

A New Form of Agent-Based Virtue Ethics

Daniel Doviak

Accepted: 7 July 2010 / Published online: 23 July 2010
© Springer Science+Business Media B.V. 2010

Abstract In *Morals From Motives*, Michael Slote defends an agent-based theory of right action according to which right acts are those that express virtuous motives like benevolence or care. Critics have claimed that Slote’s view— and agent-based views more generally— cannot account for several basic tenets of commonsense morality. In particular, the critics maintain that agent-based theories: (i) violate the deontic axiom that “ought” implies “can”, (ii) cannot allow for a person’s doing the right thing for the wrong reason, and (iii) do not yield clear verdicts in a number of cases involving “conflicting motives” and “motivational over-determination”. In this paper I develop a new agent-based theory of right action designed to avoid the problems presented for Slote’s view. This view makes morally right action a matter of expressing an optimal balance of virtue over vice and commands agents in each situation to improve their degree of excellence to the greatest extent possible.

Keywords Agent-basing · Motives · Pluralism · Right action · Virtue

1 Introduction

According to agent-based virtue ethics, the moral status of an action is determined entirely by the aretaic properties the agent exhibits in its performance.¹ In *Morals From Motives*, Michael Slote presents and defends one prominent version of agent-basing.² On this view, the motives of benevolence and care are fundamentally (i.e., non-derivatively) admirable; hence their expression in action makes the agent excellent and her actions right. By contrast, the motives of malevolence or callousness are fundamentally deplorable; hence their expression in action makes the agent deficient and her actions wrong.

¹In Slote’s words, an agent-based theory of right action “treats the moral or ethical status of an action as entirely derivative from independent and *fundamental* ethical/aretaic facts (or claims) about the motives, dispositions, or inner life of moral individuals”, Slote (2001), p. 7.

²Slote (2001).

D. Doviak (✉)
Muhlenberg College, Allentown, PA, USA
e-mail: ddoviak@muhlenberg.edu

In recent years a number of commentators have claimed that Slote's theory of right action—and agent-based views more generally—are implausible, since they cannot account for several basic tenets of commonsense morality.³ In particular, the critics maintain that agent-based theories: (i) violate the deontic axiom that “ought” implies “can”,⁴ (ii) cannot allow for a person's doing the right thing for the wrong reason,⁵ and (iii) do not yield clear verdicts in a number of cases involving “conflicting motives” and “motivational over-determination”.⁶ Despite Slote's attempts to save his view from these objections, it is not clear that he has succeeded in this regard. In fact, a growing critical literature suggests that he has not. Given the intrinsic interest in agent-basing, it might be worthwhile to try to formulate a version of the theory that avoids these objections. The aim of this paper is to develop one such view.⁷ In developing this view, however, I will not attempt to defend agent-basing as a general approach to explaining the nature of right action,⁸ nor will I argue that my proposal is (on balance) the best account of what makes right acts right. My goal is much more modest. I intend only to show that one need not abandon agent-basing for some of the more popular reasons found in the critical literature.

Here is the plan of the paper. In Section 2, I provide a summary of Slote's agent-based theory of right action and present several cases designed to illustrate the above-mentioned problems. In Section 3, I present and explain a new form of agent-based virtue-ethics. Roughly, the view makes morally right action a matter of expressing an optimal balance of virtue over vice and commands agents in each situation to improve their degree of excellence to the greatest extent possible. In Section 4, I show how this theory avoids the problems presented in Section 2.

³ For general criticism of agent-basing, see Karen Stohr and Christopher Heath Wellman (2002), pp. 49–54. See also David Copp and David Sobel (2004) pp. 518–525, and pp. 547–551, and Ramon Das (2003), pp. 324–39.

⁴ See Thomas Hurka (2001), pp. 225–226. See also Daniel Jacobson (2002), pp. 53–67, and Copp and Sobel (2004), pp. 550–551. For a defense of the idea that virtue-ethicists in general need not accept “ought” implies “can”, see Stocker (1971), pp. 303–315.

⁵ See Scott Gelfand (2000), pp. 87–88; Jacobson (2002), pp. 58–62; Ramon Das (2003), pp. 324–30; and Michael Brady (2004), pp. 1–10.

⁶ See Julia Driver (1995), pp. 281–288, and Jacobson (2002), pp. 56–57.

⁷ Scott Gelfand (2000) pp. 85–94 and Liezel van Zyl (2005) pp. 273–288 and (2009a) pp. 50–69 have formulated “hypothetical” agent-based approaches to right action. They claim that their views avoid some of the main problems listed above (van Zyl (2009a) focuses her attention solely on the right action/wrong reason objection). I will not discuss either of these views in any detail here. I will only register my doubts about their plausibility. The reason I am suspicious of these views is that each seems to collapse into one that is equivalent to Hursthouse's (1999). On Hursthouse's view, an action is morally permitted if it is what a virtuous person would do in the agent's circumstances. This problem with this theory is that it is extensionally inadequate. There are situations no virtuous person would ever be in (e.g. ones that have as prerequisites the manifestation of some vice); in such situations, Hursthouse's criterion implies (counter-intuitively) that every action is wrong. See Robert Johnson (2003), pp. 810–834 for a lucid presentation of this criticism. For an interesting virtue-based hypothetical criterion of right action that avoids this problem see Jason Kawall (2002), pp. 197–222. Kawall essentially ties right action to what an ideally virtuous person would deem right in the agent's situation and not to what the ideally virtuous person would *do* in such a situation. For another hypothetical criterion of right action that purports to avoid the criticisms of Hursthouse's view, see Tiberius (2006) pp. 247–265. Tiberius makes the rightness of an action a function of the reasons that would guide the perfectly virtuous person in a situation similar in certain respects to the agent's.

⁸ For general defenses of agent-based approaches to right action, see Watson (1990), van Zyl (2005) and (2009a), and Kawall (2009).

2 Slotean Virtue Ethics

In *Morals From Motives* Michael Slote formulates a number of different agent-based theories of right action. Slote focuses most of his attention on a class of agent-based theories he calls “warm”. Warm theories evaluate actions by appeal to altruistic and other-regarding motives like benevolence, compassion, and love. The warm views divide into two kinds: impartialism and partialism. According to impartialism, the maximally admirable or supremely good motive is egalitarian concern for others. According to partialism—the version of warm agent-basing Slote ultimately prefers—the maximally admirable or supremely good motive is one that expresses a balanced degree of care between the agent’s intimates and non-intimates.⁹

Despite their differences, partialistic and impartialistic theories share a common core: each ties the moral rightness of an action to the quality of the motive expressed in its performance. We can make this commonality explicit by introducing a schema ‘SVE’ (for “Slotean Virtue Ethics”):

SVE: an action, x , is morally right for S at t iff S , at t , performs x from a motive that is fundamentally good.¹⁰

For Slote, moral rightness is distinct from moral permissibility. An action is morally permissible on Slote’s criterion iff it is performed from a motive that is not fundamentally bad (where a motive is not fundamentally bad iff it is either fundamentally good or morally neutral).¹¹ An action is morally wrong iff it is not morally permissible, and an action is morally obligatory (i.e., something the agent *must* do) iff it is uniquely permissible in the circumstances.¹²

Three features of SVE are worth noting. First, each theory that satisfies SVE is a *single-factor* theory: thus, each of these views implies that only the quality of the agent’s motives matter in determining the moral status of an action. The epistemic quality of the agent’s beliefs do not play a role in fixing whether an action is right or wrong; nor do the properties of the agent’s emotional profile, properties like the type of emotions expressed, their felt intensities, and their phenomenal characters. Second, each theory that satisfies SVE is *monistic*: hence, each of these views implies that only the expression of a single virtue matters in determining the moral status of an action. On the warm approaches, Slote contends that benevolence alone enjoys this privilege: expressions of courage, honesty, and justice do not play a direct role in determining whether the action is right or wrong (unless

⁹ Slote (2001), pp. 29–30 and 36–37. As Slote points out, partialistic varieties of agent-basing bear close affinity to the care-based theories of Carol Gilligan (1982), and Nel Noddings (1984).

¹⁰ Slote (2001), p. 35.

¹¹ In Slote’s own words, “an act is morally acceptable if and only if it comes from a good or virtuous motivation involving benevolence or caring (about the well-being of others) or at least doesn’t come from bad or inferior motivation involving malice or indifference to humanity” (2001), p. 38.

¹² These definitions logically imply the following:

- (1) Some permissible actions are not morally right (since some permissible actions are performed from morally neutral motives).
- (2) Some obligatory actions are not morally right (since sometimes the only permissible action in the situation stems from a morally neutral motive).

See van Zyl (2009b), pp. 91–104 for a defense of claim (1) and Russell (2008), pp. 299–315 for a defense of claim (2).

the expression of these virtues had some direct connection to the expression of benevolence). Third, each view that satisfies SVE is *non-comparative*: thus, each of these views implies that acting from a motive that is intrinsically admirable is sufficient for making that action right and acting from a motive that is intrinsically deplorable is sufficient for making that action wrong. Thus, the moral status of an action does not depend upon the values of the motives for alternative actions. For example, suppose an agent has available only two alternatives, g1 and g2. Suppose that g1 would stem from a good motive if performed and that g2 would stem from one that is even better. SVE implies that each act is right and that neither is obligatory. Similarly, if an agent has available only two alternatives, b1 and b2, and each would stem from a bad motive, SVE implies that each action is wrong and that neither is permissible.

With these preliminaries out of the way, let us now turn to the objections to Slote's view. The cases presented below illustrate three alleged problems for SVE: (i) that it violates "ought" implies "can", (ii) that it cannot distinguish right action from right reason, and (iii) that it cannot coherently evaluate actions that stem from motives that are conflicting and over-determined.

Case 1: The Wicked Agent Imagine a wicked person who enjoys inflicting pain on other people and who cannot find it in himself to benefit others out of a caring motive. Suppose that on a particular occasion this wicked individual has the option of either harming or not harming an innocent person. Suppose that if the wicked agent were to harm the innocent person, he would harm out of malevolence, and that if he were *not* to harm the innocent person, he would do so only out of fear of being caught and imprisoned.

Case 1's Implications for SVE Case 1 provides an example of a more general phenomenon: bad people who cannot act from good motives. Hurka¹³ and Jacobson¹⁴ have claimed that this sort of case illustrates why Slote's view violates the deontic axiom that "ought" implies "can". According to this axiom, if one has an obligation to perform a certain action (at some specific time), then he must be able to perform that action (as of that time). To see why Slote's view violates this basic principle, we need first to establish what the wicked agent's obligation is in this circumstance. It should be obvious that the agent's obligation is to avoid harming the innocent person. But by Slote's criterion, an action is obligatory only if it is uniquely permissible, i.e., only if it is the sole option in the circumstance that is not improperly motivated. To avoid being improperly motivated, the action's motive must be either fundamentally admirable or morally neutral. Thus, Slote's view implies that the wicked agent's obligation is to

(WAO): avoid harming the innocent person from a motive that is either (a) fundamentally admirable, or (b) morally neutral.

Next, we need to consider what the wicked agent's alternatives are. By stipulation, the agent's only alternatives are to

- a1: harm the innocent person from malice
- a2: avoid harming the innocent person from a pathetic form of self-concern

¹³ Hurka, (2001), p. 227.

¹⁴ Jacobson, (2002), p. 59.

Since each of the agent's alternatives involves him acting from a deplorable motive, neither one allows him to satisfy (WAO). Thus, he cannot fulfill the obligation Slote's theory assigns to him. It follows, then, that Slote's view violates "ought" implies "can".¹⁵

Case 2: The Malicious Prosecutor A prosecutor is trying to decide whether to prosecute the man whom he knows committed a serious crime. The relevant options are to prosecute or not. If the prosecutor chooses to prosecute, he will knowingly do so out of malice, and if he decides not to prosecute, he will reveal neglect for his professional and social responsibilities. (We are assuming the prosecutor does not have the option of recusing himself from the case).

Case 2's Implications for SVE Commonsense morality dictates that the prosecutor has an obligation to prosecute even if his motive in prosecuting would be bad. Thus, commonsense morality acknowledges a distinction between doing the right thing and acting for the right reason. This distinction is important for a theory to preserve, since without it, a theory will fail to be action-guiding for agents who are badly-motivated; it will offer such agents no instructions on what they are to do.¹⁶ Jacobson,¹⁷ Brady,¹⁸ and Hurka¹⁹ have argued persuasively that Slote's theory cannot accommodate the right action/right reason distinction and that it cannot explain, in agent-based terms, why the prosecutor has an obligation to prosecute. Though these commentators present slightly different forms of this objection, the outlines of each of their arguments is essentially same and resembles closely the argument that Slote's view violates "ought" implies "can": For Slote, the moral status of an action is a function of the quality of the motive expressed in its performance. In order for an action to be obligatory, it must be uniquely permissible in the circumstances. But to be uniquely permissible, the action must stem from a motive that is either fundamentally good or morally neutral. However, none of the prosecutor's alternatives has this disjunctive property, since each alternative stems from a bad motive. It follows that none of the prosecutor's actions is obligatory; each action is instead wrong.

The reason that Slote's view cannot in general allow for a person's doing the right thing for the wrong reason is that obligatory acts will always have good or neutral motives, and wrong acts will always have bad motives; thus, SVE leaves no conceptual room for act and motive "mismatch".²⁰

¹⁵ Of course, Slote could save his view from this consequence by denying that the wicked individual has an obligation to avoid harming the innocent person. After all, if the wicked individual cannot refrain from harming the innocent person with an acceptable motive, then, by the contraposition of "ought" implies "can", he has no obligation to refrain from harming the innocent person. But, of course, accepting this implication would render the view deeply implausible.

¹⁶ Das (2003), pp. 327 and van Zyl (2009a) pp. 56 make similar points.

¹⁷ Jacobson (2002), pp. 57–62.

¹⁸ Brady (2004), pp. 4–7.

¹⁹ Hurka (2001), pp. 226–227.

²⁰ It should be obvious that the "right action/wrong reason" objection is related to the "ought" implies "can" objection. Both objections are based on individuals who are not properly motivated; accordingly, either case could be used to introduce either objection. I am here keeping the objections separated primarily because they are distinguished in the literature, but also because the objections raise distinct problems. The "ought" implies "can" objection represents a formal defect in Slote's theory; it shows how Slote's interpretation of "ought" fails to cohere with the ordinary logic of this concept. By contrast, the "right action/wrong reason" objection represents a practical defect in the theory since it shows why the theory fails to be action guiding in certain circumstances.

Case 3: The Narcissistic Philanthropist A self-obsessed philanthropist funds the construction of a hospital in an impoverished area of her favorite city. She funds this construction project for two distinct reasons: (i) doing so benefits the city she loves; and (ii) doing so further enhances her reputation. Importantly, each motive is strong enough *on its own* to compel the philanthropist to fund the construction project; thus, in this circumstance, funding the project is *motivationally over-determined*. Moreover, these motives have different moral qualities: the motive to act in the city's interest is good, but the motive to narcissistically enhance her own reputation is bad.

Case 3's Implications for SVE Jacobson²¹ suggests that motive over-determination cases show that the most natural interpretations of SVE imply that a single action can be both right and wrong. Such implications threaten the view's coherence. And Driver raises a more general indeterminacy worry for any case involving "mixed motives". She asks rhetorically: "When motives are mixed up...how are actions to be evaluated?"²² Though neither Driver nor Jacobson discusses the case of the narcissistic philanthropist, it conveniently illustrates their worries. Recall that the philanthropist has two motives. The first involves a desire to benefit her favorite city. This motive reveals that the philanthropist cares about the city's residents, at least to some degree. Thus, this motive is at least somewhat admirable. Hence, according to SVE, the action that results from this motive is right. The second motive involves the philanthropist narcissistically enhancing her reputation. This motive is obviously deplorable. But if this motive is deplorable, then, according to SVE, funding the construction of the hospital from this motive is morally wrong. But we now have a situation that seems metaphysically impossible: the philanthropist's act of funding the construction of the hospital is both morally right and morally wrong. Any adequate theory of right action must not generate this kind of practical contradiction.

3 A New Form of Agent-Basing

In what follows, I propose an agent-based account of right action that attempts to avoid the objections presented in the preceding section. Although the theory will make some simplifying assumptions and remain underdeveloped in certain respects, it shall contain enough detail to make clear how it can accommodate the basic tenets of commonsense morality introduced by the cases presented above.

On the current proposal, the deontic status of an action is a function of whether the agent maximizes a quantity I will call 'net intrinsic virtue value' (or 'net-IVV' for short). Intuitively, this quantity represents the agent's balance of virtue over vice in the performance of an action. To provide a more formal statement of this proposal and to highlight some of its main features, I will first need to introduce a few technical concepts. I will start with a brief discussion of how I shall understand virtue and vice.

For our purposes we shall assume a fairly neutral account of virtue and vice, one that characterizes the virtues and vices in a way that any agent-baser should be willing to accept. An obvious way to do this is to define the virtues and vices by direct appeal to aretaic properties like excellence (or admirability) and deficiency (or deplorability), respectively.²³

²¹ Jacobson (2002), p. 56.

²² Driver (1995), p. 286.

²³ Thus, I am more or less following Slote (1992), pp. 93–96 in my analysis of virtues and vices.

DefV+: V+ is a virtue iff V+ is a disposition whose expression in action increases the degree to which the agent is admirable or excellent (other things being equal).

DefV-: V- is a vice iff V- is a disposition whose expression in action increases the degree to which the agent is deplorable or deficient (other things being equal).

The concepts of admirability and deplorability function in the above principles as theoretical primitives. Accordingly, they are what ultimately explain the phenomena they are called upon to explain (namely, what distinguishes a virtue from a non-virtue and a vice from a non-vice). Thus, even if it turned out (e.g.) that a person is admirable iff she is disposed to maximize utility, DefV+ would still be classified as an aretaic theory and not be classified as a consequentialist theory. The reason is that DefV+ purports to do more than state a mere biconditional; it purports to *explain* what makes a virtue a virtue. Thus, in the case we are imagining, the agent-baser would insist that *admirability* plays the role of explaining what makes something a virtue; maximizing utility would merely co-vary with this more fundamental, explanatory property.

Although we shall treat the concepts of admirability and deplorability as theoretical primitives (and hence treat these concepts as incapable of definition or analysis), we can still clarify their meanings by outlining some of their broad connections to other normative concepts. For our purposes, the following should make clear how I intend to use these terms.

- Admirability A person is admirable (or does something admirably) when she deserves praise or merits esteem for being the kind of person she is (or for acting in a certain remarkable way). Hence, a person is admirable when it would be fitting or appropriate to reward or commend her.
- Deplorability A person is deplorable (or does something deplorably) when she deserves rebuke or merits disdain for being the kind of person she is (or for acting in a pathetic way). Hence a person is deplorable when it would be fitting or appropriate to reprimand or punish her.

Given these connections, DefV+ and DefV- should now be precise enough to identify some of the traits that satisfy these principles. Although I will not argue for this claim, I think it is reasonable to assume that the standard list of traits we call “moral virtues” satisfy DefV+, and the standard list of traits we call “moral vices” satisfy DefV-. This is because benevolence, justice, fidelity, temperance, courage, and so on, make their possessors better, more admirable persons, and because malevolence, injustice, infidelity, intemperance, cowardice, and so on, detract from the worth of individuals and make them more deplorable. Consider, for example, justice. Deliberately giving others what they deserve surely makes a person worthy of admiration; it enhances the degree to which she is an excellent person (other things equal) and provides grounds for praising her. By contrast, injustice increases the degree to which a person deplorable (other things equal) and makes her worthy of disdain. A deliberate failure to give others what they deserve reflects poorly on a person’s character and detracts from her excellence.

With these conceptions of virtue and vice in hand, let us now define another technical term: an agent’s “psychological configuration”. Whenever a person acts, she does so with certain thoughts in mind, with certain beliefs and intentions, with a certain pattern of motivation, and with certain accompanying emotions. I will call this collection of psychological elements the agent’s *psychological configuration* for that action. The psychological configuration of an action can be understood as the total state of the agent’s psychology at (and during) the time of action. It is a “snapshot” of the mental life of the agent as she behaves.

On the current proposal, psychological configurations link up with virtues in an important way. For each virtue, each psychological configuration will conform or fail to conform to the standards for expressing that particular virtue. This follows from the fact that each virtue requires the agent to exhibit some psychological features and that she fail to exhibit others.²⁴ (By DefV+, a condition's satisfaction will count as necessary for the expression of a given virtue only if its satisfaction will enhance the degree to which an agent is admirable, other things equal). Thus, justice requires, in part, that agents be motivated to give to each his or her due (to do so would enhance the agent's admirability). Justice also requires (e.g.) that agents *not* be motivated entirely by feelings of guilt (to do so would detract from the agent's admirability).²⁵ Of course, conforming to the standards for expressing a virtue is not an all or nothing affair. A person could satisfy some of the requirements for expressing a given virtue and fail to satisfy others. Thus, a person could give each person his due but in fact be motivated entirely by feelings of guilt. If so, one condition of justice would be satisfied but another would not. Furthermore, satisfying a given condition is a matter of degree. In many instances, a person will satisfy a given condition to some less-than-maximum degree. For example, person A might owe person B \$100 but pay back only \$80. A, then, would satisfy the condition to give each his due, but do so only partially. The overall degree to which an agent expresses a virtue, then, is a function of the number of conditions the agent satisfies for that virtue and the degree to which she satisfies each of these conditions (with perhaps some conditions being weighted more heavily than others).

Psychological configurations and vices link up in a parallel way. For each vice, each psychological configuration will conform or fail to conform to the standards for expressing that particular vice. (By DefV-, these standards will be set according to whether their satisfaction increases the degree to which a person is deplorable). Of course, as with each of the virtues, conforming to the standards for expressing a given vice is not an all or nothing affair. A person could satisfy some of the requirements for expressing a given vice and fail to satisfy others. And among those requirements that she does satisfy, she might do so to some less-than-maximum degree. Thus, the overall degree to which an agent expresses a given vice will be a function of the number of conditions she satisfies for that vice and the degree to which she satisfies each of these conditions (with perhaps some conditions being weighted more heavily than others).

It will be helpful to summarize the main points of the last few paragraphs. I have claimed that (a) virtues are traits whose expression in action increases the degree to which a person is admirable (other things being equal), and that vices are traits whose expression in action increase the degree to which a person is deplorable (other things being equal); (b) each virtue and each vice specifies a set of conditions that an agent must meet in order to act in a way that expresses that virtue or that vice; and (c) each psychological configuration satisfies (or fails to satisfy) these conditions to some degree or other. These observations support two main principles.

PV+: For each person, *p*, time, *t*, alternative, *x*, and virtue, *v*+, there is some number, *n*, that represents the extent to which *p* would express *v*+ at *t* by performing *x*.²⁶

²⁴ Cf. Thomas Hurka (2006), pp. 69–76, and Williams (1985), pp. 13–23.

²⁵ Other conditions would have to be specified by a full-blown theory of justice. It is not my intention here to offer an analysis of justice or of any other virtue.

²⁶ To be clear, on the current view, alternatives contain as constituents the agent's psychological configuration.

PV-: For each person, p , time, t , alternative, x , and vice, $v-$, there is some number, m , that represents the extent to which p would express $v-$ at t by performing x .

The numbers n and m represent, respectively, the degree to which a person expresses a particular virtue or a particular vice in the performance of a given action, and hence, by principles DefV+ and DefV-, represent, respectively, the degree to which a person is excellent/admirable or deficient/deplorable in the performance of a given action.

Of course, since PV+ and PV- introduce measures of admirability and deplorability, we must specify the range of values the variables n and m can take and make clear how the various values on these variables are to be interpreted. To this end, let us suppose that to each trait that satisfies DefV+, there is a scale that is indexed to it. Thus, we have a justice scale, a benevolence scale (and perhaps a fidelity scale, a temperance scale, a courage scale, and so on). Further, let us suppose that each scale's range is the unit interval $[0, 1]$. On each scale, we can let 0 represent the agent expressing the virtue indexed to that scale to no degree at all. At this zero-point, the agent is represented as being neither admirable nor deplorable from the point of view of the relevant virtue. We can let higher positive numbers represent expressing the relevant virtue to increasing degrees (and hence represent increasing degrees of admirability), and let 1 represent expressing the relevant virtue to the maximum degree (and hence represent being maximally admirable from the point of view of that particular virtue). Thus, actions in which the agent's psychological configuration expresses no justice but a maximum amount of benevolence are assigned a 0 on the justice scale and a 1 on the benevolence scale. And an action in which the agent expresses a moderate amount of temperance and a more than moderate (but less than maximum) amount of fidelity is assigned a .5 on the temperance scale and a .7 or .8 on the fidelity scale.

The scales for the vices are similar to those for the virtues. We can suppose that for each trait that satisfies DefV-, there is a scale that is indexed to it. Thus, we have an injustice scale, a malevolence scale (and perhaps an infidelity scale, an intemperance scale, a cowardice scale, and so on). Each scale ranges the interval $[0, -1]$. We can let 0 represent expressing a given vice to no degree at all (and hence represent being neither admirable nor deficient from the point of view of that vice), and let -1 represent expressing a given vice to the maximum degree (and hence represent being maximally deplorable). Intermediate negative numbers will represent expressions of intermediate degrees of vice (and hence represent intermediate degrees of deplorability). Thus, an action in which the agent expresses the maximum amount of malevolence and a slightly less than moderate amount of intemperance is assigned a -1 on the malevolence scale and a $-.4$ on the intemperance scale.²⁷

Now that we have some grasp for how we can use numbers to represent expressions of virtue and vice, we are in a position to define the current proposal's main concept: *net-intrinsic virtue value* ('net-IVV', for short). Intuitively, an action's net-IVV represents how virtuous (or admirable) the agent is *on balance* in performing that action. More formally, an action's net-IVV is the algebraic sum of the action's intrinsic virtue value ('IVV+' for short) and the action's intrinsic vice value ('IVV-' for short).²⁸

²⁷ The claim that expressions of virtue and vice are measurable phenomena is entirely compatible with the kind of contextualism that Swanton (2003) endorses. To make the measures of expression of virtue and vice compatible with contextualism, all that we would need to do (formally speaking) is let the standards for satisfying the conditions of the various virtues (and of the various vices) vary from circumstance to circumstance.

²⁸ We could, of course, weight some virtues more heavily than others in order to discount "easy virtue". This would complicate the computation of net-IVV but perhaps make the view more plausible. Thanks to an anonymous referee for making this suggestion.

The IVV+ for a given action is the sum of the values assigned to that action by the various virtue-indexed scales. Thus, if an action gets a .2 on the justice scale, a .4 on the benevolence scale, and 0s for each of the other virtues, then this action's IVV+ will equal $(.2 + .4) = .6$.

The IVV− for a given action is the sum of the values assigned to that action by the various vice-indexed scales. For example, suppose that the same action as the one mentioned above gets assigned −.3 on the cowardice scale, a −.1 on the infidelity scale and 0s for each of the other vices. Then this action's IVV− will equal $(-.3 + -.1) = -.4$. Since the action's net-IVV is the sum of its IVV+ and IVV−, the action's net-IVV will equal $.6 + (-.4) = .2$ ²⁹

We are now ready for an official statement of the theory. For each agent, at each time, each of his alternatives has a net-IVV. This theory makes right action a matter of maximizing net-IVV.

MEV: an action, x , is morally right for S at t iff x maximizes net-IVV for S at t .

MEV acknowledges no distinction between right and permissible actions. Thus, an action is right iff it is permissible. Actions are wrong, on this view, when they *fail* to maximize net-IVV, and actions are obligatory when they *uniquely* maximize net-IVV (i.e., they are the only right or permissible ones in the circumstance). If two or more actions maximize net-IVV, then each maximizing action is right (or permissible), and no action is obligatory. If an agent confronts “all bad options”, a situation in which each of his actions has negative net-IVV, then MEV implies that the agent ought to perform the action that on balance makes him *least* vicious. This will be the action whose net-IVV is closest to 0.

MEV is a form of agent-basing since the deontic status of an action is determined entirely by aretaic “inner” states of the agent. However, MEV differs from Slote's form of agent-basing in a number of important ways. First, MEV is a *multi-factor* view and not a single factor view. According to MEV, a person's entire psychological profile (along with the person's entire psychological profile for each of the alternatives) plays a role in determining the deontic status of an action. This whole profile contains a number of distinct elements, including the beliefs, desires, intentions, and emotions of the agent. Unlike Slote's theory, then, MEV takes more than the agent's motives into account when determining whether an action is right or wrong. Second, MEV is *pluralist*, not monist. MEV considers all of the traits that satisfy DefV+ and DefV− as relevant to determining which of an agent's options are morally right. It does not focus solely upon the degree to which the agent is excellent in expressing benevolence and malevolence in each option.³⁰ Third, MEV is *comparative* (or non-absolutistic). This implies that one cannot determine

²⁹ Some temporal issues surrounding the calculation of net-IVV are worth mentioning. To find an action's net-IVV, we have to consider the interval of time that it takes the agent to perform the action. Sometimes, actions amount only to “decisions” and are nearly temporally un-extended. In these circumstances, the action's net-IVV is quite easy to find. Just consider the time of the decision: add up the values of all of the expressions of virtue and all of the expressions of vice that occur in the performance of that action at that time. In other cases, the actions include more than just single decisions: they include a number of complex behaviors that take some longer period of time to perform. To find the agent's net-IVV for a complex action, we have to consider all of the expressions of virtue and all of the expressions of vices that occur in the performance of that action over that extended interval.

³⁰ Though the view Swanton (2003) endorses is pluralist, it is in at least two ways distinct from the present proposal. First, Swanton says her view makes actions that are “overall vicious” wrong, p. 240. Second, she claims that her “target-centered view will tolerate moral luck in the attainment of moral rightness, for rightness may depend in part on the results not entirely within the control of the agent”, p. 232. Since MEV is an internalist form of agent-basing it does not countenance luck (at least with respect to the results of actions); thus it is clearly different from Swanton's.

the moral status of a given action without examining the aretaic values of its alternatives. It is not enough to know that the agent expresses virtue (or that he is admirable in some respect) in order to know that the action is right. If there is some other alternative available to the agent that has a higher net-IVV, then any virtuous but non-maximizing action is wrong. In addition, MEV implies that an agent's being on balance vicious (or deplorable) in the performance of an action is not sufficient for that action's being wrong. If the agent is overall vicious in the performance of an action, but the performance of any of the other options would make the agent even more vicious, then the option that makes him least vicious (or least deplorable) will be obligatory.³¹

In the section below, I show how MEV's multi-factorial, pluralist, and comparative features protect it from the objections raised in Section 2.

4 Revisiting the Objections

The Wicked Masochist According to MEV, even if a person cannot bring himself to act from a good motive, there will still be a best act (or several best acts) available to him in his situation. Since this best act will maximize net-IVV, it will be the one he has an obligation to perform. Thus, MEV will not violate "ought" implies "can". To see how this reasoning works in a given case, consider again the wicked person who cannot act in a manner that expresses concern for others. Let us call him 'W'. Intuitively, we know that W has an obligation to *not* harm the innocent person (whom we will call 'IP'), even if the only way W can bring himself to do this is out of some pathetic form of self-concern. MEV explains this obligation as follows: W is more deplorable on balance under the circumstance in which he harms IP than he is under the circumstances in which he does not. This becomes clear once we consider the vices W would exhibit if he were to harm IP and we compare these to the vices W would exhibit if he were not to harm IP. If W were to harm IP, W would manifest (at a minimum) malevolence (since the action is cruel) and injustice (since the person is innocent and does not deserve to be harmed). Principle PV- would assign to each of these vices numbers close to -1 on their respective scales. The sum of these numbers (which for simplicity we shall assume represents the total amount of IVV- for this action) would be fairly close to -2 . By comparison, even though W would still express a vice if he were to choose *not* to harm IP (i.e., a pathetic form of self-concern), he would not act from the motive of malevolence. Thus, the sum of the numbers representing the total amount of IVV- for this option is less negative (i.e., closer to 0) than it is for the other. Since W is on balance less deplorable in the performance of the action in which he refrains from harming IP than he is in the performance of the action in which he actually harms him (and since there are no other alternatives available to him), W has a moral obligation to avoid harming IP.

Consideration of this case points to the following:

Conclusion 1: Since MEV is comparative, it does not violate "ought" implies "can". Thus, MEV does not require that agents do things they cannot do and hence respects the standard logic of the concept of obligation.

The Malicious Prosecutor Consider again the malicious prosecutor. The challenge here is to explain, in agent-based terms, how it is possible for the prosecutor to be obligated to

³¹ To some, this last implication may sound strange. However, it is no more strange than the implications that standard forms of consequentialism have for cases in which no matter what the agent does, he will bring about a bad result. In these sorts of situation, the agent is obligated to bring about the least worst outcome.

prosecute even though his motive in prosecuting will be bad, and how, more generally, MEV can allow right action for the wrong reason. MEV has the resources to explain both of these things. First, the prosecutor knows that failing to prosecute will result in the criminal's going free. This would not only put the public at risk, but would represent neglect for his professional responsibilities. Thus, while not prosecuting might save him from expressing malevolence, it would come at the cost of being deplorable in at least two distinct ways: being indifferent to public welfare and being disrespectful to the office he pledged to serve. If this is correct, then the alternative of prosecuting would express (at least to some degree) concern for public welfare and respect for the office he pledged to serve. It seems then that on balance the prosecutor would be more virtuous or admirable if he were to prosecute than he would be if he were to fail to prosecute. To put this point directly in the terms of MEV, the prosecutor's act of prosecuting would yield a positive amount of net-IVV while his act of not prosecuting would yield a negative amount of net-IVV. Thus, prosecuting maximizes net-IVV and constitutes the prosecutor's obligation in this circumstance.

More generally, MEV, can distinguish right action from right reason because it makes the deontic status of an action a function of a number of different psychological elements. A person's psychological configuration for a given action may contain a bad element (in the form of a bad motive) but also contain a number of good elements (in the form of other good motives or other conative, doxastic or phenomenal elements that determine the expression of other virtues). The good elements expressed in an action can sometimes outweigh the bad elements expressed in that same action. This allows MEV to countenance act and motive "mismatch" and allows it to preserve the distinction between acting rightly and acting for the right reason.

Conclusion 2: Since MEV is a multi-factor, comparative and pluralist view, it can allow a person to act rightly even if his motive in performing that action is bad. Thus, MEV is action-guiding, even for badly-motivated agents.

The Narcissistic Philanthropist This case presented a problem from Slote's view because the philanthropist's act of funding the construction of the hospital reflected two conflicting motives, each one strong enough on its own to produce the desired behavior. Since one of the motives was admirable and the other deplorable, SVE implies that funding the hospital project is both right and wrong. This implication threatens SVE's coherence. So how does MEV fare with respect to this case, and more generally, how does it fare with respect to the phenomenon of conflicting motivational over-determination?

MEV generates the right result in the philanthropist case and coherently evaluates cases of conflicting motivational over-determination. By stipulation, the philanthropist is concerned at least to some degree about the welfare of the community, so she manifests some degree of benevolence. Thus, in funding the project she displays some sort of admirability. At the same time, she also funds the project because it would enhance her reputation. We already stipulated that excessive concern for her reputation is deplorable. Given this, the act of funding the construction of the hospital is partly admirable and partly deplorable. However, the degree of admirability she manifests in funding the hospital clearly outweighs her degree of deplorability (after all, the magnitude of her benefit is so significant). So funding the hospital project yields a substantial amount of positive net-IVV. Of course, the philanthropist's alternative of not funding the project has a negative balance of admirability over deplorability and yields a negative amount of net-IVV (for doing so would be to knowingly forgo the substantial benefit to the community). Since funding the hospital project maximizes net-IVV, MEV implies that this action is right (and indeed obligatory). It does not also classify it as wrong. Thus, the motive over-determination presented by the narcissistic philanthropist does not threaten MEV's coherence.

More generally, conflicting motive over-determination will never present a problem for MEV. This is because the view is pluralist and not monist, and also because MEV is a multi-factor and not a single factor view. To determine an action's net-IVV, MEV takes into consideration the agent's entire psychology. If an agent's psychological profile contains a complex motivational structure with some elements representing virtues (and hence admirability) and others representing vices, (and hence deplorability) then the action's net-IVV will reflect this fact and do so in a way that represents the agent's *balance* of virtue over vice. In the end, each act will be assigned a single deontic evaluation based on the relative magnitude of its net-IVV. Thus, no "double deontic assignments" are possible, and no incoherence can result.

Conclusion 3: Since MEV is both a multi-factor and pluralist view, it assigns only one moral status to any given action, including those actions that stem from motives that are both over-determined and conflicting. Thus, MEV is evaluatively coherent.

Based on the analyses of these three cases and on the analyses of the more general phenomena these cases introduce, it seems reasonably clear that MEV has the resources to meet the challenges presented for standard Slottean forms of agent-basing.

5 Concluding Remarks

In closing, I wish to point out that the particular version of agent-basing I have here proposed is not the only one capable of avoiding the problems presented for Slotte's view. Any theory that is comparative, multi-factorial and pluralist is likely to survive these challenges as well. Some of these versions might choose to drop the maximizing assumption I have made and replace it instead with a satisficing principle. Others might do away with the commensurability of virtue I have here been assuming and adopt instead a principle of rough comparability. Still others may want to incorporate "holistic" considerations into the calculation of net-IVV and thus reject the simple summation approach I find attractive. However such theories are ultimately developed, I suspect they will be able to accommodate the basic distinctions, principles and intuitions introduced by the above cases as long as they retain MEV's central features: some means for ranking and subsequently selecting alternative actions; a commitment to making the rankings of alternative actions a function of more than the moral quality of a single motive; and a requirement that many different virtues and vices contribute positively or negatively to the deontic status of an action. Given the large number of agent-based theories that are consistent with these criteria, I think virtue ethicists would do well to explore some of these possibilities. A view that is ultimately satisfactory may turn up in the process.

In this paper, I have proposed an agent-based theory of right action that requires agents to choose options in which they express a maximal balance of virtue over vice, or admirability over deplorability. Since this theory is multi-factor, pluralistic, and comparative, it (i) satisfies "ought" implies "can", (ii) allows for a person's doing the right thing for the wrong reason, and (iii) accounts for cases involving mixed and over-determined motivation. Thus, this agent-based view, and any agent-based view that shares its main features, cannot be dismissed for being incompatible with some of the more basic features of commonsense morality.³²

³² Many thanks are due to Fred Feldman, Meghan Masto, and two anonymous reviewers for valuable comments and helpful suggestions on earlier drafts of this paper.

References

- Brady M (2004) Against agent-based virtue ethics. *Philos Pap* 33:1–10
- Copp D, Sobel D (2004) Morality and virtue: an assessment of some recent work in virtue ethics. *Ethics* 114:514–554
- Das R (2003) Virtue ethics and right action. *Australas J Philos* 81:324–339
- Driver J (1995) Monkeying with motives: agent-basing virtue ethics. *Utilitas* 7(2):281–288
- Gelfand S (2000) Hypothetical agent-based virtue ethics. *Southwest Philos Rev: J Southwest Philos Soc* 17 (1):85–94
- Gilligan C (1982) *In a different voice: psychological theory and women's development*. Harvard University Press, Cambridge
- Hurka T (2001) *Virtue, vice and value*. Clarendon, Oxford
- Hurka T (2006) Virtuous act, virtuous dispositions. *Analysis* 66:69–76
- Hursthouse R (1999) *On virtue ethics*. Oxford University Press, Oxford
- Jacobson D (2002) An unsolved problem for Slote's agent-based virtue ethics. *Philos Stud* 111:53–67
- Johnson R (2003) Virtue and right. *Ethics* 113:810–834
- Kawall J (2002) Virtue theory and ideal observers. *Philos Stud* 109:197–222
- Kawall J (2009) In defense of the primacy of the virtues. *Journal of Ethics & Social Philosophy* 3(2): www.jesp.org.
- Noddings N (1984) *Caring: a feminine approach to ethics and moral education*. University of California Press, Berkeley
- Russell D (2008) That “ought” does not imply “right”: why it matters for virtue ethics. *South J Philos* XLVI:299–315
- Slote M (1992) *From morality to virtue*. Oxford University Press, Oxford
- Slote M (2001) *Morals from motives*. Oxford University Press, Oxford
- Stocker M (1971) “Ought” and “can”. *Australas J Philos* 49:303–315
- Stohr K, Heath Wellman C (2002) Recent work on virtue ethics. *Am Philos Q* 39(1):49–72
- Swanton C (2003) *Virtue ethics: a pluralistic view*. Oxford University Press, Oxford
- Tiberius V (2006) How to think about virtue and right. *Philos Pap* 35(2):247–265
- van Zyl L (2005) In defence of agent-based virtue ethics. *Philos Pap* 34(2):273–288
- van Zyl L (2009a) Agent-based virtue ethics and the problem of action guidance. *J Moral Philos* 6:50–69
- van Zyl L (2009b) Accidental rightness. *Philosophia* 37:91–104
- Watson G (1990) On the primacy of character. In: Flanagan O, Rorty A (eds) *Identity, character, and morality*. M.I.T.
- Williams B (1985) Acting as the virtuous person acts. In: Heinamen R (eds) *Aristotle and moral realism*. University College London Press