Bond University

Research Repository

Agent-based theories of right action

Cox, Damian

Published in:
Ethical Theory and Moral Practice

DOI:
10.1007/s10677-006-9029-3

Licence:
Other

Link to output in Bond University research repository.

Recommended citation(APA):
Cox, D. (2006). Agent-based theories of right action. *Ethical Theory and Moral Practice*, *9*(5), 505-515. https://doi.org/10.1007/s10677-006-9029-3

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

For more information, or if you believe that this document breaches copyright, please contact the Bond University research repository coordinator.

Download date: 05 Jun 2024

DAMIAN COX

AGENT-BASED THEORIES OF RIGHT ACTION

ABSTRACT: In this paper, I develop an objection to agent-based accounts of right action. Agent-based accounts of right action attempt to derive moral judgment of actions from judgment of the inner quality of virtuous agents and virtuous agency. A moral theory ought to be something that moral agents can permissibly use in moral deliberation. I argue for a principle that captures this intuition and show that, for a broad range of other-directed virtues and motives, agent-based accounts of right action fail to satisfy this principle.

KEY WORDS: right action, virtues, moral evaluation, moral deliberation

AGENT-BASING

Over recent years there has been a resurgence of interest in agent-based ethical theory. It is an approach closely associated with the work of Michael Slote, but other virtue theorists, such as Rosalind Hursthouse and Christine McKinnon, develop views that are close to Slote's in crucial respects.[1] According to Slote, an agent-based ethics is "… one that treats the moral or ethical status of actions as entirely *derivative* from independent and *fundamental* ethical/aretaic facts (or claims) about the motives, dispositions, or inner life of moral individuals."[2] It is a view that attempts to derive moral judgment of actions from judgment of the inner quality of virtuous agents and virtuous agency.

An agent-based approach to ethical theory has two aspects to it. First, it involves a specification of relevant aretaic qualities of agents' motives, dispositions and capacities. Second, it involves a demonstration of how moral judgment of actions is derived from these qualities. According to Slote, aretaic qualities employed in agent-based ethics are fundamental. I take this to mean that they must be independent of non-aretaic moral

principles or moral values. Consider, for example, Aristotle's discussion of the virtue of equity. In characterizing this virtue, Aristotle makes use of the concept of *epieikeia*.[3] This is a sense of what would be a fair and reasonable outcome in legal situations. Aristotle appears to think of *epieikeia* as akin to perception; it is a capacity to *see* what would be a fair and reasonable outcome of a situation and is not something that can be reduced to the application of rules or laws. But this means that, for Aristotle, equity is not a fundamental aretaic fact about a just person. The virtue of equity is based in part on an agent's capacity to perceive what would be fair and so depends upon an independent account of the character and moral value of fair outcomes.[4] Slote calls Aristotle's view, and views like it in this general respect, "agent-focused" accounts of virtue.[5]

Slote further distinguishes between agent-based and agent-prior theories. An account of a virtue is agent-prior if it depends upon non-moral value judgments. Whereas agent-focused theories make essential reference to moral values like the value of fair or just outcomes, agent-prior theories make use of non-moral values like that of human flourishing or well-being. Thus, in Slote's terms, a eudaimonic virtue ethics – one in which virtues are identified with dispositions to behave in ways that are beneficial to oneself and/or to others – is agent-prior rather than agent-based.[6] Although Slote makes a pertinent distinction here, it is easy to over-draw its significance and I think it is a mistake to restrict the term "agent-based" to Slote's use of it. Slote's agent-based and agent-prior ethical theories both derive the moral significance of actions from virtuous agency and do this without appealing to non-aretaic *moral* notions. It thus seems appropriate to class them both as kinds of agent-based theory. Where they primarily differ is over the role human flourishing plays in the constitution of virtue. Slote, for

example, thinks of empathic caring as the touchstone of virtuous agency and its moral significance grounded in sentiments of admiration or approval we naturally feel towards manifestations of empathic caring.[7] The connection between virtue and well-being here is indirect: a caring world is a world in which things generally go well for us, but, on Slote's story, this isn't why empathic care is morally significant to us. Slote's sentimentalist approach differs markedly from that of philosophers, such as Hursthouse and McKinnon, who ground the moral significance of virtue directly in what it takes for a human life to go well. Nonetheless, agent-based and agent-prior accounts share a key feature. They attempt to ground all moral judgment in judgments of virtue and virtuous agency that do not themselves make use of non-aretaic moral notions. I address them both in this discussion of agent-based theories of right action.

Agent-based accounts of virtuous action identify the morally significant features of an action in terms the qualities of agency it expresses or reflects. These might include motives for the act, dispositions underlying the act or to its motive, and capacities, including epistemic capacities, that are exercised in deciding how to act. What these accounts can't be made to rely upon is adventitious features of the act or its circumstances. Being successful in one's intention is not a quality of agency but a fact about outcomes. Even if we think that a key aspect of virtue is reliability – so that possession of a virtue, say, is made to depend upon an agent's capacity to reliably succeed in ends internal to the virtue – this does not entail the virtuous agent's success on any particular occasion. Even highly reliable agents can fail in their goals if circumstances turn against them. This means that accounts of virtuous agency that build in an explicit success condition are not agent-based theories. Linda Zagzebski is a

prominent defender of this kind of account of virtuous agency, which means that her's is not an agent-based ethics.[8]

I am aware of three *prima facie* plausible ways of defining right action in terms of virtuous agency. They are:

(P_1) An agent S's performing act A in circumstances C is right if and only if it is a virtuous act.[9]

(P_2) An agent S's performing act A in circumstances C is right if and only if a virtuous person, acting virtuously, would perform act A in circumstances C.[10]

(P_3) An agent S's performing act A in circumstances C is right if and only if a virtuous person would approve of S's performing act A in circumstances C.[11]

I do not think that any of these principles are true given the constraints upon agent-based accounts of virtuous agency I have described. Each of the principles fails to accommodate what I think is a crucial relationship between right action and the possibility of deliberating about what is right.

AN ARGUMENT FROM MORAL DELIBERATION

What is the relationship between acting rightly and deliberating about what is right? It seems possible to act rightly without deliberating explicitly at all, so any relationship here isn't a direct one. Nonetheless, I think there is an important connection between acting rightly and the *legitimacy* of deliberating about what is right. Consider the following principle:

(Q) If it is morally right for S to ϕ, then there are possible deliberations that, were S to undertake them, would conclusively recommend to S that they ϕ. These deliberations must be:

(i) accurate (i.e. pick out those features of S's ϕ-ing that make it right), and

(ii) morally permissible.

I intend to use this principle to argue against various agent-based equations of virtuous agency and right action. The idea underlying Q(i) is that a right action is right in virtue of properties of the act and its circumstances; properties that can, in principle, be identified and employed in moral deliberation. Consider the example of spontaneously coming to the aid of an injured person encountered on a lonely stretch of road. If it is right to come to the person's aid, then there must be some features of the person's predicament, of the moral agent's predicament, or of the action of coming to person's aid that make it right. In principle, an account of these right-making features should be available either to retrospective justification of the action or, more importantly in the present context, to

prospective deliberation over it. Of course, time-pressures, as well as epistemic and cognitive limitations, may well rule out explicit prospective deliberation, but considerations that identify what makes an action right must nevertheless exist. Accurate moral deliberation is always at least a theoretical possibility when actions are right, because there is always a conclusive story to be told about how they are right.[12]

Q(ii) is crucial for my case against agent-based virtue ethics. The combination of Q(i) and Q(ii) generates the principle that an action is right only if one could rightly decide to perform it by deliberating explicitly in terms of what makes it right. For an action to count as right, accurate moral deliberation must not only be a theoretical possibility, it must also be morally permissible.[13] I think the principle is a robust and plausible one. One consideration in its favor is the requirement that moral theories be action-guiding. It seems that part of the motivation for articulating a theory of right action is to furnish means of deciding how to act, particularly in difficult cases. However, it is not obvious that a theory that proscribed morally deliberating in its own terms would satisfy this requirement. Such a theory would likely fail moral agents at exactly the point at which an accurate grasp of the rights and wrongs of proposed actions is most salient: when agents are deciding how to act in difficult and morally troubling circumstances. This consideration isn't decisive, however. A theory may proscribe moral deliberation in its own terms yet furnish an alternative method of deliberation, one that succeeds indirectly. In this way a theory might succeed in guiding a moral agent towards right actions without involving the agent in explicit reflection of right and wrong. For example, a consequentialist might recommend a form of moral deliberation that *has* the best consequences (i.e. leads to the best decisions being made) but which differs from

deliberation *about* the best consequences. Were such a theory possible, it would appear to guide action sufficiently well.

A more troubling consequence of failure to satisfy Q is that it puts sincere and conscientious moral agents in an implausibly invidious position. Say that I am committed to theory T, but that T proscribes its employment in moral deliberation. This means that I am under an obligation not to take my commitment to T fully and explicitly seriously. It is a commitment I ought to allow to do its work in the background, as it were, but which I am obliged to partition away when practical affairs beckon. Were I to reflect about what to do in terms of my understanding of what it would be right to do, I would, by T, be acting immorally. It seems to me that this is a very peculiar and unsatisfactory situation to force moral agents into. It requires agents to ignore what it is about their actions that would make them right at the very point at which this information is most salient: the point at which agents are trying to determining what they ought to do. The demand that one's understanding of rightness *not* be used in determining how to act forces an unwelcome and unhealthy cognitive dissonance upon us. The demand also appears to allow moral judges to condemn us on grounds that we ought not to have taken into account when deciding how to act, and this seems unfair.

A theory might recommend, not that I partition away my commitment to a moral theory when morally deliberating, but that I not explicitly commit to the theory at all. (Perhaps it is a theory only for the eyes of those who would judge my actions.) In this case, it seems that I am hostage to moral fortune in an altogether implausible way. Whatever propensity I have to act rightly would not be due to my grasp of what is right and my determination to do what is right, but due, say, to my being blessed with a

naturally right-following nature, or to my being conditioned by others in the right way. For example, I might be taught a rule that has the effect of leading me to decide upon actions that, unbeknownst to me, happen to be right. Hostage to fortune like this, it seems that a moral judge might condemn my actions on grounds that ought to be entirely obscure to me. Again, this seems unfair.

Consider the following example. A theory, D, defines right action as action that satisfies the categorical imperative. But D also makes it impermissible for agents to explicitly derive maxims from the categorical imperative. Instead, D requires moral agents to consult a book of commandments that lists all maxims entailed by the categorical imperative and the conditions of their application. (Suppose that one of the maxims recognized by D is the highly un-Kantian demand that agents trust the reasoning of their moral superiors over the impossibly opaque operations of their own conscience.) D is an example of a theory violating Q: it proscribes agents employing its account of rightness (satisfaction of the categorical imperative) in moral deliberation. Were I to act on a particular occasion by the wrong rule, say by mistaking the conditions of a rule's application, then I would have acted wrongly according to D. If the book of commandments really does track all relevant applications of the categorical imperative, then I could not have satisfied the categorical imperative when I acted by the wrong rule. By the lights of D, I must be judged to have acted wrongly because of this. My failing the categorical imperative is what makes my action wrong. Yet this would be a highly unsatisfactory way of describing my moral failure. Surely my moral failure consists in my looking up the wrong rule, not in my failing to satisfy the categorical imperative. Moral failure, and along with it the idea of fair moral blame, are intentional notions. They

need not apply to extensionally equivalent action-descriptions. This is one reason why Kant's identification of the good will with an agent's acting for duty, rather than acting in conformity with duty, has such intuitive force.[14] If I am right that moral failure is an intentional notion, then agents ought not be morally judged on grounds they are required to ignore in moral deliberation. In the example where I misapply the moral rules specified by D, I have violated the categorical imperative, but I am not morally blameworthy for this. I am morally blameworthy for looking up the wrong rule.

If we accept this condition on blaming practices, how are we to interpret D? An wrong action is still wrong because it violates the categorical imperative – violation of the categorical imperative is what makes a wrong action wrong – but agents are neither to be blamed for violating the categorical imperative (they are to be blamed for not abiding by the book of commandments) nor are they to consult the categorical imperative as they decide how to act (they are to consult the book of commandments). But if judgments of rightness or wrongness play no role in judgment of moral performance and play no role in moral deliberation, what role do they play? Perhaps it is tempting to answer that rightness plays a role in determining the content of the commandments. The moral mandarins who write the book of commandments do so in light of their judgment of what is right, namely satisfaction of the categorical imperative. Yet this is not a satisfactory response. A more cogent successor to D, D*, would define rightness in terms of obedience to the commandments, and then explain the content of commandments in terms of their satisfying the categorical imperative. Compare D* to rule-utilitarianism. A rule-utilitarian may define rightness in terms of obedience to moral rules, and then determine the content of moral rules by picking out optimal consequences of general rule

compliance. It makes little sense for a theory to articulate an account of rightness that can be used neither in moral judgment nor moral deliberation, and this is why rule-utilitarianism is the most plausible form of indirect utilitarianism and why D* is more plausible (or less implausible) than D. The upshot of this argument is that theories violating Q face an awkward dilemma. A theory violating Q may recommend unfair practices of moral blame, in which agents are blamed on grounds they ought not to have taken into consideration when deciding how to act. To avoid this result, however, a theory violating Q must define rightness in such a way that rightness plays no direct role in moral practice at all.

Theories that violate Q rule out a person's deliberately pursuing what they accurately perceive to be right. Even if one doubted the practical value, wisdom, or desirability of our *only* ever pursuing what is right deliberately, it makes little sense for a moral theory to rule out the general permissibility of our doing so. Too much is lost when we lose the possibility of deliberate and explicit moral action. The possibility of explicit moral motivation is one thing lost. That we can be motivated to act because we understand an action to be right is central to any plausible account of what it is for us to take morality seriously; it is a central feature of what it is for us to possess moral integrity. In light of all the problems that failures to satisfy Q engender, I think we should accept the principle articulated by Q. Explicit and accurate moral deliberation ought be at least a permissible option for moral agents.

We find principle Q – or something very much like it – employed in Bernard Williams's well-known objection to Kantian moral theory. Williams argues that subscription to Kantian moral theory would often involve us in having 'one thought too

many'.[15] Faced with personal demands or requests – say a demand to rescue one's own child when several children are equally in need of rescue and only one rescue can be effected; or a commonplace suggestion that we visit a friend in hospital – it seems the Kantian will require of moral agents that they justify their actions to themselves in impartial terms, by ensuring that the maxim under which their action falls satisfies the impartial demands of the categorical imperative. According to Williams, however, this leads agents to have one thought too many. A good way to interpret this objection is in terms of the principle articulated in Q. The claim is that, even if the correct conclusion is arrived at by Kantian calculation in these cases, it is arrived at in the wrong way. Kantian moral deliberation would not then be a morally permissible way to decide how to act towards friends and loved ones because it involves us in morally impermissible ways of thinking about them. Kantian moral theory thus appears to furnish an account of right action that violates Q. And if this is so, the Kantian account of right action is unsatisfactory.[16]

Q can be pressed into service against agent-based accounts of right action in similar fashion. An account of right action is adequate only if it satisfies Q. However, agent-based virtue-theoretic accounts of right action do not satisfy Q. Consider those situations in which being virtuous requires caring for another. On any reasonable conception of what is involved in genuinely caring for another, care for another requires directing your attention to their needs, not to your own virtue. Deliberating over the manifestation of your own virtue fails to instantiate a caring relation to others, even if it has an indirect effect of benefiting those who you care for.[17] Deliberating over the manifestation of your own virtue is not always compatible with manifesting that virtue.

Indeed, deliberating about the manifestation of your own virtue in circumstances that call for exercise of the virtues of care exhibits a vice – the vice of moral narcissism – rather than a virtue.

Now this charge of moral narcissism is not leveled at the intended practices of agent-based virtue ethicists. Their recommendation, it seems, is that moral agents generally eschew explicit moral deliberation.[18] However, the objection I am urging is not based on the value of explicitly morally deliberating, but on the possibility and permissibility of such deliberation. A deeply entrenched habit of always deliberating in explicitly moral terms may be a character flaw. Consequentialists, deontologists and virtue ethicists are all able to advance this claim in different ways. However, agent-based virtue ethicists ought to concede that morally deliberating in terms of the right-making properties of an action is invariably wrong when other-directed virtues are in focus.[19] Other moral theories may also find reason to discourage agents from habitually deliberating over the right-making features of their actions, but generally they must do so on different grounds. For example, a deontologist may consider the right-making property of an action to be those features that make it fall under an appropriate rule. Thinking about one's actions in this way will not itself violate the rule. The exception would be rules against explicit moral deliberation *per se*, but these tend not to figure in deontological specifications of right action. Still, there may be a rule to the effect that it is wrong to *automatically* refer to rules when making decisions. A rule of this kind would allow that explicit moral deliberation is always permissible, but ought not be allowed to become an invariable habit.

Indirect versions of consequentialism, such as the rule-utilitarian example I discussed earlier, represent an interesting contrast case. According to some indirect versions of consequentialism, moral deliberation is subject to the same consequentialist evaluation as any other action. Thus it may turn out to be wrong on consequentialist grounds for a consequentialist to morally deliberate accurately by consequentialist lights. Theories that leave the matter here violate Q. So much the worse for them, I think.[20] Yet indirect consequentialists have clear options at this point, along the lines of those exploited by rule-utilitarians. Thus the indirect consequentialist may define right action in terms of things such as rules, plans, or deliberative practices that would optimize consequences were various possible constraints on their uptake satisfied. Consequentialists therefore have ways of reconciling demands for right action with demands for accurate moral deliberation, but agent-based virtue theorists do not have this option.[21]

Another instructive contrast is with Williams's 'one thought too many' critique of Kantian moral theory. A Kantian is able to respond to Williams's critique by biting the bullet: holding that, time pressures aside, it is both possible and morally appropriate to determine the justifiability one's actions, even of highly personal actions such as the preferential rescue of one's own child, in terms of the categorical imperative. If one can settle the matter automatically, without extended reflection, then so much the better; but there is nothing morally wrong with deliberating over the justifiability of one's actions *per se*. Love for a child may certainly exist alongside a determination to act under the categorical imperative, but ought never replace such a determination. And when there is a conflict between the heart and moral reason, the Kantian will insist that moral reason

trumps. While many find this an intuitively uncomfortable position, it is not an inconsistent or incoherent one. The agent-based virtue ethicist, on the other hand, confronts a rather deeper problem. Imagine a parent deliberating over the rescue of their child in explicitly agent-based terms. In this case, they do not reason agent-neutrally that the attempt to rescue their child falls under a morally acceptable maxim. They reason agent-relatively by reflecting on whether an attempt to rescue the child would adequately reflect the fact that they are a caring parent! The danger here is not primarily a matter of having one thought too many, but of having entirely the wrong kind of thought – the wrong kind of thought by the lights of agent-based virtue theory itself. My charge against agent-based accounts of right action, therefore, is that they are, in a certain way, self-undermining. They identify right-making features of an action, but in many cases, perhaps most, they also condemn morally deliberating in these terms. It is a theory that ought not be put in practice explicitly and deliberatively, and so violates Q and breaks what I think is a natural and plausible link between right action and the possibility of deliberating about what is right.

I distinguished three agent-based definitions of right action: P_1, P_2 and P_3. P_1 identifies right action with virtuous action. Deliberating in terms of what would manifest my virtue does not itself manifest my virtue, and in circumstances calling for the application of other-regarding virtues, manifests the vice of moral narcissism. In this way, I argue that P_1 fails to satisfy Q. Much the same argument applies to P_2 and P_3. P_2 and P_3 are proposals designed to accommodate the possibility that less than fully virtuous agents might nonetheless act rightly. They do this by introducing a hypothetical virtuous agent into their formulations, a device that provides a way of identifying virtuous or

quasi-virtuous kinds of acts independently of inspecting the motives and characters of actual agents. What would it be to explicitly deliberate in terms of such theories? An agent explicitly deliberating in terms of the right-making features specified by P_2 would be searching for actions that mirror virtuous behavior externally, but not necessarily internally. They would be seeking actions that are, externally at least, virtuous kinds of act. Consider once again the virtue of particularistic care. A moral agent deliberating in terms of P_2 would inspect situations for their external compatibility with demonstrations of particularistic caring. But deliberating over whether an action demonstrates the external features of particularistic caring does not itself demonstrate the external features of particularistic caring. It demonstrates a concern that one's actions meet a certain standard, but this way of deliberating does not itself meet that standard. In many situations, truly caring people would not deliberate over what to do by focusing on the standards that their actions meet; they would deliberate directly in terms of another person's needs. Failure to do this would constitute a kind of vicious unconcern for the things that really matter in the circumstances at hand. While it is true that a concern for the external standard of one's behavior factors in the needs of others indirectly – since one can only meet standards of care by attempting to satisfying some of these needs – the focus of the deliberation is all wrong from an agent-based virtue ethicist's point of view. This kind of moral deliberation is not the kind that a hypothetical fully virtuous agent would engage in. Thus explicit deliberation in terms of P_2 does not meet standards of virtuous deliberation, and so P_2 fails to satisfy Q. The case against P_3 follows similar lines. An agent deliberating in terms of P_3 would be seeking to act so as to deserve a virtuous agent's approval. To deliberate about which of a set of potential actions would

best deserve the approval of a fully virtuous agent, would not itself generally deserve the approval of a virtuous agent. Thus P_3 also fails to satisfy Q.

AGENT-BASING THE HAZZARDS

Let me illustrate the argument against agent-basing with a fictional case, the story of the Hazzards. The Hazzards are a devout family living in 17th century New England. Alma and Quincy come to believe that their thirteen year old son, Jack, is possessed by a devil because he is given to fits of shaking and foaming, refuses to attend bible-readings and has been heard using foul language within earshot of his parents. Being caring and conscientious parents, fully convinced of the reality of satanic possession and the peril of their son's soul, Alma and Quincy spend long hours discussing their best course of action. They consult their spiritual advisor and other leaders within their spiritual community, who recommend exorcism. At the exorcism, the boy is uncooperative and so must be tied to the bed. After long hours of prayer and physical and mental chastisement it seems clear to them that the devil will not out and has reached only deeper into the boy's soul. Jack's behaviour worsens alarmingly as the exorcism reaches its climax.

Alma and Quincy interpret this as a spiritual crisis and pray for enlightenment. It comes in the form of an angel, who appears to Quincy in a dream and tells him that Jack is not after all possessed of a devil. Instead, the angel tells Quincy, Jack is one of the holy fallen: his spirit is too fragile to survive below the heavenly sphere and the boy must return to his origin before he is lost forever. Quincy and Alma are distraught at the news, but the angel was adamant that Jack's case is a matter of the greatest urgency. Yet

Quincy's visitation might have been merely a dream, perhaps it was bedevilment, so Alma and Quincy take the wise step of postponing any decision, settling on a course of further prayer, reflection & consultation.

The following night, however, the angel visits Alma, repeating the message he delivered to Quincy the night before. And on the following three nights both Alma and Quincy receive in their dreams what appear to be identical visitations from the angel, all repeating the same message and emphasizing the urgency of matters. It becomes harder and harder for Alma and Quincy to reasonably resist concluding that God's word has been sent them and that it demands immediate action. They seek a further consultation with their spiritual advisor, who although initially very sceptical of the angel's theologically strange – seemingly heretical – warning, is eventually convinced that something a great spiritual significance is occurring among the Hazzards. The night of the Hazzards' visit, the advisor also has a dream in which an angel appears to him, telling him that he must prepare the Hazzards for a terrible sacrifice, and the next day he tells Alma and Quincy of the angels' visitation to him. All three pray together and search out the voice of their conscience. Many exhausting hours of prayer and reflection eventually bring resolution to Quincy and Alma. They come to a conclusion to abide by the angel's word: to do otherwise would be to hide behind the self-deceptions of doubt, fear and selfishness. To care for Jack – the most precious part of him – requires that they take his life. It would be deeply wrong to let selfish, earthly love stand in the way of their caring for his soul. Quincy and Alma move to save Jack's soul – and so they take his life.

The story of the Hazzards presents us with what appears to be a robust counterexample to agent-basing. Here is a story of seemingly virtuous people acting in

ways that are intuitively wrong, and acting wrongly because they fundamentally misconceive the situation they face. To act rightly, it seems, one must have a secure grasp of the facts at hand. Of course, agent-based virtue theorists are unlikely to be impressed by the counterexample. They can reply that Alma and Quincy may well have acted rightly by killing Jack, and that we are inclined to condemn their action because we impose illegitimate external demands upon their deliberative processes. Given Alma and Quincy's circumstances – the boy's behavior, their dreams, their epistemic community and its traditions – what else can we reasonably expect them to have done? If we cannot answer this question in terms of the pair's manifestation of virtue, then agent-based virtue ethicists will insist on the rightness of the killing. Indeed, it seems that agent-relative judgments of rightness are an inevitable consequence of adopting an agent-based moral perspective. The dialectical situation appears to end in stalemate at this point.

To get around this stalemate, I suggest we look to Alma and Quincy's moral deliberations more closely. Let me consider the broad kinds of moral deliberation available to Alma and Quincy. From the story, it seems that they reasoned as follows:

(1) We ought to do the best we can for Jack.

(2) The best we can do for Jack is kill him.

(3) Thus we ought to kill him.

This appears to be a virtuous kind of deliberation because it expresses real care for Jack, which is the relevant virtue here, and because it involves, arguably, an admirably determined and sincere pursuit of the truth and a measure of practical wisdom. It is not

necessarily a vicious trait of Alma and Quincy that they found themselves adrift in such an epistemically problematic environment. Although their way of deliberating might be virtuous, by agent-based standards it is both unsound and inaccurate. Premise 2 is false, although Alma and Quincy are perhaps not well positioned to appreciate this. Their reasoning also fails to pick out potential right-making features of the action. Neither premise refers to the aretaic properties of Alma and Quincy's actions.

So how might Alma and Quincy have deliberated about the rights and wrongs of their actions more accurately, by the agent-based virtue theorist's lights? They would have had to deliberate along lines such as these:

(1) We ought to do what would best express our care for Jack.

(2) Killing him best expresses this care.

(3) Thus we ought to kill him.

Premise (2) now seems to be true, given Alma and Quincy's understanding of the situation. Premise (1) is also correct, according to agent-based virtue theory. The deliberation thus satisfies Q(i): it picks out relevant right-making features of the action and reasons soundly on this basis. At first sight, premise (1) also looks entirely reasonable from a moral point of view. Alma and Quincy really do care for Jack, so what could be wrong with them working out how best to express this care? Isn't this a typical and innocuous way of reflecting on moral practice? It appears an innocuous way of expressing fundamentally decent motivations, but only because we expect it to be accompanied by, or to stand as shorthand for, a host of other-directed moral reflections.

But this is not the way it functions in explicit agent-based deliberation. According to agent-based accounts of right action, the catalogue of relevant moral facts is exhausted by reference to the exhibition of Alma and Quincy's virtue: no other feature of the case – neither Jack's virtue nor his well-being – is a right-making feature of the situation. Premise (1) is therefore correct by agent-based lights only because it picks out all and only the features of Alma and Quincy's situation that count morally, and the only thing that counts morally is that Alma and Quincy emerge from the whole business as caring people (or as people who act as if caring). For them to think about their own moral status as the only matter directly at stake in Jack's case is moral narcissism; it is exactly *not* what would exhibit a caring attitude to him. So Alma and Quincy are in the following predicament. Deliberating in terms of what would be best for Jack exhibits care, but by the standards of agent-based virtue theory is an inaccurate form of moral deliberation. Deliberating in terms of what best exhibits care is accurate by the standards of agent-based virtue theory, but fails to exhibit care. It is therefore impossible for Alma and Quincy to take agent-based virtue theory seriously. Therein lies the trouble with agent-based virtue ethics.

Discipline of Philosophy

Faculty of Humanities and Social Sciences

Bond University

Gold Coast, Queensland, 4229, AUSTRALIA

E-mail: dcox@bond.edu.au

REFERENCES

Brady, Michael S., “Some Worries About Normative Metaethical Sentimentalism” *Metaphilosophy*, 34(1) (2003), pp. 144-153.

Firth, R., "Ethical Absolutism and the Ideal Observer" *Philosophy and Phenomenological Research* 12(3) (1952), pp. 317-345.

Hursthouse, R., “Virtue Theory and Abortion” *Philosophy and Public Affairs* 20 (1991), pp. 223-246.

McKinnon, Christine. *Character, Virtue Theories and the Vices* Peterborough: Broadview Press, 1999.

Railton, P., “Alienation, Consequentialism and the Demands of Morality,” *Philosophy and Public Affairs* 13(2) (1984), pp. 134-171.

Slote, M., *Morals from Motives* Oxford: Oxford University Press, 2001.

Slote, M., “Sentimentalist Virtue and Moral Judgement: outline of a project” *Metaphilosophy* 34(1) (2003), pp. 131-143.

Williams, B., *Moral Luck: philosophical papers 1973-1980* Cambridge: Cambridge University Press, 1981.

Zagzebski, L., *Virtues of Mind* Cambridge: Cambridge University Press, 1996.

NOTES

[1] Hursthouse (1999), Slote (2001), McKinnon (1999).

[2] Slote (2001, 7) (italics in original)

[3] Aristotle *Nicomachean Ethics* 1137b - 1138a

[4] The independence at issue here might be interpreted as conceptual independence or as value independence. On the former interpretation, the claim would be that our concept of equity is dependent upon our concept of fair outcomes, whereas our concept of fair outcomes is not dependent on our concept of equity. On the latter interpretation, it would be that we find moral value or significance in the virtue of equity itself; we do not morally value equity as a means of generating fair outcomes, fair outcomes are morally significant to us because they reflect the good character of those who brought them about. We needn't settle the interpretative issue for the purposes of this discussion.

[5] Slote (2001, 5-7).

[6] See Hursthouse (1991) and McKinnon (1999) for accounts of eudaimonic virtue theory.

[7] Slote (2003).

[8] Zagzebski claims that "an act is an act of virtue A if and only if arises from the motivational component of A, it is something a person with virtue A would (probably) do in the circumstances, and it is successful in bringing about the end (if any) of virtue A because of these features of the act." Zagzebski (1996, 248).

[9] This appears to be Slote's preferred formulation. It appears to capture his intuition about the full moral significance of motives, in particular the idea that it is always wrong to act from vicious motives (even if what is done might also have been done from good motives).

[10] See Hursthouse (1999, 28) for a formulation along these lines.

[11] This is a formulation suggested by Michael Brady (2003, 147). Brady is concerned to accommodate the case in which a less than fully virtuous agent attempts to improve

herself. This, he claims very plausibly, is the right thing for such an agent to do though it is ruled out by the other formulations on offer.

[12] Moral particularists might wish to dispute this claim. However, the agent-based virtue ethicist I am discussing offers a more or less explicit account of the right-making features of actions, namely the aretaic qualities that underlie the agent's acting.

[13] There may be circumstances in which accurate moral deliberation is obligatory as well as permissible, but these would be special circumstances. In general, I think it is enough to insist that accurate moral deliberation ought to be at least permissible.

[14] Kant *Groundwork of the Metaphysics of Morals* 4: 398.

[15] Williams (1981, 17-18).

[16] Williams presentation of the objection contrasts the impartial demands of morality with partial and agent-relative commitments which he thinks of as essentially non-moral yet constitutive of a person's allegiance to life itself, including moral aspects of life (1981, 18). Williams held the very task of articulating a theory of rightness to be misconceived. However, the 'one thought too many' objection to Kantian moral theory has an application beyond Williams's anti-theory stance.

[17] Michael Slote calls this effect a doubling back to the world. He writes that "one's inward gaze effectively "doubles back" on the world and allows one … to take facts about the world into account in one's attempt to determine what is morally acceptable or best to do" (2001, 39).

[18] Nevertheless, Slote's main example of the practical efficacy of agent-based ethics involves the case of a women reflecting on whether allow heroic surgery for her gravely ill mother. Slote sets out the relevant deliberations in terms of the women's reflecting on

what kinds of actions would best express her care and argues that although a truly benevolent or caring person *need* not reflect on the matter this way, "there is nothing unusual or inappropriate about ...[this]… as an expression of moral problem-solving" (2001, 41).

[19] The case may be different with self-regarding virtues like integrity, but even in the case of integrity it is not at all clear that a person manifests integrity by explicitly directing their deliberative concern at their own integrity.

[20] See Railton (1984) for an example of such a theory.

[21] The nearest theory in the vicinity appears to take the form of a theory that defines right action in terms of properties such that, were moral agents to act deliberately to bring them about, they would be acting virtuously. An action would then be right if and only if it instantiates these properties or causes their instantiation. Plausible versions of such a theory – versions that involve essentially other-directed virtues – would not be agent-based theories, because the properties agents seek to instantiate when they act virtuously would not be an aretaic properties of themselves.