JASON KAWALL

VIRTUE THEORY AND IDEAL OBSERVERS

(Received 27 April 2000; received in revised version 19 April 2002)

ABSTRACT. Virtue theorists in ethics often embrace the following characterization of right action: An action is right iff a virtuous agent would perform that action in like circumstances. Zagzebski offers a parallel virtue-based account of epistemically justified belief. Such proposals are severely flawed because virtuous agents in adverse circumstances, or through lack of knowledge can perform poorly. I propose an alternative virtue-based account according to which an action is right (a belief is justified) for an agent in a given situation iff an unimpaired, fully-informed virtuous observer would deem the action to be right (the belief to be justified).

Virtue theorists in ethics, while primarily united simply in their rejection of familiar deontological and consequentialist theories, do share some common positive ground.[1] A standard position among virtue theorists is that what constitutes a morally right action is to be derived (in some way) from the behaviour of virtuous agents. Similarly, in Linda Zagzebski's recent virtue-based approach in epistemology, justified belief and knowledge are understood in terms of the behaviour of epistemically virtuous agents.[2]

In what follows I argue that the relation between right action (or justified belief) and virtuous agents espoused by many virtue theorists is severely flawed. However, I also show that a related position, making use of the judgements of virtuous idealized observers, remains true to the virtue theorists' insights, but is not subject to the difficulties which beset the standard virtue theory approach. I first focus on virtue approaches in ethics; I turn to virtue theory in epistemology in section IV, and consider objections to the alternative approach in the final section of the paper.

Philosophical Studies **109:** 197–222, 2002.

I

Virtue theorists in ethics generally hold that judgements about good character or virtue are primary or basic, in the sense that other moral notions such as right action, or good states of affairs can be explained or defined in terms of virtues or character. We can characterize a right action as that action (or one of a set of equally acceptable actions) which an agent morally ought to perform, all things considered. An account of right action in terms of virtuous agents which is commonly held by virtue theorists was given by Rosalind Hursthouse in a 1991 paper:

P.1. An action is right iff it is what a virtuous agent would do in the circumstances.[3]

While Hursthouse has since modified her position,[4] her original account of right action in terms of what a virtuous agent would do in like circumstances is not only common to many virtue theories, but is sometimes held to be an essential feature of all virtue theories. Consider the following passage from Justin Oakley, in a recent survey article on virtue theory:

> I shall outline *six* claims which appear to be essential features of *any* virtue ethics view. The first, and perhaps best-known claim, which is central to any form of virtue ethics, is the following:
>
> > An action is right if and only if it is what an agent with a virtuous character would do in the circumstances.[5]

It is this standard virtue theory account of rightness (SVAR), supposedly essential to virtue ethics, which will be my target in what follows.

The underlying flaw in such accounts of right action (in terms of what a virtuous agent would do in like circumstances) is that there can be unfavourable circumstances in which virtuous agents would make moral mistakes. Robert Louden presents an example of a truly virtuous agent performing what is intuitively a morally wrong action:

> There are cases in which a man's choice is grounded in the best possible information, his motives honourable and his action not at all out of character. And yet his best laid plans may go sour [. . .] My point is that virtue ethics is in danger of blinding itself to the wrongful conduct in Oedipal acts [i.e. wrong acts performed

by good agents], simply because it views the Oedipuses of the world as honorable persons and because its focus is on long term character manifestations rather than discrete acts.[6]

Oedipus is presumably a virtuous agent, and thus on the standard virtue theorist's account of right action, his actions (such as sleeping with his mother) are morally right. He does what a virtuous agent (Oedipus himself!) would do in those very circumstances. Intuitively, however, this seems quite wrong. Oedipus severely regrets his actions, which suggests that, at the very least, a virtuous agent would not consider all of Oedipus' actions right.

Louden focuses on a case in which "a man's choice is grounded in the best possible information, his motives honourable and his action not at all out of character",[7] but the problem can be extended. Consider a case in which a moral saint is given a large dose of an hallucinogen, or a drug which causes extreme aggression. Surely many or most of his actions would not be morally right when he is under such influences, but it seems that the standard virtue ethicists' account of right action cannot provide us with an explanation of this fact. On the SVAR, if a saint would turn to fistfighting in such circumstances, then it is morally right to fistfight in such circumstances. If a saint, while hallucinating, would try to encourage a child to fly by tossing the child out of a tall building, then it is morally right to toss a child out of a tall building to encourage him to fly (while one is hallucinating).

Note that there are further problems with the SVAR when we focus on the condition that an action can be right *only if* it is such that a virtuous agent would perform it in like circumstances. Suppose, plausibly enough, that no virtuous agents are drug addicts with an extremely high tolerance for hallucinogens. Consider a case in which an agent is given a large dose of an hallucinogen, and then confronted with a burning house in which there is a child who could easily be rescued. We could well imagine that an addict in these conditions could act to save the child (given his high tolerance for the drug, he is not strongly affected), while no virtuous agents would be capable of this. Under the SVAR, acting to save the child would not be morally right because no virtuous agent would do so under such circumstances.[8]

II

What could a defender of the SVAR say in response to such charges? First, she could perhaps emphasize that in the unusual situations presented above, there is a sense in which the actions of the virtuous agents would be morally justified. It is hard to find Oedipus blameworthy for his actions – he was not culpably ignorant, etc. He was doing the best he could, within his limitations. Similarly, if a virtuous agent were injected with an hallucinogen, we cannot expect her to act appropriately, etc. Her faculties are not working properly, and she can hardly be blamed for not having a higher tolerance to an hallucinogen.

Still, while there may be a sense in which these actions are morally justified, it does not seem that such actions are *right*. We are tempted to say that Oedipus is not morally blameworthy, but this is a far cry from declaring his actions to be right. A basic tenet of commonsense morality is that a person can accidentally or unintentionally perform wrong actions. If you give a person a pill which you justifiably believe to be aspirin, but which is in fact arsenic, your action seems to be best described as being perhaps subjectively morally justified, or non-blameworthy.[9] But it is certainly not a right action.

The moral justification of actions like those of the virtuous agent on hallucinogens is quite minimal, at best. Consider the case where a saint, due to the effects of a drug, attempts (out of benevolence) to teach a child to fly by throwing her from a tall building. Obviously such an action is not right. Furthermore, it seems quite possible that, given the effects of the drug, the saint would not even be concerned with morality. If so, the saint's action does not seem to be terribly justified – it is not a case in which a person is trying to do the right thing, but fails.[10] At best we might say that the saint would not be morally culpable, due to the influence of the drug.

What of Oedipus? Perhaps his actions could be seen as morally right, on a more subjective reading of 'right'. He does the best he can, in accordance with his moral standards. But even here there are problems. Consider a virtuous agent who looks back on an action she has performed, but now with more complete information, and in calmer circumstances. Surely she could conclude that she did the wrong thing, though she was justified at the time in the action

she performed, given her good intentions but poor information. The same considerations apply in the case of Oedipus. Even if we can identify some form of moral justification which would accrue to the actions which would be performed by virtuous agents, virtue theories would be left without an account of what constitutes a right (or wrong) action. The SVAR would need to be abandoned as an account of right action, and a new account developed. Thus, this line of response on behalf of the SVAR fails.

Hursthouse suggests an alternative line of response in the following brief remark: "It [the SVAR] also intentionally allows that in some circumstances – those into which no virtuous agent could have got herself – no action is right".[11] Thus, perhaps the SVAR theorist could maintain that virtuous agents would not allow themselves to be injected with various drugs, and so on. In cases in which an agent is injected with such drugs, there is no right action, because there is no action which a virtuous agent would do (she would have avoided the situation entirely).

As it stands, however, this response is quite implausible. While a virtuous agent may wish to avoid situations in which she cannot function as she normally would, it seems terribly unlikely that virtuous agents will never find themselves forced into such situations. Consider a dining companion surreptitiously placing an hallucinogen into a virtuous agent's drink. Such a scenario seems quite possible, and also quite out of the control of the virtuous agent. It is hard to think of any situations into which a virtuous agent could not be forced – through various forms of bad luck, drugs, or what-have-you. Thus, the suggestion fails as a defense of the SVAR. Virtuous agents can be forced into the sorts of undesirable situations described in section I, and thus be led to perform what we would standardly take to be morally wrong actions.

A further response on behalf of the SVAR extends what I take to be the basic insight which motivates the previous suggestion from Hursthouse. Rather than holding that virtuous agents simply would not become involved in certain situations, we could instead hold that in such situations they would not be acting in character. Gandhi under the influence of a strong hallucinogen is not acting as the virtuous agent Gandhi. Thus, his actions would not be those of a

virtuous agent, and we would not need to hold that his actions would be morally right.

One could argue against this that even in the unfortunate situations described in section II, the virtuous agents are acting in character. There are ways in which a virtuous agent will respond to another person's unjustified anger towards her, there are ways in which she will respond to an emergency when there is adrenaline in her system, and (similarly) there are ways she will respond when she is under the influence of a drug. These latter responses form a part of who she is (her character), just as much as her behaviour under any other circumstances. If so, it seems that we again arrive at the actions of hallucinating virtuous agents constituting right actions.

Still, a defender of the SVAR may have a plausible rejoinder available. Consider: there is, presumably, a way in which a virtuous agent will behave when thrown into an active volcano, or when placed on the surface of Mercury, and so on. But the agent's virtuous character does not seem to be involved in such situations. With this in mind, the defender of the SVAR could claim that a virtuous agent's being injected with a hallucinogen is akin to a virtuous agent being placed on Mercury: there is some way that the agent will behave, but the agent's virtues are not involved (nor should we expect them to be) under such circumstances. Indeed, this is precisely the sort of modification Hursthouse has made in more recent work. She now proposes the following:

> An action is right iff it is what a virtuous agent would characteristically (i.e. acting in character) do in the circumstances.[12]

We can refer to this modified account as the SVAR′. Under the SVAR′ a virtuous agent's behaviour under extreme conditions will not constitute right action because the virtuous agent's virtues do not function under such conditions, and thus the agent is not acting in character.

Under what conditions are we to say that a person is not acting in character? We can consider three proposals. First, it could be that the situation is so unusual and rare that one's reactions in the case cannot be seen as reflecting one's character. It is not clear why this should be the case. What if an agent were to happen upon a village in which an army commander will kill twenty people unless

the agent kills an intelligent unicorn? This would, of course, be an extremely unusual situation. But why wouldn't the agent's virtues be involved in her decision-making? More generally, it seems that conditions merely being unusual or rare would not suffice to show that an agent's character is not involved in the given situation.

Second, it could be that the situations at hand are not such as to involve moral virtues and standard decision-making. While this might be a good description of an agent placed on the surface of Mercury, it doesn't seem to be true of the sorts of cases under consideration here. Here agents are making decisions, in situations with moral significance. Note also that in any case, we can be acting in character in situations, which do not call for the use of moral virtues. A person can be acting in character by making a bad pun, or quickly coming to the solution of a problem in algebra. There is more to our character than our moral virtues and vices.

Third, it could be that the agent is impaired in such a way that her character cannot come into play. This is perhaps the most plausible account of why an agent's character is not exhibited in a given situation. For example, a person who does not drink could be given several glasses of whiskey – and thus the normally clear-thinking, reserved, careful individual winds up dancing (none too gracefully) on a pool table. The original objection to this general line of thought again seems relevant. Why shouldn't we see an agent's behaviour under the influence of drugs or who is otherwise impaired as constituting or reflecting part of her character? I must admit that at this time I don't see a non-question-begging way of answering this question either way. I will thus grant for the sake of argument that there can be cases of the sort that the defender of the SVAR′ would espouse – cases in which an agent is impaired in such a way that her character is not truly reflected in her behaviour. Still, I will argue that even allowing this, the SVAR′ faces serious difficulties.

Before turning to these difficulties, we can introduce a proposal similar to the SVAR′, one suggested by Michael Slote:

> A view can be agent-based and still not treat actions as right or admirable simply because they are done by a virtuous individual or by someone with an admirable or good inner state.
>
> [. . .] a benevolent agent is typically *capable* of choosing many actions that *fail to express or exhibit* her benevolence. [. . .] Thus if one is benevolent and sees an individual who needs one's help, one may help and, in doing so, exhibit one's

benevolence. But it is also presumably within one's power to refuse help, and if one does, then one's actions won't exhibit benevolence.[13]

Given that the actions of a virtuous agent may not always express the agent's character, Slote holds that

> [. . .] Acts therefore do not count as admirable or virtuous for an agent-based theory of the sort just roughly introduced merely because they are or would be done by someone who in fact is admirable or possessed of admirable motivation; acts have to exhibit, express, or further such motivation, or be such that they *would* exhibit, express, or further such motivation if they occurred, in order to qualify as admirable or virtuous.
>
> [. . .] In order to avoid wrongdoing, one must (on agent-based theories of the sort just mentioned) avoid actions that exhibit bad or deficient inner motives.[14]

The key intuition that Slote draws upon is that an action is not right or admirable merely because a virtuous agent performs it; rather, the virtuous agent must be properly motivated in her action in the particular situation.

Still, neither the SVAR′ nor Slote's proposal constitutes an adequate response to the problems raised for the original SVAR. First, they do nothing to show how Oedipus' actions are not morally right, given that Oedipus is acting in character when he performs the actions. More generally, it seems highly unlikely that we can redescribe all cases in which a virtuous agent apparently performs a wrong action as a case in which the agent is acting out of character, or from a poor motive. For example, many tragic actions involve virtuous agents acting in character and from a virtuous motive, but also (tragically) lacking some key piece of information. Thus, even if the SVAR′ or Slote's proposal eliminates actions which are out of character for a virtuous agent, crucial counterexamples (tragic actions) remain.

Second, if the defender of the SVAR′ or Slote's proposal claims that a virtuous agent on hallucinogens is acting out of character (or cannot be virtuously motivated), then he cannot account for the rightness of such actions as an addict's saving a child from a burning building. Surely the addict (who has taken a dose of an hallucinogen which would incapacitate any virtuous agent) performs a right action in these circumstances. But the SVAR′ would be unable to account for this. Even if some virtuous agent on hallucinogens were to stagger into the building to rescue the child, the agent would

be considered to be acting out of character, and thus his behaviour could not constitute a right action. On the other hand, if virtuous agents do act in character (and thus have virtuous motivations) while on hallucinogens then a saint's well-intended throwing of a child from a building would be morally right. The defender of the SVAR′ or Slote's proposal is thus forced into a vicious dilemma.

Third, what if a virtuous agent were given a drug which greatly expanded her mental faculties, far beyond their normal capacity – indeed, so expanded that no ordinary human could approach such mental acuity? Imagine this agent placed in a morally demanding, complex situation, and an ordinary virtuous agent placed in similar circumstances. With only her ordinary, non-drug enhanced faculties the ordinary agent is unable to foresee a number of terrible consequences which will result from the action she chooses to perform. On the other hand, the agent with the drug-enhanced faculties is aware of these consequences, and is able to find a far superior alternative action. The SVAR′ would force us to hold that the former action is morally right, while the latter, intuitively superior action is at best morally neutral, and quite possibly morally wrong. Why? The agent with the drug-enhanced mind is acting out of character – she has no such mental acuity without the drug.[15] Thus, the ordinary agent's action while in character (without the drug) is right, while the action which would be endorsed by the agent with superior, drug-enhanced faculties would not be considered right. Given that no other virtuous human would be able to see the alternative action as better (as they lack the requisite mental faculties), the action could well turn out to be morally wrong, according to the SVAR′ – no virtuous agent acting in character would perform the action. This cannot be correct. Surely the judgements of a person with superior faculties are better grounds for what is morally right than those of an average person who cannot understand the full consequences of an action (given a similarly morally virtuous nature). I take it therefore, that the SVAR (or SVAR′) still faces severe difficulties.

While this last problem does not apply to Slote's proposal, his emphasis on having appropriate inner motivations ('expressed' in an action) creates difficulties of its own. It classifies any action, no matter how inept or poorly performed, as morally right, as long

as it would be performed by a properly motivated virtuous agent and thus 'expresses' the virtue. While we might hold that virtuous agents must generally be skillful in their actions, Slote's proposal doesn't require an agent to be acting in character. The agent's skill need not be in evidence for an action to be considered right. Thus, for example, a virtuous, but temporarily confused doctor who benevolently gives a patient an improper treatment is considered to have performed a right action on Slote's proposal, to the extent that the treatment exhibits the doctor's benevolence. We could even stipulate that the treatment, while improper, is skillfully performed (e.g. the doctor skillfully performs an unnecessary and potentially dangerous surgery). But again, this seems at best a morally justifiable or excusable action. We do not deem such inept and potentially damaging actions morally right; Slote's proposal cannot be correct.

A variant of the SVAR′-Slote response is to hold that the SVAR must be restricted to normal conditions. Thus:

> SVAR*: Under normal conditions, an action is right iff it is what a virtuous agent would do in like circumstances.

We can presumably eliminate cases in which virtuous agents are on drugs, or in extremely unusual situations as these are not normal circumstances. The proposal is akin to secondary-quality analyses of colours ("An object is red iff it appears red to normal observers under normal circumstances").

However, the SVAR* does not fare better than its predecessors. First, while Oedipus' particular situation is perhaps outside the range covered by the normal conditions clause, his general situation is common enough – being in a situation in which we lack relevant information. It is highly implausible to hold that all such cases must be abnormal or deviant. Thus, a wide range of intuitively wrong actions would still be declared right on the SVAR*. Second, what are we to say of cases in which conditions are not normal? We would require an additional account of rightness for such circumstances, and it seems preferable to have account which works for all circumstances.[16] Alternatively, the proponent of the SVAR* could maintain that in unusual circumstances there simply are no right actions. However, this would make the account implausibly gappy – particularly if we restrict the range of normal conditions

significantly in an effort to avoid the counterexamples which plague the SVAR. Finally, the dilemma raised above for the SVAR′-Slote response also applies to the SVAR*. If being under the influence of (large quantities of) alcohol or hallucinogens are deviant conditions, we cannot account for the rightness of actions by addicts, while if they are not deviant, we cannot account for the wrongness of well-intentioned but misguided actions by virtuous agents.

III

If we abandon the SVAR and its variants, must we also abandon hope of an explanation of right action in terms of virtuous agents? I believe there is a much better alternative position available to virtue theorists. The motivation for the position becomes clear when we examine the source of the difficulties for the SVAR.

There seem to be two key problems at the base of the counter-examples to the SVAR. First, in many cases it seems that virtuous agents can non-culpably lack crucial, morally relevant information. This plays a key role in the case of Oedipus. Humans have limited belief-forming faculties, and are incapable of foreseeing all of the consequences of their actions, and so on. Even if virtuous agents intend well, they can perform actions which are morally wrong through ignorance.

A second set of difficulties arises in cases in which an agent's faculties would be adversely affected when immersed in a given situation. On the SVAR we appeal to what virtuous agents would do in given circumstances to determine what constitutes right action. Problems arise because some situations will involve hallucinogens, brainwashing, and so forth. The virtuous agent's behaviour will often be adversely affected by such factors, and thus not a suitable guide to right behaviour.

In light of these difficulties we can see what an adequate account of rightness requires. First, the virtuous agents to whom we appeal must have all relevant information about the situation. When agents lack crucial information they can make mistakes. Second, we need to appeal to a well-placed observer of the situation, not an agent who is immersed in it. Thus, we arrive at an alternative virtue-based account of rightness:

IOAR: An action is right for an agent in a given set of circumstances iff an unimpaired, fully-informed virtuous observer would deem the action to be right.

We can see that this "ideal observer account of rightness" (IOAR) avoids the difficulties which plague the SVAR. Oedipus' actions were not right because a fully-informed, unimpaired virtuous (hence 'ideal') observer would not approve of the actions. Oedipus himself, when fully-informed, disapproved of his actions. The actions of a saint on hallucinogens would not be considered right, assuming that a fully-informed, unimpaired virtuous observer would not approve of the actions. An unimpaired observer of the situation would not be required to take an hallucinogen. Such an observer could judge an addict's saving of a child as morally right, even if this observer would be unable to do this were she placed in similar circumstances. I will consider objections to the IOAR in section V. For the moment, I hope only to have shown that there is a *prima facie* plausible alternative account of rightness available to virtue theorists.

IV

We can now turn to a recent virtue-based approach to epistemology – an approach which naturally leads to an idealized observer theory. Linda Zagzebski's characterizations of basic epistemic notions are grounded in the beliefs and behaviours of virtuous people. I will focus on her definitions of justified and unjustified beliefs, which parallel the SVAR account of morally right action:[17]

A *justified belief* is what a person who is motivated by intellectual virtue, and who has the understanding of his cognitive situation a virtuous person would have, might believe in like circumstances.

An *unjustified belief* is what a person who is motivated by intellectual virtue, and who has the understanding of his cognitive situation a virtuous person would have, would not believe in like circumstances.[18]

To begin, it is worth noting that Zagzebski seems not to capture *justified* or *unjustified* beliefs, but rather *justifiable* (or *unjustifiable*) beliefs. Imagine an intellectually virtuous agent drinking tea and listening to the radio. She might form a belief that it is going to snow tomorrow, based on the forecast of a respected meteorologist

she hears on the radio. According to Zagzebski such a belief is justified. Consider then a person who, in the same circumstances, also forms a belief that it will snow tomorrow, but on the basis of the pattern of tealeaves left in his cup. Surely this belief is not justified, yet it would be considered so, given Zagzebski's characterization of justified belief. It is justified to the extent that it is a belief that a virtuous agent in like circumstances might form.

Zagzebski states that "The definition of a justified belief is exactly parallel to the definition of a right act".[19] Herein lies the source of the above problem. It is commonly accepted that we can perform right acts by luck, accident, or for inappropriate reasons. Zagzebski herself notes that

> we will need to distinguish between the evaluation of the act or belief and the evaluation of a person *for* doing an act or having a belief. Such a distinction may seem peculiar on some ways of looking at the psychology of action, but ethicists often have no trouble speaking of an act in abstraction from the agent of the act. So if a person does the just thing out of a motive of gain, a common response is to say that the act itself is right, but the agent is not praiseworthy for doing it; he gets no moral credit for it.[20]

A lunatic might try to poison a city by putting fluoride in the city water supply, a person might go to a restaurant and by chance keep an appointment for lunch with a friend, a shopowner might not cheat her customers in order to maximize her profits, and so on. These are standard examples of right actions that reflect no credit upon the agents who perform them.

However, the forming and sustaining of justified beliefs is incompatible with most forms of luck, accident, or inappropriate reasons. In the example given, the agent uses a method (reading tealeaves) which is entirely unreliable and it is pure luck that the output of the method in this case matches the output of reliable methods. Such a belief is not justified. Epistemic justification is not simply a matter of the content of a belief; how the belief is formed and sustained are crucial factors. The alleged parallel between justified belief and right action is incorrect. However, a stronger parallel could be drawn between epistemically justified beliefs and morally justified actions. A morally justified action is one which an agent performs for good moral reasons (an objective notion), or for what she takes to be good moral reasons (a subjective notion). The motivations and processes

which lead to the action are relevant to its status as justified or unjustified.[21]

Similarly, right actions and justified beliefs are not analogous in the way supposed by Zagzebski. A stronger analogy here would be drawn between right actions and true beliefs; neither is guaranteed to obtain simply as a result of a given agent's motivations or attitudes. We can accidentally form true beliefs or perform right actions, regardless of our motivations or justification. We can perform morally justified actions that are not right, and we can form epistemically justified beliefs that are not true. At best, we might say that beliefs formed on the basis of tealeaves, etc., are *justifiable* beliefs, to the extent that there are reliable methods (available to the agent) which would have led to beliefs with the same content under the same circumstances.

An initial step in improving Zagzebski's account would be to modify it in the following manner:

> A *justified belief* is what a person who is motivated by intellectual virtue, and who has the understanding of his cognitive situation a virtuous person would have, might believe in like circumstances, and is a belief which is formed in the same manner as the belief which would be formed by a virtuous person.

With the final clause in place, the modified account would not be satisfied by beliefs which merely share the content of beliefs which would be formed by virtuous persons. Rather, beliefs would need to be formed in an appropriate manner (i.e. in the manner of an intellectually virtuous agent) in order to be considered justified.

We can now turn to a second difficulty, one that leads us towards an idealized, virtuous observer account of justification. Consider a situation in which an epistemically virtuous agent is given a large dose of an hallucinogen. She tries to be careful, etc., in forming her visual beliefs (and thus is intellectually virtuously motivated), but still comes to believe that there are faces moving in her shower curtain and that the walls around her are bending. Such beliefs would be characterized as justified, given Zagzebski's definition of justified belief, but this seems quite implausible. At best, such beliefs would simply be ones commonly formed under such circumstances. Surely we'd expect very few, if any beliefs, to be justified under such dreadful conditions.

The source of the difficulty lies in the fact that in dreadful circumstances even virtuous people can do things which are quite unjustified. This, of course, parallels one of the difficulties raised earlier for the SVAR. A simple step towards improving Zagzebski's position would be to consider the judgements of virtuous *observers* about the beliefs in question, rather than considering the beliefs of virtuous agents immersed in a similar situation. A well-placed, unimpaired observer of a situation will not be subject to the detrimental factors which may influence agents in the situation under consideration. An outside observer can assess the beliefs of an agent undergoing brainwashing without herself being brainwashed.

A related point concerns the understanding a virtuous person might have of various unfortunate cognitive situations. A virtuous agent's understanding of her cognitive situation when she has just received a severe head injury could be quite defective. It seems plausible to hold that most (perhaps all?) of her beliefs in such a situation are quite unjustified. Yet her beliefs would still be justified on Zagzebski's account of justified beliefs, assuming she is virtuously motivated.

Notice that this difficulty arises regardless of whether Zagzebski intends to capture an objective notion or a subjective notion of epistemic justification.[22] Imagine that the head injury causes the agent to think that she is in a situation in which the Gambler's Fallacy is a good method of reasoning. She then carefully forms beliefs using the fallacy. In such a case the agent's beliefs are not objectively justified.[23] This is not captured by Zagzebski's definition as the agent is virtuously motivated and has the understanding (however flawed) of her cognitive situation that an epistemically virtuous agent would have.

Next, imagine that the agent has incorporated modus ponens into her epistemic standards prior to the head injury. Thus, if she forms beliefs in accordance with this rule of reasoning, these beliefs will be subjectively justified for her (in addition to being objectively justified). Following her head injury the agent attempts to carefully form beliefs in accordance with modus ponens, but due to her disorientation she becomes confused about which proposition is the consequent, etc. Thus the beliefs she forms are not in accord with the rule. Here, it seems that her beliefs won't even be subjec-

tively justified (given that they do not accord with her subjective epistemic standards) due to her poor grasp of her cognitive situation. But again, such beliefs would be considered justified on Zagzebski's account as she has the same understanding of her cognitive situation that a virtuous agent would have in this situation.[24]

These cases point to the need for a better understanding of the agent's cognitive situation than that which a virtuous agent would have in similar circumstances, if we are attempting to determine whether the agent's beliefs are justified in any sense. Just as virtuous agents can act (or form beliefs) quite inappropriately in some situations, they can also have a very poor understanding of their cognitive situation. The difficulty with Zagzebski's proposal arises as agents can lack important information about their cognitive situations, information which is highly relevant to the justification of beliefs formed in such situations. Given this, they will not be sufficiently reliable guides as to what beliefs or actions are in fact justified (or even justifiable) in these situations.

Thus, Zagzebski's account of justified belief runs afoul of difficulties parallel to those which afflict the SVAR. The behaviour of epistemically virtuous agents can be severely impaired – by drugs, injury, and so on. Furthermore, such agents may also lack crucial information about their cognitive situations – and the absence of such information can lead to poorly-formed beliefs. We arrive at an alternative account of justified belief:

> IOAJ: A belief is justified for an agent in a given situation iff a fully-informed, unimpaired, intellectually virtuous observer would deem the belief justified.

This "idealized observer account of justification" parallels the account given previously of moral rightness. With this in hand, we can now consider certain questions which arise in assessing the IOAR and IOAJ.

V

(i) *What does 'fully-informed' amount to?* I've spoken of fully-informed virtuous observers, but how much information is this? Further, we risk circularity in our account of justified belief if a fully-informed observer requires justified beliefs.

Response: It seems that an ideal observer must have true beliefs concerning any facts relevant to appraising a given action or belief. If an observer lacks relevant information her judgement could be quite flawed. Similarly, an appraisal based on falsehoods would surely be quite suspect. What determines relevance? Here we can appeal to an account of relevance modified from one developed by Thomas Carson:[25]

> A proposition *x* is relevant to a judgement about *y* if and only if either (1) believing that *x* would make a difference to an ideal observer's reaction to *y*, or (2) *x* is a member of a group of propositions *G* such that believing *G* would make a difference in an ideal observer's reaction's to *y*, and there is no subset of *G* the believing of which would have exactly the same effect on his reactions to *y*.

Intuitively, a proposition is relevant to judging a situation iff having a belief about it (or about a group of propositions of which it is a member) would have an impact upon an ideal observer's reaction to the situation. A fully-informed observer is one who has true beliefs about all propositions relevant to the situation being appraised (and no false beliefs about relevant propositions).

(ii) *What does 'unimpaired' amount to?* We can appeal here to a somewhat extended sense of 'unimpaired'. In order for an observer to be unimpaired in this extended sense, the observer must have adequate time to fully consider the situation, be free from distractions, be free from the influence of hallucinogens, depressants, etc. (any of which could have a detrimental impact upon the observer's judgements), as well as being free from coercion, personal interests in individuals involved in the situation being assessed, and so on. For example, we would want to exclude cases in which an observer has been threatened that 10000 people will be killed if she deems a certain action to be morally right. Clearly, we would worry in such cases that the observer's judgements would be altered and adversely affected as the observer's judgements would not be based on the situation under consideration, but rather on the external factors involved. We would not have the observer's genuine, unimpaired appraisal of the situation. Intuitively, we want to consider the judgements of virtuous observers who are focused on the situation at hand (without conflicts with other interests), and who are

without any influences which will bias or impair the observer in her deliberations.[26]

(iii) *Is this still a virtue theory?* It is often suggested that one of the key improvements of virtue theories over deontological or consequentialist theories is that they work with real, embodied agents, and look to their actions in particular situations, rather than fumbling with clumsy, abstract rules. Agents immersed in a situation will have a vivid understanding of the situation which cannot be captured by broad rules; the agents will have relationships, feelings, commitments, etc. which are particular and unique, and not adequately accounted for under rival moral theories. A worry for the current proposal is that this sensitivity to particularity is lost if we shift to virtuous observers of a situation, rather than agents directly embedded within a situation.

In response, note that it is crucial that the virtuous observers of the situation be both virtuous and fully-informed. Being truly fully-informed will require observers to know the relationships of the agents involved in a situation, the felt quality of pain that an agent will suffer if a certain action is performed (not just knowledge that this agent will suffer in some way if the given action occurs), the depths of the personal commitments of the persons involved, and so on. Thus, with such full information, our virtuous observers will be extremely sensitive to the particulars of different situations. Indeed, as we have seen, a fully-informed observer could well have crucial, particular information that an agent immersed in a situation might lack. A virtuous observer can still be sensitive to the friendships, personal commitments, etc. held by the agents in the situation being assessed. There is no reason to think these would be dismissed by a sensitive, virtuous observer.

In addition to possessing full information, the ideal observers must be virtuous. The virtues themselves require, as essential components, a range of concerns and emotional dispositions. An honest person will have a concern for the truth, and will tend to look favourably upon those who tell the truth in standard circumstances. Lawrence Blum characterizes the virtue of compassion as follows:

> Compassion is not a simple feeling-state but a complex emotional attitude toward another, characteristically involving imaginative dwelling on the condition of the

other person, an active regard for his good, a view of him as a fellow human being, and emotional responses of a certain degree of intensity.[27]

Other virtues will require additional emotional dispositions. Thus, our ideal observers are not mere centres of computation; they must also possess complex emotional attitudes. In this they are like any other virtuous agents, and insofar as the current approach relies essentially on the judgements of observers with the virtues, it remains a virtue theory. The possession of the virtues by the observers, and the associated emotional capacities is what allows them to make use of their full information in forming their judgements.

What of cases in which the judgements of the embedded agent differ from those of a virtuous observer? Why prefer the judgements of the observer to the actual, embedded agent with her natural reactions? The guiding intuition is that we don't want to blindly follow the behaviours of an impaired or ill-informed saint. We are concerned with their judgements – their best judgements. And recall – virtuous agents can look back on their own actions and see them as well-intentioned, but morally wrong, when they are able to fully reflect on their actions.

The judgements of virtuous agents in most situations reflect their best judgements – they have enough information, and are unimpaired. In turn, their actions are a reflection of their best judgement. Thus, it is tempting in standard cases to appeal to what a virtuous agent would do in the circumstances – the actions reflect their best judgement. But this only holds true insofar as we are getting proper judgements from them. When circumstances are such that the virtuous agent is impaired, acting out of character, or ill-informed, we can no longer trust her judgements or actions. Thus, the current theory need not be seen as a radical revision of virtue theories. It is simply a theory that holds that under ordinary circumstances we can see virtuous agents' actions as reflecting their best judgement, and so we might be tempted to appeal directly to what they would do in given circumstances; we basically have a suitable observer immersed in the situation. But – the key claim is that what is really doing the moral work here is the judgements of fully-informed, unimpaired virtuous observers.

(iv) *On what basis do the virtuous ideal observers judge?* It might appear that IOAR and IOAJ face a trilemma of devastating circularity, vacuousness, or a crucial reliance on a deontological or consequentialist framework. After all, we might wonder on what basis the virtuous ideal observers might decide whether or not to deem an action morally right (or a belief justified). One possibility is that they try to figure out whether ideal virtuous observers would deem the action to be right. But then we have reached an empty circularity – they *are* the observers at stake. They would be told to judge as they would judge. On the other hand, if they are not appealing to this, it might appear that there is nothing else to appeal to, and thus they would be left with no guidance as they attempt to determine the status of a given action or belief. Finally, we might propose that the virtuous ideal observers could appeal to an independent set of deontological principles, or a consequentialist theory, and judge the action on the basis of this theory. However, if this is the case, then we no longer have a virtue theory, and have simply arrived at an underlying deontological or consequentialist theory that is doing all of the work.

This is a problem for all pure virtue theories. Consider the SVAR. If we say that actions are morally right iff they are what would be performed by a virtuous agent in the circumstances, we can ask: on what basis would these virtuous agents decide to act? Would they try to determine what virtuous agents would do in the circumstances? But then they are themselves these agents, and they are given no guidance. It might seem that they must appeal to further deontological or consequentialist theories. And if this is the case, it is hard to see that we have anything worth calling a virtue theory, as distinct from these others.

The solution to the problem lies in recognizing that the agents with the virtues will *ipso facto* be concerned with certain sorts of outcomes and modes of behaviour. A person who is honest will have a concern for the truth; a person with the virtue of gratitude will have a concern for her benefactors and their actions, and so on. Gary Watson suggests that

> To be sure, a concern for outcomes will be internal to certain virtues. For instance, the benevolent person will be concerned that others fare well. But the moral significance of this concern stems from the fact that it is part of a virtue, not

from the fact that misery and well-being are intrinsically or ultimately bad and good respectively. To put it another way, it will follow from an ethics of virtue that virtuous people care about things (and outcomes) for their own sakes (as final ends in themselves). There is no further commitment, however, to the idea that these concerns are virtuous ones because their objects are inherently valuable or desirable for their own sakes.[28]

Thus, the virtues will lead us to value certain outcomes, etc., but the value of the outcomes is derivative from the virtues. Virtuous ideal observers then, need not appeal to consequentialist or deontological theories. Their very possession of the virtues will lead them to value (or disvalue) various outcomes and modes of behaviour. We then appeal to the overall judgements of fully-informed, unimpaired virtuous agents to determine the normative status of actions and beliefs; for example, they will deem a certain action or range of actions available to a given agent in a given situation to be right – what the agent ought to do in the circumstances. Note that in deeming an action to be right, an ideal observer (or anyone else) does not *mean* that it satisfies IOAR (after all, this would again lead us to circularity or a regress).

Virtuous ideal observers need not appeal to the IOAR or IOAJ as decision-procedures. That is, they need not look to the IOAR or IOAJ for guidance in determining their attitudes towards actions or beliefs (etc.). Rather, it is the virtues themselves that will shape their attitudes. As David Solomon puts it, "within an EV [ethics of virtue] it is not the theory of the virtues which is supposed to be primarily action guiding, but rather the virtues themselves."[29]

We might compare the IOAR to a sophisticated psychological theory of vision. Ideal observers who possess the virtues need not appeal to the IOAR to guide their actions, just as persons with good visual systems need not appeal to a theory of vision in order to see well. It is the virtues or visual systems themselves which guide these agents, not the theories which are built upon the behaviour of these agents. The IOAR and IOAJ don't provide us the virtues, or with wisdom. Rather, they tell us that what is morally right will be determined by those who are unimpaired, fully-informed, and virtuous.

(v) *What are the virtues?* This question goes rather beyond the scope of the present paper, but does point to an issue, which (ulti-

mately) must be addressed. Note that providing an account of the virtues is a project which would need to be undertaken by defenders of the SVAR and Zagzebski's account of justified belief also – it is not a project peculiar to the IOAR or IOAJ. Indeed, it is a project shared by all virtue theorists.

For the moment, note that we can appeal to any of the accounts of the virtues which virtue theorists already accept. Thus, we could take the virtues to be those traits essential to lead a flourishing human life – a well-known position espoused by Aristotle, Foot, Hursthouse, Wallace, and others. Other alternatives are also available.[30]

One might object that we cannot sensibly attribute human traits or virtues to the ideal observers, given their potentially very different capacities, etc.[31] But the objection is ill-founded. Consider – sensitive hearing may be an essential trait for a cat to lead a flourishing feline life, but we can still sensibly attribute sensitive hearing to other animals, even those which are otherwise quite different from cats. Similarly, there is nothing to prevent an ideal observer from possessing virtues which are essential to human flourishing (regardless of whether these same traits are in any sense essential to the ideal observer).

On the other hand, consider the following standard objection to ideal observer theories, as presented by Jonathan Harrison:

> If we could discover what the reactions of an ideal observer to any kind of action were, and it turned out that his reaction was wholly different from the one we expected it to be, and that this difference between his reaction and ours was not due to ignorance of matters of fact or a mistake about matters of fact I do not think we would just accept, without argument, that an action we had always thought was right was wrong. [. . .] we are not prepared to say that since a standard or ideal person would have the reaction of approval, that settles the matter; we would want to be shown that he was right to feel approval of the action in question.[32]

The objection is telling against traditional ideal observer theories such as those proposed by Roderick Firth or Charles Taliaferro.[33] On such theories ideal observers are characterized as omniscient, vividly aware of the facts, and impartial. Such an ideal observer could still be mildly sadistic, hateful, etc.[34] As Harrison suggests, it is implausible to hold that we should simply allow the judgements of ideal observers who could have such flaws to constitute moral rightness, etc.

Note that by attributing the virtues to ideal observers, we need not see the reactions of these observers as flawed, irrelevant or distant from us. If an ideal observer disagrees with us (and not on the basis of false beliefs, or a lack of knowledge on our part) we still have reason to accept the IO's judgement because she embodies the virtues.[35] The IO's judgment can be seen as superior to our own insofar as she is benevolent, compassionate, unbiased, and so on – we need only recall our human flaws and limits. Insofar as we value the virtues, we should also value the judgements of virtuous ideal observers.

Thus, we see that there is good reason for a virtue theorist to abandon the standard virtue account of rightness or justified belief in favour of the virtuous ideal observer approach presented here. This alternative account still grounds moral and epistemic norms in the judgements of virtuous agents, but avoids the problems caused by adverse conditions and impaired faculties. In addition, attributing virtues to ideal observers in this way strengthens traditional ideal observer theories by providing the ideal observers with a basis for their judgements, which in turn gives us (humans) reason to abide by these judgements. The two approaches are natural complements.

ACKNOWLEDGEMENTS

I am very grateful to the many people who have provided me with feedback on this paper, including: Dan Brock, Teresa Celada, James Dreier, Patrick Durning, Jeremy Fantl, Jennifer Lackey, Thidwick Lorignal, Baron Reed, Ernest Sosa, John Stone, and Jim Taggart. Versions of this paper have been read at the 1999 Canadian Philosophical Association meetings, the University of Louisville, and York University. I would like to thank these audiences, and Wayne Fenske, my commentator at the CPA, for their helpful input. Finally, many thanks to an anonymous referee for several insightful comments.

NOTES

[1] See Gregory Trianosky, 'What is Virtue Ethics All About?,' *American Philosophical Quarterly* 27 (1990), pp. 335–344, for a summary of the common negative ground. Justin Oakley's 'Varieties of Virtue Ethics,' *Ratio* 9 (1996), pp. 128–152, is a recent survey focusing on shared positive claims made by virtue theorists.

[2] See Linda Zagzebski, *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge* (New York: Cambridge University Press, 1996).

[3] Rosalind Hursthouse, 'Virtue Theory and Abortion', *Philosophy and Public Affairs* 20 (1991), p. 225. She presents the same account in her 'Normative Virtue Ethics', in Roger Crisp (ed.), *How Should One Live?* (New York: Oxford University Press, 1996), pp. 19–36.

[4] See Rosalind Hursthouse, *On Virtue Ethics* (New York: Oxford University Press, 1999), p. 28. I will discuss her revised account in section II.

[5] Oakley, p. 129. For further examples of this sort of virtue-based account of rightness, see Zagzebski, p. 235; Michael Slote, 'Agent-Based Virtue Ethics', in French, Uehling, and Wettstein (eds.), *Moral Concepts*, Midwest Studies in Philosophy, vol. XX (Notre Dame: University of Notre Dame Press, 1996). Virtue ethics is also characterized in this way, though not endorsed, by James Dreier in his 'Structures of Normative Theories', *The Monist* 76 (1993), p. 34, and by Gary Watson, 'On the Primacy of Character', in Owen Flanagan and Amélie Oskenberg Rorty (eds.), *Identity, Character, and Morality* (Cambridge, MA: MIT Press, 1990), pp. 459–461.

[6] Robert Louden, 'On Some Vices of Virtue Ethics', in Kruschwitz and Roberts (eds.), *The Virtues: Contemporary Essays on Moral Character* (Belmont: Wadsworth, 1987), pp. 71–72.

[7] Ibid., p. 71.

[8] Compare Owen Flanagan's insight that oftentimes it is a vice which allows an agent to perform morally right actions. See his 'Admirable Immorality and Admirable Imperfection', *Journal of Philosophy* 83 (1986), pp. 41–60.

[9] Note that the term 'morally justified' is often used in a rather different sense, to refer to an action that is morally right, all things considered. For example, we might say that an action is justified from the moral point of view (as opposed to the aesthetic, or pragmatic point of view, etc.). In this paper I use the term, broadly, to refer to an action for which the agent has some reason to believe that it is morally right or permissible. It is because such an agent has such reasons and justification that she can be seen as non-blameworthy (in typical cases), even if her action is not morally right. In this usage of the term I follow such authors as Alvin I. Goldman, 'The Internalist Conception of Justification', in French, Uehling, and Wettstein (eds.), *Studies in Epistemology*, Midwest Studies in Philosophy, vol. V (Minneapolis: University of Minnesota Press, 1980), pp. 27–51; Richard Feldman, 'Subjective and Objective Justification in Ethics and Epistemology', *The Monist* 68 (1985), pp. 407–419; James Montmarquet, *Epistemic Virtue and*

Doxastic Responsibility (Lanham: Rowman and Littlefield, 1993); and Jonathan Dancy, *Moral Reasons* (Cambridge: Blackwell, 1993), p. 113.

[10] Of course, I do not mean to suggest that a saint in such circumstances could not be benevolently motivated. I wish only to draw attention to the possibility of a saint ignoring moral concerns due to the influence of a drug.

[11] Hursthouse, 'Virtue Theory and Abortion', p. 225.

[12] Hursthouse, *On Virtue Ethics*, p. 28.

[13] Slote, p. 86. When Slote speaks of an "agent-based theory" he is referring to virtue theories such as the SVAR, where the moral status of actions is entirely derivative from evaluations of character or virtue.

[14] Ibid., pp. 86–87.

[15] Note that Slote's response would be immune to this sort of case, as it requires only that the virtuous agent be properly motivated and that the act express this motivation – it does not require that the agent act in character.

[16] I hope to show in the following sections that such a unified account is available to virtue theorists.

[17] There is much to say about her characterizations of other epistemic concepts; however I will not consider these in the present paper.

[18] Zagzebski, p. 241.

[19] Ibid., p. 241.

[20] Ibid., p. 235.

[21] A morally justified action need not be a right action. For example, an agent may perform a terrible action, but one which seemed to her morally appropriate, given the limited information she had. Here we could say her action was morally justified (at least subjectively), but that she did not perform a right action. Similarly, a right action need not be a morally justified action. The examples presented of right actions performed by accident or by luck could all be described as cases of unjustified morally right actions.

[22] Given that she views justified beliefs as parallel to right actions, it seems that Zagzebski intends an objective notion of justification.

[23] I assume, of course, that the Gambler's Fallacy is not a reliable vehicle for producing true beliefs.

[24] The beliefs may be justified in the sense that the agent is virtuously motivated in forming them (a form of epistemic responsibility), but Zagzebski clearly wants more than this. After all, if she only wanted to capture this sense of justification her definitions of justified and unjustified belief could be greatly simplified – justification would be a matter of intellectually virtuous motivations alone.

[25] Thomas Carson, *The Status of Morality* (Dordrecht: D. Reidel, 1984), p. 58. Carson's original proposal is that

> A fact x is relevant to a judgement about y if and only if either (1) knowing x would make a difference to an ideal observer's reaction to y, or (2) x is a member of a group of facts G such that knowing G would make a difference in an ideal observer's reaction's to y, and there is no subset of G the knowledge of which would have exactly the same effect on his reactions to y.

[26] There is, of course, a certain lack of precision in this characterization of being unimpaired. Still, I believe that we have adequate intuitions here to guide us, even lacking a strict definition. (This is not to say that I believe that we can simply rest easy – ideally we should have a more rigorous characterization of 'unimpaired', but I hope that for the purposes of this paper, at least, an intuitive account is adequate.)

[27] Lawrence Blum, 'Compassion', in Amélie Oskenberg Rorty (ed.), *Explaining Emotions* (Berkeley: University of California Press, 1980), p. 509.

[28] Watson, p. 459. Philippa Foot makes a similar claim in her 'Utilitarianism and the Virtues', *Mind* 94 (1985), pp. 196–209. Reprinted in Samuel Scheffler (ed.), *Consequentialism and Its Critics* (New York: Oxford University Press, 1988), pp. 224–242.

[29] David Solomon, 'Internal Objections to Virtue Ethics', in French, Uehling, and Wettstein (eds.), *Moral Concepts*, p. 439.

[30] Elsewhere I argue that attempts to ground the virtues in terms of human flourishing are ultimately untenable. Still, the position serves adequately for expository purposes here.

[31] We might expect ideal observers to require certain abilities beyond the normal human range – especially in order to have access to all information relevant to judging a situation.

[32] Jonathan Harrison, *Our Knowledge of Right and Wrong* (London: George Allen & Unwin, 1971), pp. 178–179.

[33] See Roderick Firth, 'Ethical Absolutism and the Ideal Observer', *Philosophy and Phenomenological Research* 12 (1952), pp. 317–345, Charles Taliaferro, 'The Environmental Ethics of the Ideal Observer', *Environmental Ethics* 10 (1988), pp. 233–250, and Charles Taliaferro, 'Relativising the Ideal Observer Theory', *Philosophy and Phenomenological Research* 49 (1988), pp. 125–126.

[34] Firth also requires an ideal observer to be (beyond omniscient, etc.) "an otherwise normal human being". Allowing that this is even possible, it is clear that a certain degree of malevolence, dishonesty, etc. can fall within the range of human normalcy. Indeed, Firth might have an additional problem to the extent that saint-like benevolence, etc. go far beyond what is normal for humans, and thus could not be traits of a Firthian ideal observer.

[35] Attributing virtues to ideal observers need not be circular. For example, if we hold that the virtues are those traits which are essential to leading a flourishing human life, there is clearly no circularity in then attributing these traits to ideal observers.

Department of Philosophy and Religion
University of Tennessee at Chattanooga
615 McCallie Ave.
Chattanooga, TN 37403
USA