

Zagzebski. Linda Trinkaus. Exemplarist Moral Theory.

Oxford: Oxford University Press. 2017. Pp. 274. \$69.00 (cloth).

What role should heroes, saints and sages play within moral theory? While it would be unfair to say that contemporary moral theory ignores such figures, it certainly does not grant them a central role in moral theorizing. In this ambitious and innovative book, Linda Zagzebski seeks to change that by developing a moral theory in which moral exemplars play a foundational role.

Zagzebski has five aims for her book. Her primary aim is to provide an original, comprehensive ethical theory based upon moral exemplars that can serve the same purposes as its rivals: consequentialism, deontology and virtue ethics. Second, Zagzebski seeks to provide a moral theory that is capable of morally improving people. The third aim is to provide an ethical theory that fits with the moral development of individuals and communities. Fourth, Zagzebski's aim is to construct a theory that gives empirical sciences, particularly psychology and neuroscience, a central role. Finally, the book aims to provide a theory that is capable of providing different versions for different communities that have different needs.

The key idea behind Zagzebski's theory is that a foundational moral theory can be based on direct reference to exemplars of moral goodness. Zagzebski appeals to the Direct Reference theory of semantics developed by Hilary Putnam and Saul Kripke to show that a moral theory can be built upon non-conceptual foundations. According to this theory of semantics, knowing the meaning of a word that picks out an object need not require possessing an adequate description of the object being referred to (as is required by Descriptivism). Knowing the meaning of a natural kind term such as 'water' or 'gold', for example, need not require possessing any such description. Instead, as Zagzebski explains (11), the Direct Reference Theory holds that

these terms refer to “whatever that is the same kind of thing as some indexically identified instance.” On this account, ‘Gold’ refers to anything that is the same element as *that*, where the demonstrative term ‘that’ refers directly to some object in the world.

This theory of reference opens up the possibility of basing morality upon non-conceptual foundations, as a speaker can refer successfully without possessing an adequate description of what they are referring to. Applying this to moral terms allows Zagzebski to say (14) that we should not seek to identify the referents of such terms through providing an adequate description of them. Rather, the semantics of such terms is fixed by direct reference (15). Zagzebski is not the first to have pursued this approach (See Richard Boyd, “How to Be a Moral Realist,” In Geoff Sayre McCord (Ed.) *Essays in Moral Realism* [Ithaca, NY: Cornell University Press], 181-228.) However, Zagzebski makes the original move of appealing to this semantic theory to justify a moral theory based upon moral exemplars.

In order for this theory to work though, we need some way of recognizing moral exemplars in a way that does not depend on already possessing a descriptive account of what such people would be like. Zagzebski claims that we identify moral exemplars directly through the emotion of admiration. Zagzebski claims that , “there is something in us that detects the excellent, and that is the emotion of admiration,” (2). It is through admiration that we recognize the admirable and identify the morally exemplary. This identification of exemplars through admiration is a recognizable feature of our moral practice and one that precedes any theoretical knowledge we have of morality. Moreover, we can be more confident in our knowledge of the admirable nature of these exemplars than we can be of any theoretical moral knowledge we possess. In Zagzebski’s words, “we are more certain that Confucius, Jesus, and Socrates are admirable than we are of claims about the good of pleasure, or what human flourishing is, or the good of one’s

duty, or any of the other claims that are used to ground a moral theory,” (10). Basing a theory on exemplars then provides a foundation for moral theory that is based on a trusted part of our moral practice.

In addition, admiration also provides us with the motivation to become morally better people. Zagzebski claims (35-40) that there is an important difference between admiration for “natural talents” (those that people are born with) and admiration for “acquired excellences” (those that people have had to work to develop). While we admire both kinds of excellence, the way in which the admiration is experienced is different in both. Admiration for acquired excellences brings with physical feelings of being uplifted and brings with it a desire to emulate the person we admire (43). Zagzebski takes moral and intellectual excellences to be an acquired rather than in-born. This means that when we admire someone’s moral or intellectual excellence we will have a desire to become like them.

Given this desire to become more like moral exemplars, how are we to go about achieving this? An interesting upshot of direct reference semantics is that empirical investigation can reveal natures. Zagzebski claims this as an advantage for her view. We can pick out who the moral exemplars are prior to having much of an understanding of the nature of those exemplars. This though does not give us a detailed insight into the psychological make-up of these exemplars. We can achieve this through empirical investigation of moral exemplars, together with detailed narratives and descriptions of their lives. This Zagzebski claims will allow us to discover facts about the nature of good people through empirical investigation (18). Having discovered these facts we will then be in a position to become more like the exemplars.

Zagzebski then goes on to argue that a moral theory based on moral exemplars can provide us with definitions of all of our basic moral terms. First, value terms such as ‘virtue’, ‘good motive’,

‘good end’, and ‘virtuous act’ can be defined in terms of the traits and motives that we admire in exemplars, the ends they act for and the acts they perform (105). Zagzebski claims that defining virtue terms in reference to exemplars has the advantage of helping us discover the psychological structure of the virtues and the ways in which different virtues can fit together (128).

Next, a good life is defined as one that would be desired by admirable people (159). Zagzebski claims that this account shows how an account of the good life can be grounded in the emotion of admiration (168) which, because it is accompanied by a desire to emulate, also provides us with a motivation to act morally (169).

Finally, Zagzebski explains how to define deontic terms. Zagzebski defines a duty as an act that exemplars who possess practical wisdom would “judge to be the only option,” feel guilty if they did not perform it and “blame others if others did not do it,” (196). A wrong act is one that such an exemplar would judge not to be an option, would feel guilty if she performed it and blame others who performed it (196). Finally, a right act is one that such an exemplar would characteristically take to be most favoured by the balance of reasons (201). Zagzebski claims several advantages for defining deontic terms in this way. First, these definitions allow for a greater chance of agreement about the class of duties and wrong acts than their competitors do (206). In order to reach agreement on which acts are duties all that is needed is that all of the moral exemplars identified by people in a society agree on which acts are duties. They do not even need to agree on the reasons why the acts are duties so long as they agree on how to respond to them. Second, Zagzebski claims that defining deontic terms in reference to exemplars, “can serve as a psychological force to (a) get agreement within a community about the identification of negative duties and any positive duties, and (b) motivate persons to act on their

duty,” (208). The reason for this is that these terms are defined in terms of moral exemplars who are capable of inspiring admiration rather than abstract moral concepts.

There is much to admire about this book. First, it is impressively ambitious. Many books in moral philosophy aim to provide a new articulation or defence of an existing moral theory. Much important and worthwhile work is done in this vein. However, it makes it particularly noteworthy when an author promises to provide an original and comprehensive ethical theory. Moreover, In addition to explaining and motivating her theory, Zagzebski also explains how this theory can illuminate issues in moral education (Ch.5) and Meta-ethics (Ch.8). Zagzebski deserves great credit for having the ambition to pursue this kind of project. Moreover, I think that Zagzebski succeeds in her primary goal of showing how a moral theory based on exemplars can serve as a viable alternative to existing moral theories. Another praiseworthy feature of Zagzebski’s theory is that it lays the foundations for an exciting, interdisciplinary research project into the nature of moral exemplars. Of course, consequentialists, deontologists and virtue ethicists can also motivate the study of moral exemplars but a theory with moral exemplars makes this study central to understanding morality. It is also worth noting that Zagzebski’s book is one that may serve as a source of moral inspiration, a rare virtue in a work of moral philosophy. By including detailed descriptions of the lives of moral exemplars, Zagzebski’s book may well inspire admiration of such figures. If her claims about the motivational powers of admiration are correct, then this may well inspire moral inspiration in some of her readers.

Nevertheless, I have a number of concerns about Zagzebski’s theory that I do not think are adequately addressed. First, it seems reasonable to wonder about the extent to which we are able to trust our judgements about who is and who is not a moral exemplar. History is full of morally reprehensible people who have been admired by many, and indeed it is not hard to think of

several such people who exist today. This creates a clear problem for Zagzebski's theory. Given that admiration for exemplars is foundational, it is far from clear how it can have the resources to differentiate this admiration from the admiration that is directed towards real moral exemplars. Moreover, the place of admiration at the foundation of this theory means that if our admiration is directed at the wrong target in this way then this places all of our ethical thinking in jeopardy. If we misidentify the moral exemplars then our view of what the virtues are, which acts are right and wrong and the nature of a good life will all be off-target.

Zagzebski does of course consider this problem. She responds in three ways. First, she claims that our admiration for someone will disappear if we discover new information about them that shows them not to be exemplars (46). Second, Zagzebski claims that we can trust our admiration for people when it survives reflection over time (47). Zagzebski suggests that if an admirer of Hitler who subjected her admiration to critical reflection may well lose this admiration. Even if this is not the case, Zagzebski takes comfort in the fact that, "hardly anybody reacts to the Nazi admiration for Hitler by saying that maybe the Nazis were right and we are wrong," (47). Finally, empirical investigation into common ways in which our admiration might be biased can help us critically reflect on our admiration.

These responses help but I don't think they fully answer the worry. The most obvious worry is that people do still admire fascists and racists of various kinds (even a cursory look at contemporary politics will tell us this much). Are all of these people simply failing to critically reflect on their admiration? It is comforting to think so but there are many thoughtful and reflective people who have admired racists and fascists in the past, so it is far from clear that we are justified in thinking this.

The deeper problem though is that admiration seems capable of making such critical reflection difficult. Think of the lengths that people go to in order to defend their favourite actor, sport star or politician from some incriminating allegation made about their moral character. In these cases it seems as if these feelings of admiration may lead to a desire to defend the person being admired which in turn can inhibit a critically reflective stance. Of course, this is something that could potentially be tested empirically. At the very least though, it seems worth questioning whether admiration will always play a positive epistemic role.

Another worry that might be raised against this theory is that in giving a central role to admiration it may marginalize other important moral emotions. Zagzebski certainly seems right to claim that admiration is an important moral emotion but so are guilt, shame, resentment and regret (and perhaps envy, pride, reverence and awe). These have an inferior status in Zagzebski's theory. There will times where it is appropriate to feel guilt. The way to identify this though is by looking at what a moral exemplar, someone identified through the emotion of admiration, would feel in that situation. There may also be times where it is appropriate to feel resentment, if this is involved in blaming someone. Again though, we can only find out if this is appropriate by examining whether a moral exemplar would feel resentment or endorse such feelings in this situation. So while other moral emotions do have a role in Zagzebski's moral theory, it is a subordinate one. This may strike some people as problematic, as it seems like the emotion of admiration is being granted a privileged epistemic position in this theory. To see why imagine that I feel resentment towards someone but judge that someone I morally admire would not feel resentment or endorse my feelings of resentment in this situation. It seems that on Zagzebski's theory, I should reject my feelings of resentment here, as they are not what a moral exemplar would feel. At the very least I should not blame the person for acting wrongly, as according to

Zagzebski's definition of wrong action (196) they will not have acted wrongly if a moral exemplar would not blame people who acted in that way. Accepting this means placing greater trust in my feelings of admiration than in my feelings of resentment. At this point it seems reasonable to ask whether the emotion of admiration is always more trustworthy than other moral emotions. To be fair to Zagzebski, she is not committed to saying that it will always be more trustworthy. Perhaps this kind of emotional conflict can form part of the process of critical reflection that may lead us to reject our feelings of admiration. Nevertheless, it would have been helpful to hear more about the role that other moral emotions are supposed to play in this theory. None of these issues though, changes the fact that this is an important and groundbreaking book that will be essential reading for anyone working on issues in moral philosophy that relate to moral exemplars. In addition, this book is highly recommended reading for anyone interested in moral philosophy, moral psychology and moral education.

Alfred Archer

Tilburg Institute for Logic, Ethics and Philosophy of Science (TiLPS), Tilburg University