then we risk ascribing a problematic kind of massive error to ordinary folk.

The point is not that there's any problem ascribing massive systematic error to the folk in general: people can obviously be systematically wrong about what the correct moral theory is. But we *should* avoid ascribing *two* massive systematic errors to the folk when one will do. Here, we're being told not just that the folk are in systematic error regarding which moral theory is true. (That's an error that, at least for present purposes, I'm happy to say people make.) We're also being told that the folk are in systematic error regarding the pattern of their judgements about what doing the right thing for the right reason comprises. For manifestly some of those folk think that it's possible to do the right thing for the right reason without doing the right thing for the normative reasons. (Manifestly, since some people *really are* self-effacing consequentialists. And despite holding such (possibly mistaken) moral theories these people *really do* engage in talk of doing the right thing for the right reason.) According to the present suggestion, those people are not just mistaken in their pattern of judgements about what the normative reasons comprise (i.e. they're wrong about the correct moral theory), they're mistaken in their understanding of what doing the right thing for the right reasons comprises (i.e. they're wrong about what the phenomenon is they care about). But, again, we shouldn't ascribe that error unless we must. Since it's not necessary to ascribe this error in order to solve the problem – that is what the availability of the special status view Light reveals – we shouldn't do so.

<sup>26</sup> Google searches on the phrases 'for the [right/wrong] reason' turn up almost a million results in sum.

<sup>27</sup> Compare Markovits, 'Acting', p. 236.

V.4. *Non-identity of right and normative reasons undermines the normative significance of match*

*Objection:* In order to make itself available to first-order moral theories such as self-effacing consequentialism, the special status view Light allows that the right reasons for action can come apart from the normative reasons favouring that action. But in allowing this, it vitiates the explanatory power of the original special status view. This is because, if the right reasons are not identical to normative reasons, then it's mysterious why match between the right reasons and an agent's motivating reasons is normatively significant in the way outlined in the special status view: namely, by reflecting on the agent's character and (hence) meriting praise. If we adopt the special status view Light, then, contrary to appearances, we haven't solved our original problem since we've failed to explain the normative significance of cases of match.

*Reply:* It's best to think about this objection as attempting to saddle the special status view Light with an extra explanatory burden it can't bear. The special status view Light owes us some explanation of why, if the right reasons are not identical to normative reasons, match between the right reasons and an agent's motivating reasons is normatively significant. Two points in reply.

First, the special status view Light can point out that match between an agent's motivating reasons and the right reasons (whether or not these comprise the normative reasons) is normatively significant precisely because the right reasons are the reasons *it's normatively correct to be motivated by*. In other words, it's not at all surprising that we care about an agent's motivating reasons matching the right reasons (even if the right reasons are not identical to normative reasons). After all, if we care about comporting ourselves as the first-order normative theory we accept says we *ought* comport ourselves, then we'll *ipso facto* care about being motivated in the way that theory says it's correct to be motivated, i.e. we'll care about being motivated by what that theory says are the right reasons for action.

Second, whether or not the right reasons are identical to normative reasons, being motivated by the right reasons *does* reflect on an agent's character and *does* thereby merit a kind of praise. After all, the fact that an agent is in some case (not) motivated in the way the first-order normative theory we accept says it's correct to be motivated is presumably a function of that agent's character, and this is true *regardless* of whether this is a matter of the agent's (not) being motivated by the normative reasons. Why would the content of the

correct motivation make any difference to the source or nature of that motivation? If an agent's motivation flows from and so reflects on her character, then it does so however we think that motivation ought to be. So, cases of match (or failure to match) will still reflect on an agent's character. And it's also true that an agent's comporting herself in her motivational profile in the way the correct first-order normative theory says it's correct to do is something that, *ceteris paribus*, will merit praise: she is, in that respect at least, doing – or, more accurately: being motivated – as she is supposed to. In other words, the idea that the normative significance of match is that in a case of match the action reflects on an agent's character and is thereby due praise doesn't seem to depend on any view about the content of the right reasons, and certainly doesn't depend on the view that the right reasons are always identical to normative reasons. Instead, such normative significance seems to derive from the fact that, in cases of match, agents are being motivated *in the way they ought*. And this is true just as much for someone who adopts the special status view Light as for someone who adopts the unamended special status view.

## VI. SUMMARY

Agents sometimes act rightly, and sometimes they act rightly for the right reason. We face a puzzle of explaining what the structure of this phenomenon is like, and what its normative significance is. The orthodox special status view says that when agents act for the right reason, their actions are thereby *worthy*. Proponents of the special status view have argued that certain forms of consequentialism, in particular versions of the view that are self-effacing, cannot adopt the special status view, and that the structure of their view prevents them from offering any alternative account. Hence these views are supposed to face a dilemma: either fail to account for the normative significance of match, or give up their consequentialist ambitions.

The solution for consequentialists is to adopt the special status view Light. The crucial move is to notice that, in making the original charge, the special status view assumes that the *right* reasons for action are always necessarily identical to *normative* reasons for action. This is what causes the trouble for self-effacing views. But this assumption is both illicit in the present context and, in any case, not compulsory. The special status view Light that rejects this assumption is therefore perfectly compatible with self-effacing consequentialism's first-order normative theory, and such consequentialists have no difficulty – at least, no *special* difficulty – solving our original problem. Moreover, as I've argued, we might *all* have reason to prefer the special status view

Light in so far as we are interested in allowing that it's always possible for agents to do the right thing for the right reason.<sup>28</sup>

[natesharadin@gmail.com](mailto:natesharadin@gmail.com)

<sup>28</sup> Thanks to Ben Bradley, Patrick Connolly, Luke Elson, John Lawless, Daniel Layman, Hille Paakkunainen and Dave Sobel for helpful feedback on earlier drafts of this article. Thanks also to two anonymous referees for their thoughtful comments.