Croatian Journal of Philosophy
Vol. XIV, No. 40, 2014

Consequentializing Moral Responsibility

FRIDERIK KLAMPFER
Faculty of Arts,
University of Maribor

In the paper, I try to cast some doubt on traditional attempts to define, or explicate, moral responsibility in terms of deserved praise and blame. Desert-based accounts of moral responsibility, though no doubt more faithful to our ordinary notion of moral responsibility, tend to run into trouble in the face of challenges posed by a deterministic picture of the world on the one hand and the impact of moral luck on human action on the other. Besides, grounding responsibility in desert seems to support ascriptions of pathological blame to agents trapped in moral dilemmas as well as of excess blame in cases of joint action. Desert is also notoriously difficult, if not impossible, to determine (at least with sufficient precision). And finally, though not least important, recent empirical research on people's responsibility judgments reveals our common-sense notion of responsibility to be hopelessly confused and easily manipulated.

So it may be time to rethink our inherited theory and practice of moral responsibility. Our theoretical and practical needs may be better served by a less intractable, more forward-looking notion of responsibility. The aim of the paper is to contrast the predominant, desert-based accounts of moral responsibility with their rather unpopular rival, the consequence-based accounts, and then show that the latter deserve more consideration than usually granted by their opponents. In the course of doing so, I assess, and ultimately reject, a number of objections that have been raised against consequentialist accounts of moral responsibility: that it (i) doesn't do justice to our common-sense theory and practice of responsibility; (ii) ties responsibility too closely to influenceability, thereby exposing itself to the charge of counter-intuitivity; (iii) assigns undeserved responsibility (praise, blame) to agents; (iv) confuses 'being responsible' with 'holding responsible', and (v) provides the wrong-kind-of-reason for praise and blame. My negative and positive case may not add up to a knockdown argument in favour of revising our ordinary notion of responsibility. As long as the considerations adduced succeed in presenting the consequentialist alternative as a serious contender to a pre-arranged marriage between moral responsibility and desert, however, I'm happy to rest my case.

Keywords: Moral responsibility, desert, blame, (reasons for) reactive-attitudes, consequentialism.

1. *Introduction*

Moral responsibility is one of our core moral notions. It figures prominently in our concept of moral agency as well as in some influential accounts of moral wrongness and justice (mainly retributive, but also distributive). Still, in the last century or so, the idea that we are (at least sometimes) genuinely responsible—in the sense of being praise- or blameworthy—for what we think, feel and do has increasingly come into disrepute. My aim in the paper is not to defend the notion of moral responsibility against either traditional or more recent objections. My ambition is more modest. I try to cast some doubt over traditional attempts to define, or explicate, moral responsibility in terms of deserved praise and blame. Desert-based accounts of moral responsibility, though no doubt more faithful to our common-sense/ordinary notion of moral responsibility, tend to run into trouble in the face of challenges posed by a deterministic picture of the world on the one hand and the impact of luck on human action on the other. Besides, grounding responsibility in desert seems to support ascriptions of pathological blame to agents trapped in moral dilemmas as well as of excess blame in cases of joint action. Desert is also notoriously difficult, if not impossible, to determine (with sufficient precision). And finally, though not least important, recent empirical research on people's responsibility judgments reveals our common-sense notion of responsibility to be hopelessly confused and easily manipulated.

So it may be time to rethink our inherited theory and practice of moral responsibility. Our theoretical and practical needs may be better served by a less intractable, more forward-looking notion of responsibility. The aim of the paper is to contrast the predominant, desert-based accounts of moral responsibility with their rather unpopular rival, the consequence-based accounts, and then show that the latter deserve more attention than usually granted by their opponents. In the course of doing so, I consider, and ultimately reject, a number of objections that have been raised against consequentialist accounts of moral responsibility: that (i) it doesn't do justice to our common-sense theory and practice of responsibility; (ii) ties responsibility too closely to influenceability, thereby exposing itself to the charge of counter-intuitivity; (iii) assigns undeserved responsibility (praise, blame) to agents; (iv) confuses 'being responsible' with 'holding responsible', and (v) provides the wrong-kind-of-reason for praise and blame. My negative and positive case may not add up to a knockdown argument in favour of revising our ordinary notion of responsibility. As long as the considerations adduced succeed in presenting the consequentialist alternative as a serious contender to a pre-arranged marriage between moral responsibility and desert, however, I'm happy to rest my case.

Here is the plan of my paper. In section 1, I elucidate the core notion of moral responsibility, the so-called R(eactive)-responsibility. In section 2, I discuss the strengths and weaknesses of the rival, desert-based account of R-responsibility (DMR). Section 3 introduces a consequentialist alternative (CMR) and claims certain comparative advantages for it. In sections 4 and 5, I defend the consequentialist account of R-responsibility against some common objections and misconceptions. In the course of doing so, I sketch an (rather minimalist) account of the nature and function of blame. In conclusion, I summarize and qualify my view.

But let me begin with some stage-setting.

2. *A multitude of meanings of 'responsibility'*

Ascriptions of moral responsibility serve a variety of purposes in a host of different contexts. Sometimes, when trying to apportion responsibility, we are really after authorship ("whodunnit?"). At other times our goal is to determine the appropriate sanction or reward; or to establish potential duties of reparation or restitution; or even just to make the agent feel bad about himself for having done something we dislike or resent. To make things worse, ascriptions of responsibility sometimes serve as a ground for statements about someone's obligations to others. More often, though, responsibility ascriptions ground claims about someone's accountability or answerability for her behaviour.

This variety of purposes and meanings of "responsibility" is reflected in the rich vocabulary we employ when we try to allocate moral responsibility in the broad sense of the word: we hold agents accountable, blameable, culpable, liable, appraisable, reprehensible, deserving of praise or reproach… for something they have done or failed to do. A close examination of ordinary language use reveals that the word "responsibility" refers to a number of closely related, yet nevertheless different ideas.[1]

2.1. *The notion of R(eactive-)responsibility*

The first task, then, is to isolate, or disentangle, from this interconnected heap, a core notion of moral responsibility. In the paper, I will be interested in the notion of responsibility that has been labelled blame-, remedial-, liability- or reactive-responsibility. To be morally responsible for some action (or state of affairs) in this, R(eactive) sense, is to be

[1] Faced with these variety of meanings of the word 'responsible', and a rich array of ideas that find expression in it, different authors distinguish between as little as two (Miller 2007) and as many as six (Hart 1967/2007, Vincent 2011) different meanings of "responsibility": (i) virtue-responsibility, (ii) role-responsibility, (iii) outcome-responsibility, (iv) causal-responsibility, (v) capacity-responsibility and (vi) blame and remedial/liability-responsibility. I myself have found it useful to distinguish between four different meanings of 'responsibility' in the past (Klampfer 2004: 152–154). What Hart and Vincent call virtue- and role-responsibility I prefer to characterize in either descriptive ("Smith is a morally reliable person, i.e. the kind of person you can rely upon for doing the right thing.") or deontic terms ("In his capacity as the captain of the ship, Smith ought to have protected people's lives and possessions better than he in fact did.").

an appropriate object of a particular kind of reaction—praise, blame, or something akin to these—for having performed that action (or having brought that state of affairs about). And to hold someone R-responsible is to single that person/agent out as the (only, most) appropriate object of such reactive attitudes.[2]

Throughout the paper, I will employ the following notion of blame- or reactive-responsibility (in short, R-responsibility), unless indicated otherwise:

MR: to say that A is morally (R-)responsible for X is to say that it is appropriate (fitting, correct, fair, and the like) to single A out as the object of (positive or negative) reactive attitudes, primarily praise and blame, in virtue of the fact that A did X (or some other fact(s) about either A or X).

Three brief points. First, note that the notion of 'appropriateness' or 'fittingness' of reactive attitudes used here—which is the main source of MR's intuitive appeal—is deliberately vague, so that it can be dished out in different ways, of which the one in terms of fairness or desert is not better simply by default; hence, the consequentialist proposal to understand these terms as substitutes for 'usefulness' or 'efficiency' is not to be dismissed off-hand. Secondly, even though MR is pluralist in the sense that it allows a variety of reactive attitudes to play the defining role, I will, for the sake of simplicity, concentrate on only one such attitude, blame (and its positive correlate, praise). Thirdly, the basis of, or the reason for, a reactive attitude, the thing that accounts for it being an appropriate, or fitting, or correct response, is some fact about the object of the attitude, i.e. the agent. Which fact(s) about agents can serve this function, the definition should leave open. This once again makes room for a substantive account of when the attitude of blame is appropriate or fitting rather than the desert-based one. A consequentialist account, for example, which I prefer, thus meets this formal requirement, since it nominates the expected impact of blame on the agent's future behaviour for this role, clearly a fact about the agent.

So here is the adjusted version of the notion of moral responsibility that I will work with in the paper:

MR*: to say that A is morally (R-)responsible for (having done) X (in C) is to say that it is appropriate (fitting, correct, fair, and the like) to single A out as the object of blame (or praise), in virtue of the fact that X was the wrong (right, admirable) thing to do (for A in C).

2.2. *Blame and punishment*

Before proceeding, let me briefly address another important issue. Blame is often conceived of as a younger sibling of punishment, as a milder, in-

[2] Thus, in focusing on the reactive-type of responsibility, I follow Bruce Waller (and many others): "*...Moral responsibility provides the moral justification for singling an individual out for condemnation or commendation, praise or blame, reward or punishment.*" (Waller 2011: 2, my emphasis)

formal type of sanction. Consequently, it often receives the same kind of treatment, in the sense that whatever conditions are said to make blame (and praise) appropriate, they are also said to render punishment (and reward) appropriate. I disagree. I believe the two should be treated separately, and this for a host of reasons. One is the following:

> To say that an agent is morally responsible (for an act, omission or attitude) is to say that the Strawsonian reactive attitudes are justified in relation to her with regard to that act, omission or attitude. That is, it is appropriate for observers to have certain attitudes in relation to her and her act, especially the attitudes, partly cognitive and partly constituted by emotion, of praise and blame. It is a *further* question whether it would be appropriate to punish or reward the agent for her act, or even whether it would be appropriate to *express* the judgment. It may be that the expression of the reactive attitudes is justified under stronger, or merely different, conditions than those under which it is appropriate merely to have them, and it is with the latter that we are here exclusively concerned. …For instance, the (quasi-) expression of attitudes the *having* of which is not justified might be justified on consequentialist grounds, as indeed might be the imposition of sanctions. But it does not follow, from the putative fact that treating an agent *as if* she were responsible is justified, that the agent *is* responsible. (Levy 2005)

Hence, we might have consequentialist reasons for *expressing* attitudes (such as blame) that we are not justified in *having*. So from the mere fact that it would be appropriate to hold someone responsible (i.e. blame her), it doesn't follow that she is responsible (i.e. blameworthy). The same seems to be true of punishment—we may have good consequentialist reasons for punishing someone despite having no good reason to believe that they are responsible (blameworthy, deserving of punishment). By the same token, then, from the mere fact that it wouldn't be appropriate to punish someone, it doesn't follow that she isn't morally responsible.

This is but one disanalogy between blaming and punishing, or praising and rewarding. There are many others that speak in favour of separate treatment. Blame has an irreducible psychological dimension and it is often warranted when punishment isn't. Furthermore, if the agent's behaviour is beyond influence, or irresponsive to blame, then perhaps we ought not to blame her; whereas if it is irresponsive to punishment, or beyond repair by legal means, we ought to choose a different form of punishment, one that secures the accomplishment at least of detention, the subordinate goal of punishment.

3. *Desert-based accounts of moral responsibility (DMRs)*

The most common and popular way of spelling out conditions of R-responsibility is in terms of deserved or fair blame.

(DMR) An agent, A, is morally responsible for X, if and only if she is praise-/blameworthy for it, i.e. deserving of praise/blame in virtue of having done, or brought about, X (freely, knowingly, intentionally, as a reflection of one of your stable character traits, and so on).

Is DMR conceptually true? There is a natural temptation to treat DMR as the, to borrow Vargas' term (Vargas 2011), correct 'diagnostic' account of MR, i.e. as a description of the content and use/conditions of application of our common sense notion of moral responsibility.[3] But I want to resist this temptation and urge the reader to do the same. Ascriptions of moral (R-)responsibility are best understood as claims about certain reactions being justified, warranted and so on. This, however, doesn't by itself commit us to a further, more substantive claim that the only proper justification for reactive attitudes is one cast in terms of fairness or justice.

Despite its undeniable appeal, DMR is wrought with problems. Here is a tentative, but by no means exhaustive list: (i) strong metaphysical commitments; (ii) epistemological obstacles; (iii) methodological quibbles; (iv) pathological (self-)blame in moral dilemmas; (v) excess blame in cases of moral bad luck; and (vi) indeterminacy of personal desert in joint action.

These issues are neither all equally pressing nor all equally damaging. (i), (ii) and (vi) all arise because of the ineliminable desert element in DMRs. (iii), on the other hand, plagues only those DMRs that define the 'fittingness-relationship' in terms of 'fairness' or deservedness of reactive attitudes. And, finally, (iv) and (v) may turn out to be more or less accidental, affecting only very specific versions of DMR. How serious are, then, these problems? Let me briefly address each in turn.

3.1. *Strong metaphysical commitments*

Metaphysical concerns about the basis of moral desert are not just the most basic, they are also by far the most serious. The argument for the groundlessness of (moral) desert is pretty straightforward—since in our deterministic world the necessary conditions for (moral) desert can in principle never be met, there is no such thing as desert.

(1) For someone to deserve something on some basis, B, he or she needs to be responsible for B.

(2) Responsibility presupposes genuine (libertarian) free will.

(3) There is no (libertarian) free will in our deterministic world.

Hence,

(4) no one is really responsible for anything.

But then (from 1 and 4),

[3] Recall Waller's hasty conclusion: „...whatever the conditions required for moral responsibility, *it is meeting those conditions that makes punishment (and reward, blame, and praise) fair and just.*" (Waller 2011: 2, my emphasis) Waller advances this as a purely formal account of moral responsibility, which can then be supplemented by various substantial accounts, i.e. sets of conditions that need to be met for someone to be morally responsible in this sense. My worry is that he has already built in substantive assumptions in his purely formal definition and thereby restricted, for no good theoretical reason, the range of plausible substantial accounts.

(5) no one *really* deserves anything, including praise and blame (and hence no one is *really* blameworthy or praiseworthy for his or her attitudes and actions).

Now, of course, the above argument is only straightforward at the cost of its accuracy. What do we mean by 'genuine, or libertarian, free will'? Do agents really need to be free in the sense of possessing such a capacity in order to be responsible, and is it even true that such freedom of will cannot be found in a causally deterministic world? These are notoriously difficult questions that have been pondered by generations of the finest philosophical minds. (With little success, if I may add.) Libertarians have, of course, flatly denied the truth of causal determinism. Compatibilists, on the other hand, have chosen to reject either premise (3), or, more recently, premise (2). Harry Frankfurt lifted their spirits, when he famously argued that you can be morally responsible for your action, even if you couldn't have acted otherwise. Finally, determinists come in two variations: the more pessimistic ones (Pereboom 2001) tend to write obituaries to our ordinary notions of responsibility, guilt, blame(worthiness) and even wrongdoing, while the more optimistic (Smilansky 2000) try to curtail the damage done to our ordinary moral thought and action by determinism. Not even premise (1) with its undeniable intuitive appeal has been spared criticism.[4]

I cannot enter these and related metaphysical disputes. Neither do I want to suggest that the scores in this game are settled once and for all. After all, powerful arguments have been advanced in support of the claim that the (non-)existence of free-will is largely irrelevant to moral responsibility (Strawson 1962, Frankfurt 1967). But the prospects of the so-called soft-compatibilism for securing our traditional notions of free agents and moral agency are still unclear and until this is so, determinism will continue to pose a threat to our ordinary notion and practice of holding people morally responsible for what they do and omit.

3.2. *Epistemological obstacles*

It is an odd feature of an otherwise excellent systematic treatment of the topic in George Sher's classical book on desert (Sher 1987) that while he devotes enormous energy and space to uncovering justificatory grounds of various types of desert-claims, not a single chapter addresses the thorny issues of the epistemology of desert. Desert may have its own, independent normative force and making sure people get what they deserve (either because this is dictated by some plausible moral principle—the principle of gratitude or respect for persons as moral agents—or simply because it is a valuable state of affairs) may even come close to a self-evident moral truth, but can we ever deter-

[4] See Feldman (1999) for a number of ingenious counter-examples to the idea that moral desert presupposes moral responsibility.

mine what punishment, reward, opportunity, success, good or bad luck, prize, pay, compensation, and so on anyone deserves?

Doubts about the possibility of accessing the desert-bases come in two variations. According to the epistemological argument against desert, the influence of natural and social factors on people's actions and traits undermines *some* desert-claims (or all desert-claims *to a certain extent*). But since we do not know which ones (or to what extent), and since we cannot measure people's deserts, we cannot reward or punish them (Moriarty 2005). The argument begins with a characterization of human agency. People's actions, traits and achievements—their potential desert-bases—are the product of two forces. The first is what Sidgwick calls 'gifts of nature' and 'favouring circumstances' (what we would nowadays classify under the label 'good luck'). These are the native abilities and social circumstances that vary so much from person to person. The second force is free choice. The argument next says that determining what people deserve requires prying these two forces apart, for, it assumes, people can be deserving only in virtue of that part of their achievement which is the product of their own free choices, not in virtue of that part which is the product of natural gifts and favouring circumstances. But, it continues, prying these forces apart is 'impossible in practice', 'impracticable', or 'intractable'.

So here is the epistemological argument presented in a more formal way.

(1) Conscious effort, rather than achievement (which is to a large extent influenced by luck), is the only ground of desert.

(2) Conscious effort, however, results partly from the agent's free will and partly from her undeserving natural and social endowments.

(3) Rewarding desert requires separating the contributions that these two/three factors make to the agent's conscious effort.

But since

(4) this cannot be done (with sufficient precision and/or confidence),

(5) claims of desert can never be sufficiently justified.

But then, on the plausible assumption that

(6) unjustified propositions cannot provide grounds for other propositions,

it follows that

(7) desert claims cannot provide grounds for ascriptions of responsibility.

Considerations other than these also speak in favour of moderate scepticism about desert in general, and moral desert in particular. As Norvin Richards (Richards 1993) points out, given our limited epistemic access to the bases of desert, what a person deserves for a particular deed can, and often will, differ considerably from the criticism we are actually entitled to level against her for doing it. An agent, for example, can be much more culpable than anyone has grounds to realize, and in that case no one is entitled to criticize her as harshly as she deserves. Conversely, we may only have a very restricted epistemic access to either the exculpating or mitigating circumstances. So there will often be an indeterminately big gap between the agent's degree of responsibility and the degree of blame that we are entitled to ascribe to her, and this will provide one example of how the issue of responsibility and the practice of blaming can come apart. Things get even more complicated, if desert is, as some have argued (Hurka 2011), essentially holistic in nature, i.e. if it depends on a host of other facts in addition to those that concern the agent himself.

And as if that were not bad enough, when we turn to factors that typically shape people's judgments of desert, we find even less room for (self-)confidence:

> When we say that a person deserves a positive or negative outcome, we are making a judgment that is influenced by a number of variables. We might be influenced by the person's own positive or negative characteristics, by our knowledge of what kinds of groups or social categories the person belonged to, and by whether we like or dislike the person. Information about these different variables has to be considered and integrated in some way, and our judgment of deservingness follows that psychological process, a process that involves the cognitive-affective system. (Feather 2002)

In making desert-judgments, people seem to pay close attention to things, such as the value of outcome, the value of the action that brought this outcome about, whether the agent was responsible for it and to what extent, her perceived characteristics even her group membership and whether we like her or not, which should make very little or no moral difference at all.

3.3. *Methodological concerns*

Contemporary philosophical theories of moral responsibility attempt to develop universal criteria for fair assignments of blame and praise. With that aim in sight, they crucially rely on appeals to shared intuitions about key principles and cases to justify these criteria. Consequently, theories of moral responsibility must make empirical assumptions about the universality or convergence of the intuitions upon which their theories rely. But are these assumptions warranted? Not necessarily. Empirical evidence suggests that there are fundamental intuitive differences regarding the conditions for fair assignments of moral responsibility, differences sufficiently deep and well-motivated to make it implausible that reflection, concept disambiguation, dialogue, and agreement about non-moral facts could resolve them. But if so, then we have no principled means of establishing the truth of 'universalist' theories of moral responsibility and we are left with 'metaskepticism about moral responsibility' (Sommers 2012).

The methodological argument, then, takes the following form:

(1) Conditions of moral responsibility are conditions of fair assignment of praise and blame.

(2) In identifying conditions of fair praise and blame, we must rely heavily on intuitions.

But since

(3) there is not much convergence in these intuitions (across cultures), or at least not enough to identify a single, unified set of criteria,

(4) no theory of moral responsibility can claim universal validity.[5]

Consequentialist alternatives to DMR are, at least prima facie, immune to these sorts of concern, since they don't rely, in identifying the conditions of moral responsibility, on intuitions at all—let alone on intuitions about the fairness of praise and blame—but rather on empirically informed predictions about the impact of praise/blame on the agent's future performance instead.

3.4. *Addicted to blame?*

We share a deep psychological need for blaming (mostly) others and (only occasionally) ourselves (and vice versa with praising—we tend to praise ourselves excessively and others rather sparsely). Judgments of deserved blame and the corresponding reactive emotions that they validate may even be, as Strawson (1962) famously argued, an essential part of the fabric of our social world. And yet, psychologists have identified a tendency of exaggerated or pathological (self-)blame. People, they say, are stubborn moralists, inclined to blame other people for their actions ahead, and even in spite, of the evidence of the absence of intention and/or control, ascribe agency and goal-directed behaviour even to inanimate objects, and even readily accommodate judgments of causality and intentionality to suit their moral judgments (Pizarro & Helzer 2010). And philosophers are by no means immune to this, as can be seen in their readiness to blame dilemmatic choosers and victims of bad luck. Also, in contexts of joint or coordinated action, we tend to apportion excess blame (Goodin 1995). So, there is at least some philosophical need for trying to tame (our 'natural' need for) blaming agents for things they do, or reduce blame to its right proportions.

[5] Angela Smith (Smith 2007), in her critique of attempts to define moral responsibility in terms of when it would be fair to blame the agent from an outside perspective of an unbiased, emotionally detached judge, combines the two perspectives. Our intuitive judgments about whether and when it would be fair to blame the agent for her actions, she argues, are sensitive to factors that are too contextual to admit of a uniform account, and too irrelevant to be of real epistemic merit.

3.4.1. *Pathological (self-)blame in genuine moral dilemmas*

According to a fairly popular view, an agent in a genuine moral dilemma is required to perform or refrain from performing both of two incompatible actions and since she will inevitably have done wrong, she is right to judge herself blameworthy, either for failing to perform the foregone action or for performing the prohibited action she did choose.

The key feature of the moral psychology of the agent who takes herself to be in a moral dilemma is that she will judge herself to have done something wrong no matter how she chooses. It is because this is how things look to such agents that the 'remainder thesis' is said to apply to them. The agent who has chosen in a dilemmatic situation is said to be subject to a moral remainder or residue: the moral force of the required but foregone option or the prohibited choice remain, and ought to do so, to haunt the conscience. The agent has failed morally, and ought accordingly to be blamed (i.e. blame himself) and to feel profoundly guilty. The agent comes to possess an objective moral taint after her choice.

Or so the popular story goes. The logically problematic move is, of course, from 'the agent (cannot help but) believe(s) that she has done wrong' to 'the agent cannot but see herself as blameworthy'. We should resist this move. There are at least two ways of blocking this inference. One could block it by denying that the dilemmatic choosers have done any wrong at all. Those among them who agonize over their choice are, on this view, simply mistaken or confused. But there is another strategy that I find no less promising. It is to reject the implicit equation of wrongness with blameworthiness. So even if what's at issue is not the objective facts of the matter (did the agent do something wrong or not?), but rather the agent's perception of the situation (can she escape believing that she did something wrong or not?), I want to question the suggestion that insofar as she cannot help but see herself as a wrongdoer, she must also see herself as blameworthy. It is fairly clear to me that you can consistently believe to have done some wrong and nevertheless deny deserving blame for that wrong.

As Byron Williston has convincingly shown (Williston 2006), dilemmatic choosers may not help but see themselves as wrongdoers. Still, they both can and should divorce this judgement from an ascription of self-blame. It is possible, as Williston correctly observes, to blame an agent robustly for what he chooses, even if the choice takes place in a dilemmatic situation, as long as the values guiding his choice are themselves morally bad. However, the sort of agent that the literature on moral dilemmas is overwhelmingly concerned with is the one whose core values are all morally sound, or at least not obviously corrupt. If an agent's character is both strong and good, then so far as her actions are the product of that character, she will act blamelessly. Thus dilemmatic choosers are morally sui generis in that although their actions involve a diminishment of personal integrity, their characters can still be described as both strong and (at least moderately) good. This characterization allows us to insert fully the wedge between wrongdoing and blame, and to treat dilemmatic choosers as blameless wrongdoers. Not only is the notion of blameless wrongdoing as such perfectly consistent,[6] moral dilemmas are its most natural environment.

3.4.2. *Excess (self-)blame in cases of moral bad luck*

Our differential intuitive judgments about the cases in which agents are beneficiaries of moral good luck, or victims of moral bad luck, suggest that we tacitly accept that:

(i) In order to deserve praise or blame for something (i.e. be a proper object of reactive attitudes, such as blame, reproach, contempt, and so on), the agent must be responsible for it; ('responsibility constraint on desert'); and that

(ii) you can only be responsible for those things that are within your control (things you can influence, affect, make a difference to); ('control condition on responsibility')

and yet,

(iii) we often blame (as well as praise) people for things over which they clearly have no or very little control (their character traits, the choices they make, what they do or fail to do and the outcomes of their actions). ('pathological blame')

Now, even though the tendency to blame victims of bad moral luck (and, to a much lesser extent, to praise beneficiaries of good moral luck) is neither confined to, nor rooted in DMRs, the two look very much like natural allies. It is, after all, the conflation of two sorts of agent evaluations that facilitates ascriptions of excess blame: (a) evaluations concerning the agent's moral record (or performance), and (b) evaluations concerning the agent's moral worth (or virtue). The former are judgments about what the agent did (what he can be credited for as the author), the latter judgments about what sort of person he is (virtuous or vicious, better or worse than someone else). It is a common mistake to understand the agent's worth as closely related to (or even entirely determined by) the agent's record. But this is clearly false, given that the agent's record depends on what the agent actually does, whereas his worth (virtue or vice) depends on what he is disposed to do (Greco 2006). We tend to blame victims of bad luck, when we do, it seems, because we falsely assume that if A did some wrong (if he, as in a well-known example from the literature, swerved his car onto the pavement and killed a pedestrian) that B didn't do, then A must be a worse person (for that reason), and hence deserving of more blame, than B, even if it is true that had B been less fortunate, he would have done the same. And such an inference would surely appeal more to

[6] This is not a universally accepted view, of course. For a dissenting voice, see Mason 2002.

those who equate wrongness with blameworthiness and blameworthiness with deserved blame than those who insist on keeping action- and agent-judgments apart.

3.4.3. *Excess aggregate blame and/or indeterminable individual desert in cases of joint action*

One of the tasks of philosophical accounts of R-responsibility, such as the DMR, is to reliably guide the distribution of R-responsibility, or blame, among several agents (wrongdoers). This purpose, however, seems to be rather poorly served by DMRs in those cases of joint actions where the undesirable outcome is either under- or over-determined (Goodin 1995). In the former, the actions/omissions of individual agents are either not individually necessary or jointly sufficient to produce the outcome. In the latter, the actions/omissions of individual agents are either all necessary or each individually sufficient to produce, or prevent, the undesirable outcome. How, then, should blame be distributed in such cases? Intuitively, individual responsibilities should sum to less than one in cases of under-determination (where, for example, both drivers are 25 per cent responsible for a car crash, their liabilities shouldn't add to one) and to more than one in cases of over-determination (where one person poisoned and other shot the victim, they are both fully liable for murder, making the sum of liabilities bigger than one). DMR, on the other hand, requires that the degree of each individual's blame (liability) strictly reflect his or her individual contribution to, or responsibility for, the outcome. Accordingly, whenever individual responsibilities for the outcome sum up to more than one, so will the liabilities that we will need to distribute among contributing individual agents.

Why, if DMRs fare so badly, not follow Waller's advice and simply abolish the reactive type of moral responsibility altogether? Why not say, when the shortcomings of desert-based accounts become so apparent, that we should *never (again)* hold *anyone* morally responsible? Well, for one, holding people R-responsible is part of the fabric of our social world and it serves an important social function. Also, most of the problems with DMR have more to do with the presumed link between R-responsibility and desert than with R-responsibility or appropriateness of blame as such. So why not try to salvage this useful notion before we dispense with it once and for all?

4. *Consequence-based accounts of moral responsibility (CMRs)*

So, if we are to preserve the concept and practice of moral responsibility, we should sever the link between responsibility and desert and put responsibility on a firmer footing. And what better ground is there than the impact of holding people responsible on their performance? Which brings us to our next candidate, consequentialist accounts of moral (R-) responsibility.[7]

(CMR) An agent, A, is morally responsible for X just in case praising or blaming, rewarding or punishing her for X would produce good consequences by improving her (own as well as other people‘s) future moral performance.[8]

On the face of it, CMR has many attractive features. It promises to avoid (i) the threat of determinism—since it insulates moral responsibility against the threat of determinism (admittedly, at the cost of severing its tie to freedom of choice and action); (ii) the threat of moral luck—since it insulates moral responsibility against the vagaries of luck (admittedly, at the cost of dissociating it from its natural ally, moral desert); (iii) epistemological obstacles—since effects of praise/blame on people's future are in principle epistemicaly easier accessible than the bases of desert; (iv) methodological worries—since identifying conditions of moral responsibility no longer depends on convergent intuitions about the fairness of praise/blame; (v) ascriptions of pathological (self-)blame to agents in moral dilemmas—since no amount of praise and blame can possibly improve agents' behaviour in future moral dilemmas; (vi) ascriptions of excess individual blame in joint action—since it doesn't insist on degrees of individual blame closely mirroring degrees of individual contributions.

On the other hand, CMR provokes even more criticisms. It is said to (vii) fail to do justice to our common-sense theory and practice of responsibility; (viii) tie responsibility too closely to influenceability, thereby exposing itself to the charge of counter-intuitivity; (ix) assign undeserved responsibility (praise, blame) to agents; (x) confuse 'being responsible' with 'holding responsible'; and (xi) provide the wrong kind of reason for ascriptions of responsibility. In the remainder of the paper, I will focus on CMR's putative weaknesses, rather than its advantages. My aim is to demonstrate that weaknesses are exaggerated and objections misfired.

[7] CMR has been revived by J.J.C. Smart (Smart 1961) and recently defended, in a somewhat revised form, by Richard Arneson (Arneson 2003) and Manuel Vargas (Vargas 2006). But the same idea can already be found in Moritz Schlick (Schlick 1932) and Henry Sidgwick's *Methods of Ethics*: "From a Utilitarian point of view, as has been before said, we must mean by calling a quality, 'deserving of praise', that it is expedient to praise it, with a view to its future production: accordingly, in distributing our praise of human qualities, on utilitarian principles, we have to consider primarily not the usefulness of the quality, but the usefulness of the praise" (Sidgwick 1962: 428).

[8] More precisely, a revised, modified Smart account that Richard Arneson defends looks like this: "To say that an agent is responsible for an act she has done is to say that she is accountable, that is, a fit object of praise or blame, reward or punishment, *depending on its quality*. The condition that renders an individual a fit object of praise and blame and so on for what she has done is influenceability. And an agent is influenceable with respect to what she has done if imposition on her of praise or blame and so on for doing it would improve the future by affecting the likelihood that the agent will act in a similar way in the future" (Arneson 2003; my emphasis).

4.1. *Incompatibility with our ordinary notion of R-responsibility*

Let's start with the unfamiliarity objection. According to a common complaint, CMR is difficult, if not impossible, to square with our ordinary, common-sense notion and practice of moral (R-)responsibility. Our common-sense notion is, critics of CMR insist, meritocratic in that it requires either that blame be well-deserved or that it passes the fairness test, if it is to be determinative of responsibility. CMR, on the other hand, allows for the possibility that a person, A, is responsible for X, even though it would be either unfair to blame A, or A would not deserve to be blamed, for X.

But should we really try to accommodate our ordinary notion of R-responsibility?[9] Perhaps not. There is now ample empirical evidence that our ordinary notion of moral responsibility is simply too inconsistent and confused for any philosophical account to possibly do justice to it. Ordinary people randomly switch criteria when attributing moral responsibility; there is no one, identifiable set of criteria that they would predictably apply to all cases; their judgment is highly context-sensitive; and their judgments seem to be influenced by features of situations that bear little or no relevance to the issue of agent's R-responsibility, such as whether the situation is described in very abstract or more concrete terms, whether the agent under consideration is a close friend or a complete stranger, whether her behaviour is morally good or morally bad, whether harm caused (or the risk of harm imposed) by her action is serious or relatively trivial, and so on.[10] A further mark of the ordinary notion of R-responsibility is a stark asymmetry between praise and blame. It seems that (a) the fact that an outcome was merely a foreseen side-effect reduces the responsibility attributed to agents for morally good behaviours but not for morally bad ones, (b) the fact that a behaviour was the product of an overwhelming emotion reduces the responsibility attributed for morally bad behaviours, but not also for the morally good ones, and (c) the fact that the agent intended to do (but eventually didn't do) something increases the degree of his responsibility for morally bad, but not also for morally good intentions. And that's

[9] Note that this is but one of the many interrelated features of our ordinary notion and that we should judge any account of R-responsibility by how many features of our common practice of praise and blame it can accommodate. So, for instance, both the DMR and the CMR would have to explain, in addition to that, why we excuse agents for underperforming, when we do, i.e. come up with a plausible account of exculpating and mitigating circumstances as well. Another important feature of our ordinary concept and practice of blame is grading—we blame some agents more than others. Yet another is its distinct temporal dimension, or what Miranda Fricker (Fricker 2010) following Bernard Williams, calls 'the relativism of blame'. In contrast to judgments of wrongdoing, which seem to stand the test of time, judgments of blame tend to be sensitive to the passage of time, in the sense that the more remote in the past certain wrongdoing lies, the less we are inclined to blame the agent for it.

[10] For a full list, see Knobe & Doris 2010.

not even the end of bad news. Judgments that are supposed to precede, and ground, judgments of moral responsibility, aren't immune to this moralizing virus either—so much that people's causal judgments ("Did the agent cause this thing or not?") and judgments of intent ("Did the agent bring such and such about intentionally or not?") also crucially depend on a host of features that no serious theory of causality or intentional action would consider relevant.

Knobe & Doris prefer to characterize our ordinary notion of moral responsibility as 'variantist' and 'contextualist', rather than simply 'confused'. But it is not so much variantism that we should find disturbing about our ordinary concept of moral responsibility, as the arbitrary nature of the conditions of its application. I have no qualms about the fact that ordinary people consider certain features of situations morally salient, i.e. such that they either establish agent's moral responsibility, or increase or reduce its degree, and others not. This fact alone doesn't make them suspect. It is the morally arbitrary character of these features that disqualifies them as reliable indicators of moral responsibility. So even if we grant that variantism per se does not compromise our ordinary notion of R-responsibility, people are simply mistaken in their judgment of moral responsibility and hence the latter cannot be relied upon in our philosophical inquiry.[11]

4.2. *Responsibility is not the same as influenceability*

CMR seems to equate responsibility with influenceability. But aren't these two rather different ideas? Thomas Scanlon is equally puzzled, it seems, by this kind of proposal, when he states, "The usefulness of administering praise or blame depends on too many factors other than the nature of the act in question for there ever to be a good fit between the idea of influenceability and the idea of responsibility *which we now employ*" (Scanlon 2008; my emphasis).

Here is another complaint of a similar kind:

> The deepest oddity about Pereboom's world (i.e. the world of hard determinism), however, lies… in the fact that the only problems wrongdoing appears to present to its inhabitants are future oriented. That, at any rate, is the

[11] As an alternative to dismissing the ordinary notion of moral responsibility as hopelessly confused and theoretically worthless, one could perhaps argue that at least some empirical findings are quite consistent with, if not even supportive of, a consequentialist approach to responsibility. The praise-blame asymmetry, for example, lends itself well to a consequentialist justification, since there are obviously good consequentialist reasons for giving priority to the elimination of moral defects over the strengthening of moral excellences. An alternative, non-consequentialist account of the praise-blame asymmetry is developed in Hindricks (2008). Hindricks accepts the suggestion implicit in the ordinary notion of intentional action, that the concept of intentional action is normative, not descriptive, and then offers his own, normative account of intentional action, according to which an agent does something intentionally, if she fails to be motivated to avoid the bad effect. I find the idea of the normativity of the concept of intentional action difficult to fathom, but cannot afford to open another front here.

> clear implication of the three responses to wrongdoing—admonish, ignore, walk away—that Pereboom is willing to countenance; for all three recommend themselves primarily as methods of preserving our future tranquillity. *This exclusively future-oriented stance toward wrongdoing, reminiscent of some of what Strawson says about the objective attitude, is bound to seem profoundly strange to anyone to whom the primary significance of wrongdoing lies not in what it augurs but simply in what it is. To such a person—that is, to all of us in our philosophically unguarded moments—nothing less than actual blame will do.* (Sher 2006: 6; my emphasis)

So what is it that some critics find so disturbing about the idea of responsibility as influenceability? If Scanlon's and Sher's misgivings are representative of this kind of worry, then CMRs, or any similar revisionist proposal, will, by turning our attention away from the relation between the agent and her past action to the presumed future effects of reacting to that action in a particular way, inevitably fail to engage with wrongdoing per se (as such, in itself). R-responsibility is about, or should reside in, the agent's contribution to the wrong-making features of her action which makes her liable to a particular sort of moral criticism. By subscribing to CMR, however, we should expect our moral outlook, or at least the part of it that governs our reactions to wrongdoing, to undergo substantial transformation, in that wrongdoing will come to be seen as, to borrow Kurt Baier's (Baier 2003) words, a symptom of a system-failure to be repaired.

Severing the tie between the agent and his past action renders CMR vulnerable to counter-examples. Here are some implications of the account that most people will no doubt find awkward, if not plainly absurd: (a) If you can't improve someone's future behaviour by blaming her (for her past behaviour), then you ought not to do it; (b) of the two people who did pretty much the same thing for pretty much the same reason and with pretty much the same consequences, we ought to blame the first one a lot more (since he is more resistant to incentives for a change of attitudes and/or behaviour) than the second one (who is much more responsive); correspondingly, (c) some wrongdoers will get away with no or comparatively little (self-)blame, and this seems unjust and/or unfair. Also, (d) it wouldn't make sense to blame wrongdoers as long as they are already dead.

So how bad is this news? I'm willing to bite this bullet. After all, we seem to tolerate relatively big differences in legal sentences administered to people who are found guilty of the same or similar offence. Also, when it comes to one-time offenders, the fact that there is no need to correct their future behaviour (at least not in this one particular respect) speaks in favour of a mild, reduced punishment. The proposed account, then, is no more of a challenge to logic, reason and imagination than current provisions of law which we readily accept and seldom question.

4.3. *Undeserved blame?*

The previous objection could also be understood differently, as a complaint over the undeservingness of blame—some people will get the blame that they don't deserve and others will get away with no blame, even though they would clearly deserve it. In addition, many people are going to end up getting more blame than they deserve, while others will get less. Recall: when two people do pretty much the same thing for the same reason and with similar consequences, CMR may imply that we blame one a lot more than the other, as long as the first one proves to be more resistant to incentives for a change of attitudes and/or behaviour and the second one is fairly responsive. Correspondingly, some wrongdoers will get away with no or comparatively little (self-)blame, and this strikes most people as plainly unjust and/or unfair.

How serious is this flaw? Not particularly, I guess, and this for the following reasons. (a) Desert might be, for all we know, baseless (in which case there won't be any such thing as deserved or undeserved blame anyway). (b) At any rate, desert is indeterminable (and so it is practically impossible to assess whether and how much blame the agent deserves for his or her wrongdoing). (c) And even if some judgments of moral desert are warranted, considerations of desert could still be, and probably often are, overrated.

The first two points I've made before and need no repetition: all the necessary conditions for either moral or reactive desert are not fulfilled in our deterministic world; and even if we could put (claims of) moral desert on a solid metaphysical footing, desert could never be determined with sufficient precision to warrant any judgment of moral desert. Let me therefore briefly elaborate on the third. We usually take considerations of desert and merit rather seriously. Charges of undeserved reward or punishment, praise or blame cannot be easily dismissed or sidestepped. But is such a high degree of normative priority really warranted? That will depend on what we believe desert-claims amount to. McMahan (2005) and Boonin (2008), for example, propose an analysis of moral desert in terms of intrinsic betterness. In their opinion, "A deserves D" boils down to something like "The world in which A gets D is intrinsically better than the world in which A doesn't get D". Which, it seems, would rank undeserved blame among relatively minor moral offence(s)—on the assumption that blaming B for X would be unfair, the world in which A holds B responsible, would be intrinsically worse (by how much? to what degree?) than the world in which this isn't the case. Now, on the assumption that the world, in which people perform better, morally speaking, is intrinsically better than the world in which their moral records are comparatively worse, the outcome of an 'all-things-considered' moral equation won't be easy to predict. Of course, the above account of moral desert in axiological terms can be, and is, disputed. When interpreted in deontic terms ("A deserves D" only if "prima facie, it ought to be the case that A gets D", or "prima facie, it would be wrong if A didn't get D"), desert carries much more normative weight and cannot be lightly dismissed or exchanged for just about any other value. Still, this is only part of the whole picture. For even on this assumption, while it may be true that if someone deserves X, then prima facie she ought to get X, it is not equally evident that if someone doesn't deserve Y, then prima facie she ought not to receive Y (particularly if her getting Y will not deprive of Y someone else who is more deserving of Y than her); or at least it will be relatively easy to advance other considerations in favour of providing Y to her that may override, or trump, the (lack of) desert consideration.[12]

4.4. *Being responsible vs. holding responsible*

CMR is not so much an account of when someone is responsible for something, the critic might object, as an account of when it is appropriate to hold her responsible regardless of whether she is responsible or not. In other words, what consequentialism can plausibly provide is an account of when, or under what conditions, we are justified in *holding* someone responsible (or, to use Levy's words, in 'expressing our reactive attitude of blame'), whereas what we are primarily interested in is when we are justified in believing that someone is responsible (or, to use Levy's words, in 'having those attitudes').

It is one thing to prove someone responsible (i.e. the appropriate object of reactive attitudes) and quite another to show that it is appropriate to hold her responsible, and yet still another to justify punishing her. Hence, we might have good consequentialist reasons for *expressing* or communicating to others attitudes (such as blame) that we are not justified in *having*. So, from the mere fact that it would be appropriate to hold someone responsible (i.e. blame her), it doesn't follow that she is responsible (i.e. blameworthy). The same goes for punishment—we may have good consequentialist reasons for punishing someone who we have no good reason for believing is responsible (culpable, guilty, or deserving of punishment). By the same token, then, from the mere fact that it would be appropriate to punish someone, it doesn't follow that she is culpable or that she deserves to be so punished.

This presents a serious challenge to the proposed solution, because it threatens to restrict the scope of CMR and open the door for a hybrid account, one that combines a desert-based account of being responsible with a consequence-based account of holding responsible.[13] As long as

[12] For instance, demands of equality or equal treatment. Is it fair to hold someone, A, responsible for something, X, if X was not in A's control? Well, given that desert presupposes agent control and that by assumption A did not exert control over X, it cannot be fair in the sense of being deserved. It may, however, be fair in another sense, namely from the point of view of the principle of equality or equal treatment, as long as the alternative, namely not holding anyone responsible, would have been even more inegalitarian. For an argument along these lines, see Stemplowska 2008.

[13] Reminiscent, perhaps of the hybrid accounts of the morality of punishment, where the retributivist idea that in order to be liable to legal punishment, one has to

the former can be plausibly assigned some sort of lexical priority (in the sense that holding responsible is parasitic on being responsible, or that it can only be appropriate or fair to hold those responsible who are in fact responsible, or that we can justify public expression of only those feelings that we are justified in having), the prospects of CRM for finishing the race ahead of DMR suddenly begin to look bleak.

The objection has a certain prima facie appeal. Note, however, that it doesn't target specifically CMRs. Certain desert-based accounts will be equally hard-pressed to answering it. For if conditions for someone being responsible differ from conditions for (justifiably) holding her responsible, an account of the latter will be only approximately true of the former. Furthermore, it seems perfectly legitimate to ask why *hold* anyone responsible unless that person really *is* responsible. However, it may turn out, upon reflection, that in the actual world no one is really morally responsible for anything (either because there can be no free agency in a causally deterministic world and we need to be free in order to be responsible, or because given the prevalence of luck nothing we do or affect in this world is properly under our control and control is necessary for responsibility) and hence no responsibility ascription is literally true. In this case, we may still want to preserve the practice of holding people responsible because of its many psychological and societal benefits. For example, if, as a matter of empirical fact, blaming and praising people for who they are and what they do actually improved their moral performance (or strengthened their compliance with moral norms),[14] that would give us a good pragmatic, or practical, or even moral reason for maintaining this practice despite the fact that it lacked any ontological footing.

But we don't even need to pull the determinism card to block the above objection. All we need to establish is that while it is undoubtedly true that being responsible makes you the primary candidate for being held responsible, this connection is defeasible. We can see this by considering an example discussed by David Miller. Miller draws a similar distinction as the one above, namely between identifying and assigning responsibility:

> In the case of both of these notions of responsibility, we can distinguish between identifying responsibility and assigning it. Identifying responsibility is a matter of looking to see who, if anybody, meets the relevant conditions for being responsible. What these conditions are will depend on the form of responsibility at issue. In the present case, for the teacher to identify Johnny as outcome responsible for the messy classroom, she would at the very least have to establish certain matters of fact, such as whether he had been in the classroom during break. She could get this wrong, and judge Johnny responsible for the mess when in fact it was Katy who was respon-

be (proven) guilty of some criminal offense, is combined with the utilitarian insight that in determining the sentence one ought to be guided by its expected effects.

[14] Which, admittedly, is still subject to dispute. For some doubts, see Springer 2008.

> sible. Assigning responsibility, by contrast, involves a decision to attach certain costs or benefits to an agent, whether or not the relevant conditions are fulfilled. The teacher may lack any concrete evidence about Johnny, but because she harbours suspicions based perhaps on past incidents, and because she feels the need to pin responsibility on someone, she says to him, 'I'm holding you responsible for the state of this room; you'll be in big trouble if it happens again'.
>
> Or, in the absence of any information about which child was in fact responsible for the chaos, she might assign responsibility to the whole class and impose some form of collective punishment or liability. Unlike identifications, assignments of responsibility can be justified or unjustified, but they cannot be correct or incorrect.
>
> A parallel distinction can be drawn in the case of remedial responsibility. Remedial responsibilities can be identified where there are reasons for attaching them to one agent rather than another. In the classroom case, remedial responsibility would naturally fall on the children who were outcome responsible for the mess by virtue of having created it. …However, the teacher might also simply assign remedial responsibility, picking out one child at random or choosing a child she dislikes. Here she would naturally say 'I'm making you responsible for clearing up this room'. Again such an assignment might be justified or unjustified—it would be unjustified if the teacher kept picking on a particular pupil, for instance—but it could not be correct or incorrect in the way that an identification could be. (Miller 2007: 84–85)

So here is Miller's idea in a nutshell—identifying someone as (either outcome or remedially) responsible for something and assigning (either outcome- or remedial-) responsibility to someone are two distinct practices, distinguished by the respective goals they serve and the respective set of rules that govern them. Hence, even though we are most likely to assign responsibility for some action or state of affairs to the agent that we have identified as responsible for it, this need not be so (and Miller, in the above paragraph, describes a couple of situations in which we might be justified in pulling them apart). But if we can, at least in principle, justifiably assign responsibility to those who we cannot (correctly) identify as being responsible, then this opens room for the possibility of a complete detachment of assignments of responsibility from its identifications, one which would help to preserve the practice of assigning responsibility to agents in the absence of either required freedom of will or determinable personal desert. So the distinction, if sound, will be useful both to soft determinists and to advocates of consequentialist accounts of moral responsibility.

4.5. *A wrong kind of reason?*

The foregoing discussion has opened up another avenue of criticism for the foes of CMR. Their discontent with the kind of justification of either our attitude of blame or our practice of blaming agents for what they think, feel or do, typically provided by CMR, is reminiscent of another famous controversy in contemporary philosophy, the dispute between proponents and opponents of the so-called fitting-attitude analysis of value (or FAVs, for short). FAVs try to give an (reductive) account of axiological properties in terms of deontic properties. More precisely, they define (different kinds of) value, or goodness, in terms of normative reasons for (various kinds of) pro-attitudes towards the carrier/ locus of value.

Here is the classical FAV scheme:

FAV: X is valuable/good = def X has properties which make it a fitting object of certain pro-attitudes, such as favouring, desiring, admiring, and the like.

There is a striking structural similarity between FAVs and CRMs. Just as FAVs try to give an account of value in terms of fitting pro-attitudes, CRMs (as well as DMRs) offer an account of moral (R-)responsibility in terms of a fitting attitude of blame. Let's call this group of views of MR the fitting-attitude analyses of moral responsibility, or FAMRs for short. FAMRs, then, offer an analysis of the concept of moral responsibility in terms of whether having and/or expressing certain reactive attitudes, foremost blame, is an appropriate (fitting, warranted, justified, and the like) response to the agent's wrongdoing.

The following scheme is shared by all such views:

FAMR: An agent, A, is morally (R-)responsible (i.e. praise- or blameworthy) for X, just in case, and whenever A is a fitting object of (a particular subset of) reactive attitudes, such as blame, resentment, and the like.

And if we replace the notion of 'fittingness', 'appropriateness' or 'what is called for', with that of a 'there being a reason for', we can formulate the above view in the language of reasons:

FAMR*: Agent A is morally (R-)responsible for X, just in case, and whenever there is a (sufficiently good) reason for blaming A for X.

In their influential paper 'The Strike of the Demon: On Fitting Proattitudes and Value' Wlodek Rabinowicz and Toni Rønnow-Rasmussen present a forceful objection against FAV. They call it the 'Wrong Kind of Reason' (or the WKR for short) problem for FAV.[15] Here is how it goes. FAVs define value or goodness as a second-order property of having properties which make it an appropriate, fitting object of pro-attitude. Such an analysis, however, is too generous. It seems to commit the advocates of FAV to saying that if there are reasons for favouring X, then X is good/valuable. The problem is that we may have good reasons for adopting a pro-attitude towards something that is clearly worthless (just as we may have reasons not to have pro-attitudes to objects that are obviously valuable). For instance, if the Evil Demon demanded that we admire a saucer of mud or else he would punish us, this would pro-

[15] More precisely, their point is that there is no non-circular, informative way of drawing the line between those reasons for various pro-attitudes towards an object that are constitutive, or indicative, of its value, and those that aren't.

vide us with a good reason for admiring a saucer of mud,[16] even though the object of our admiration is clearly worthless. Hence, the inference from "there is a good reason for a pro-attitude to X" to "X is good/valuable" must be invalid. Examples like this compel proponents of the FAV to draw a distinction between the right and the wrong kinds of reasons for pro-attitudes—those that are and those that aren't determinative/indicative of value. That Evil Demon is going to punish me unless I admire a saucer of mud, may give me a good reason for admiring a saucer of mud, they say, but it is simply the wrong kind of reason.

The trouble with this otherwise logical move is that it is not easy to identify a principled ground for telling these two categories of reasons apart. What makes a certain fact the right kind of reason and another one the wrong kind of reason for a certain positive or negative attitude to a particular object, as opposed to simply *different* kinds of reasons? Here are some proposals for how to conceive of this difference: it is supposed to coincide with the distinction between a) object- and attitude-given reasons; (b) reasons for first- and second-order attitudes (favouring something vs. bringing it about that I favour something); (c) reasons that play a dual role (i.e. properties that both justify the attitude and are represented in its intentional content as those properties for the sake of which I adopted that attitude) and those that don't; and (d) reasons why the attitude is correct vs. reasons for adopting it.

So, does any of this help to solve the WKR conundrum? Hardly. First of all, the dividing line between these pairs of reasons is far from precise and clear-cut.[17] Secondly, and even more troublesome, none of these proposals succeed in insulating the FAV against all counter-examples—for it seems that certain facts are going to constitute reasons for pro-attitudes such that even though the latter manage to pass the right-kind-of-reason test, they will nevertheless favour adopting pro-attitudes towards clearly worthless objects. In other words, the above characterizations of the difference between the right and the wrong kinds of reasons for pro-attitudes all fail to distinguish cases of genuine value from cases of merely putative value. There will always be cases where we seem to have a reason of the first kind for favouring something, and yet that thing is clearly not good/valuable. But if FAVs fail to characterize reasons of the right kind, as opposed to reasons of the wrong kind, then it is difficult to see how the WKR-type of objection could be used to settle the dispute between the DMR and the CMR about the correct account of moral (R-)responsibility. For in order to do so, the friends of the former would have to explain why the fact that 'A deserves the blame for X' is the right kind of reason for blame, whereas the fact that 'blaming A will improve A's future moral performance'

[16] Which we will, if this is what the Evil demon demands from us on pain of punishment.

[17] Every state-, or attitude-given reason, for example, can be in principle formulated in terms of object-given reasons. See Rabinowicz & Ronnow-Rasmussen 2004.

isn't. If there is no principled way of telling, from the perspective of a FAV, 'good' cases of valuing from the 'bad' ones, why hope that we can find a principled, non-ad hoc way of identifying, within the class of appropriate, fitting blame, cases where blame is not only appropriate but where its propriety is determinative, constitutive or indicative of 'genuine' moral (R-)responsibility?

Perhaps such expectations are indeed naive. But dismissing them outright would be premature. The WKR-objection against CMR was initially inspired by the WKR-objection against FAV. CMRs offer an account of moral responsibility of A for X in terms of the beneficial impact that blaming A for X will have on A's future moral performance. But since, so the objection goes, the only good reason for blaming wrongdoers is that they are in fact blameworthy, i.e. deserving of blame, consequentialist analyses of moral responsibility in terms of influenceability provide us with the wrong kind of reason for blame. Consequently, CMRs will validate, as fitting, appropriate, or called for, the having, or expressing, of a Strawsonian type of reactive attitude towards people when such attitudes are clearly not fitting to have and/or express (i.e. when the agent is not in fact blameworthy, or R-responsible). Influenceability, or the agent's responsiveness to blame, may be, and often is, a perfectly legitimate reason for blame; it is just not the kind of reason that is relevant to *defining* the reactive-type of moral responsibility, in the sense that it should figure in a correct analysis of this concept.[18]

Two quick rejoinders are called for. First, on one reading, the above objection conflates the criteria for an accurate *diagnostic* account of our ordinary notion of responsibility with the criteria that we reasonably expect any plausible *critical* account will meet. It is probably true that moral desert is much more part of our ordinary, everyday notion of moral (R-)responsibility than influenceability. But since our ambition is not to give an accurate diagnostic account of B-responsibility, why feel bound, in attempting a critical analysis of this concept, by its currently predominant mode of employment? Furthermore, the WKR-objection already presupposes what it would have to prove first, namely that influenceability, unlike deservingness, is not the right kind of reason for adopting the attitude of blame towards wrongdoers. Such a verdict should not be passed lightly, however. Instead, and ideally, it should follow from a general account of what kinds of objects call for, and validate, what kinds of attitudes. In assessing the merits of the WKR-objection against CMR, we need to look more closely into the general topic of reasons that one might have for adopting a particular kind of attitude towards a particular kind of object. So we can no longer ignore the issue that we have postponed so far, of the nature and logic of blame.

[18] Just as many perfectly legitimate reasons for, say, desiring an object are irrelevant to its evaluation, and hence shouldn't play a role in the analysis of the concept of value.

4.5.1. *The nature of blame*

Three accounts of blame seem to prevail in philosophical literature:

(i) Blame as a negative character-evaluation/appraisal.
(ii) Blame as a (milder) form of punishment/sanction.
(iii) Blame as a re-evaluation of one's relation with the agent in the light of the (perceived) meaning of her action.

For the purpose of evaluating the force of the WKR-objection against CMR, I will sketch an account of praise and blame that closely parallels David Alm's account of admiration and disdain.[19] My account belongs to the first group, and characterizes blame as

(i) a form of negative (moral) appraisal
(ii) of an agent
(iii) for his substandard behaviour (i.e. wrongdoing in case behaviour falls short of moral standards)
(iv) that was due to some (corrigible) deficiency in agent's character, will or motivation.

Thus, I suggest to take (moral) blame to be (a) an attitude that we adopt and/or express towards (b) a wrongdoer for (c) his substandard moral performance (e) insofar it was due to some (fixable) defect of will or motivation and, I'd like to add, (f) in order to effect a change in her behaviour to the better.

This, of course, is more of a sketch than a fully-fledged, detailed account of what blame is. But since my aim is to assess the force of the WKR-objection against CMR, and I merely stipulate the above as a plausible enough account of what characterizes the attitude and practice of blame, it will have to do. Let me just fill in a couple of details. Blame, on my account, presupposes both that the agent underperformed (i.e. that his behaviour fell short of certain reasonable expectations) and that *he* is at fault for underperforming. As such, it is subject to the control condition—the agent could and should have done better than he did (if only he had tried harder);[20] and it draws this failure back

[19] See Alm 2007. With one important difference—Alm is interested in an account of when someone is *deserving of either admiration or disdain*. I, on the other hand, am offering (a sketch of) an account of when someone is an *appropriate or fitting* object of *praise and blame*.

[20] Admittedly, the control condition is somewhat controversial. If 'having been able to have done better, i.e. otherwise, at that particular occasion' threatens to reintroduce thorny issues of freedom and determinism that plagued DMRs, I would suggest to rephrase it in a metaphysically neutral way, i.e. in words that bring no such metaphysical commitments. For all we need to make sense of the blame in the above sense is the truth of the proposition 'he could have done better had he tried harder', whether or not it is also true that he could have tried harder on that particular occasion. Alternatively, we could weaken the control condition to something like 'the agent's ability to do better next time he tries', which clearly doesn't assign to the agent a mysterious power to change either the past or the laws of nature. There is an extensive literature on how to unpack the notion of 'ability to do otherwise' which attempts to answer Harry Frankfurt's (Frankfurt 1969) original

to some flaw or deficiency in the agent's character or will (rather than skill or talent, which may be beyond his control). Alm, following Grice, adds another condition of propriety of the attitude in question, namely 'maxim of relevance'—by criticizing the agent for not living up to a certain standard of evaluation, we imply that the person could have lived up to that standard. The reason for this is pragmatic—there is no point in criticizing people if they cannot do better anyway.[21] I concur.

The described account of the nature of blame is perfectly consistent with CMR, as far as I can see. The last-mentioned feature, its orientation towards the future, even speaks strongly in its favour. At the same time, however, it manages to accommodate most of the central features of our ordinary notion of blame: it conceives of blame as an attitude and practice that is directed at agents, typically triggered by a belief that their behaviour was in some respect substandard, and adopted and/or expressed with the aim of effecting a change, to the better, in their future behaviour. As such, it is oriented both backward and forward, balancing both elements without sacrificing either.

So when is it inappropriate to adopt the attitude of blame towards someone, according to the proposed account? Well, for one thing, when we have set the standards by which we judge him unreasonably high; also, when the basis of negative appraisal doesn't really apply to the agent, i.e. when he didn't underperform for reasons suggested or when his underperformance didn't really uncover any defects of will or motivation. And, finally, blame is inappropriate when criticism implicit in the evaluative attitude is pointless, i.e. cannot possibly improve the agent's future performance—when, for example, substandard performance was due to some incorrigible factor beyond the agent's control, such as stupidity, low intelligence, and the like. With this background, let's return to our initial puzzle: how damaging is the WKR-objection against the CMR? If this account of both the nature and the function of blame (i.e. of what blame is and what it does) is more or less correct, is the fact that blame will affect a change in the agent's behaviour the right or wrong kind of reason for blame? On which side of the divide does it fall? With respect to Parfit's method of drawing this distinction, it seems to be both an attitude- and an object-reason—the impact of blame is, first of all, a fact about the attitude of blame; however, how blame will affect the agent is at the same time also clearly a fact about the agent, and so object-given. How about other proposals? Is the

challenge. For some examples see van Inwagen 1999 and Šuster 2012.

[21] Reasons for this may not be entirely pragmatic. Whenever we blame someone we impute, among other things, some sort of defect to her (character or will). The exact nature of this defect can be left to a theory of blameworthiness to specify, but as Richard Brandt (Brandt 1992: 232) plausibly suggests, it will be some sort of "dispositional feature of the agent, an incapacity or deleterious tendency". If behaviour shows no such incapacity or objectionable tendency in the agent, it must be excused. By the same token, then, unless this sort of defect is corrigible, what's the point of blaming the agent for it?

agent's influencebility a reason why the attitude of blame is fitting, or rather a reason for holding such an attitude, for bringing it about that we adopt it? Is it that property of the object which makes the attitude correct, or a property of the situation which speaks in favour of adopting such an attitude? I find these questions puzzling, primarily because the aforementioned distinctions are so blurred.

Is there a better way, then, to solve this issue? We could, I guess, ask the following question instead: Is CMR vulnerable to the same sort of counter-examples as the FAV? Would it assign moral responsibility to someone who is clearly not morally responsible, the same way FAV identifies as valuable something that clearly lacks value?[22] I don't see that it does. True, it is a common complaint over CMR that as long as blaming the agent will improve his future moral performance, it doesn't really matter, on this account, whether he has done what we blame him for or even whether what he did was wrong at all. And surely an agent who is morally and causally innocent of X, clearly shouldn't turn out to be morally (R-)responsible for X on any plausible account of R-responsibility. Hence, if CMR endorsed blaming morally innocent agents as (at least occasionally) appropriate, it would assign moral responsibility for things to people who are clearly not (R-)responsible for those things. However, the impression that this is exactly what CMR does, is mistaken. To see this, we only need to recall what we said about the nature of blame and its cognitive or doxastic preconditions—namely that a belief about the agent's wrongdoing is either causally necessary for, or constitutive of, blame. Though you can in a sense blame someone without sincerely believing either that what he did was wrong or that he was at fault for doing the wrong thing, by uttering the words "It's all your fault!", thereby mimicking (outward expressions of) blame, you cannot genuinely blame someone unless you believe that he failed to do the right thing and ascribe his failure to some (corrigible) deficiency of his will, character or motivation. But if this is correct, then CMR condones neither pretence blame (behaving as if we blamed someone for something) nor imposing blame on the morally innocent, and is in fact not vulnerable to the sort of counter-examples that threaten to undermine the FAV.[23]

[22] Let me note my unease even with respect to the original case—is a saucer of mud really so clearly worthless, as assumed in the WKR-objection, if by admiring it I can escape severe punishment? Doesn't the bizarre situation render it at least instrumentally valuable?

[23] What about the kind of examples mentioned by Scanlon (Scanlon 2008: 125), where wrongness and blameworthiness seem to come apart and yet blame may be appropriate? For instance, when we blame someone for having done the right thing, but for a wrong reason? These, I suggest, are parasitic on the paradigmatic cases of character-flaw-revealing-wrongdoing, in that the intuition about the appropriateness of blame in such cases is fuelled by the intuition, if indeed one shares that intuition, that doing the right thing for a wrong reason is a piece of morally substandard, and insofar morally objectionable, behaviour.

5. *Conclusion*

In the paper, I compared the advantages and disadvantages of two rival accounts of moral (R-)responsibility, DMRs and CMRs. Even though the former are still considered the default option and the latter looked upon with suspicion, I have offered a much more balanced assessment. The advantages of the DMR are typically overestimated and their problems ignored. At the same time, the strengths of the CMR are downplayed and putative shortcomings overblown. Once both are seen in the correct proportions, CMR becomes a much more serious candidate for the correct account of moral responsibility. To establish this unambitious claim was the modest aim of my paper. I can now rest my case.

Literature

Alm, D. 2007. "On an apparent asymmetry in attitude desert." In: *Hommage a Wlodek. Philosophical Papers Dedicated to Wlodek Rabinowicz.* Accessed at: http://www.fil.lu.se/hommageawlodek/site/papper/AlmDavid.pdf

Arneson, R. 2003. "The Smart theory of moral responsibility and desert." In: Olsaretti, Serena (ed.). *Desert and Justice*. Oxford: Oxford University Press: 233–258.

Baier, K. 2003. *The Rational and the Moral Order. The Social Roots of Reason and Morality*. La Salle: Open Court Publishing.

Boonin, D. 2008. *The Problem of Punishment*. New York: Cambridge University Press.

Brandt, R. B. 1992. "A utilitarian theory of excuses." In: *Morality, Utilitarianism and Rights*. Cambridge: Cambridge University Press: 215–34.

Feather, N. T. 2002. *Values, Achievement, and Justice. Studies in the Psychology of Deservingness*. New York: Kluwer Academic Publishers.

Feinberg, J. 1970. "Justice and personal desert." In: *Doing and Deserving*. Princeton: Princeton University Press: 55–94.

Frankfurt, H. G. 1969. "The Principle of Alternate Possibilities." *The Journal of Philosophy* LXVI: 829–839. Reprinted in: Fischer, J. M. (ed.). *Moral Responsibility*. Ithaca: Cornell University Press, 1986: 143–152.

Fricker, M. 2010. "The Relativism of Blame and Williams's Relativism of Distance." *Aristotelian Society Supplementary Volume* 84: 151–177.

Goodin, R. 1995. "Distributing merit and blame." In: *Utilitarianism as a Public Philosophy*. Cambridge: Cambridge University Press: 88–99.

Greco, J. 2006. "Virtue, Luck and the Pyrrhonian Problematic". *Philosophical Studies* 130: 9–34.

Hart, H.L.A. 1968/2008. *Punishment and Responsibility. Essays in the Philosophy of Law*. 2nd Edition, Oxford: Oxford University Press.

Hindriks, F. 2008. "Intentional action and the praise-blame asymmetry." *Philosophical Quarterly* 58: 630–641.

Hurka, T. 2011. "Desert: individualistic and holistic." In: *Drawing Morals. Essays in Ethical Theory*. Oxford: Oxford University Press: 154–177.

Van Inwagen, P. 1999. "Moral responsibility, determinism and the ability to do otherwise." *The Journal of Ethics* 3: 341–350.

Klampfer, F. 2004. "Moral Responsibility for Unprevented Harm." *Acta Analytica* 33: 119–161.

Knobe, J. & Doris, J. M. 2010. "Responsibility". In: Doris, J. M. (ed.) *Moral Psychology Handbook*. New York: Oxford University Press: 321–354.

Mason, E. 2002. "Against blameless wrongdoing." *Ethical Theory and Moral Practice* 5: 287–303.

McMahan, J. 2005. "The basis of moral liability to defensive killing." *Philosophical Issues* 15: 386–405.

Mcnamara, C. 2011. "Holding others responsible." *Philosophical Studies* 152: 81–102.

Miller, D. 2007. *National Responsibility and Global Justice*. Oxford & New York: Oxford University Press.

Moriarty, J. 2005. "The epistemological argument against desert." *Utilitas* 19 (2): 205–221.

Levy, N. 2011. *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*. Oxford: Oxford University Press.

Pizzaro, D. A. & Helzer, E. G. 2010. "Stubborn moralism and freedom of the will." In: Baumeister, Roy et al. (eds.). *Free Will and Consciousness: How Might They Work?*. Oxford: Oxford University Press: 101–132.

Pereboom, D. 2001. *Living Without Free Will*. Cambridge: Cambridge University Press.

Rabinowicz, W. & Ronnow-Rasmussen, T. 2004. "The Strike of the Demon: On Fitting Pro-attitudes and Value." *Ethics* 114: 391–423.

Richards, N. 1993. "Luck and desert." In: Statman, D. (ed.). *Moral Luck*. Albany: State University of New York Press: 167–180.

Rosen, G. 2004. "Skepticism about Moral Responsibility." *Philosophical Perspectives* 18: 295–313.

Scanlon, T. 2008. *Moral Dimensions. Permissibility, Meaning, Blame*. Cambridge: The Belknap Press.

Schlick, M. 1939. *Problems of Ethics*. Tr. by David Rynin. New York: Prentice-Hall.

Sher, G. 1987. *Desert*. Princeton, New Jersey: Princeton University Press.

_______2006. *In Praise of Blame*. Oxford: Oxford University Press.

Sidgwick, H. 1962. *The Methods of Ethics*. 5thEdition, Indianapolis: Hackett Publishing.

Smart, J. J. C. 1961. "Free Will, Praise, and Blame." *Mind* 70: 291–306.

Smilansky, S. 2000. *Free Will and Illusion*. Oxford: Clarendon Press.

Smith, A. M. 2007. "On Being Responsible and Holding Responsible." *The Journal of Ethics* 11: 465–484.

Springer, E. 2008. "Moral feedback and motivation: revisiting the undermining effect." *Ethical Theory & Moral Practice* 11 (4): 407–423.

Stemplowska, Z. 2008. "Holding people responsible for what they do not control." *Politics, Philosophy and Economics* 7: 355–377.

Stocker, M. 2007. "Shame, Guilt, and Pathological Guilt." In: Thomas, A. (ed.). *Bernard Williams*. Cambridge: Cambridge University Press: 135–154.

Strawson, P. 1962. "Freedom and resentment", *Proceedings of the British Academy* 48: 1–25. Reprinted in: Fischer, J. M. & Ravizza, M. (eds.). 1993. *Perspectives on Moral Responsibility*. Ithaca and London: Cornell University Press: 45–66.

Šuster, D. 2012. “Lehrer and the consequence argument.” *Philosophical Studies* 161: 77–86.

Vargas, M. 2011. “Moral influence, moral responsibility.” In: Trakakis, N. & Cohen, D. (eds.). *Essays on Free Will and Moral Responsibility*. Cambridge Scholars Publishing: 91–123.

Vincent, N. A. 2011. “A structured taxonomy of responsibility concepts.” In: *Moral Responsibility. Beyond Free Will and Determinism*. Dordrecht: Springer: 15–36.

Wallace, J. R. 1994. *Responsibility and the Moral Sentiments*. Cambridge: Harvard University Press.

Waller, B. N. 2011. *Against Moral Responsibility*. Cambridge: The MIT Press.

Williston, B. 2006. “Blaming Agents in Moral Dilemmas.” *Ethical Theory and Moral Practice* 9: 563–576.