Review essay

Citation for published version:

Digital Object Identifier (DOI):
10.1163/17455243-01606003

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Journal of Moral Philosophy

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Review Essay: Exemplarist Moral Theory

Lani Watson and Alan T. Wilson

Linda Zagzebski. Exemplarist Moral Theory


ISBN: 9780190655846 (hardcover)

Abstract: This review essay provides a critical discussion of Linda Zagzebski’s (2017) Exemplarist Moral Theory (EMT). We agree that EMT is a book of impressive scope that will be of interest to ethical theorists, as well as epistemologists, philosophers of language, and philosophers of religion. Throughout the critical discussion we argue that exemplarism faces a number of important challenges, firstly, in dealing with the fallibility of admiration, which plays a central role in the theoretical framework, and secondly, in serving as a practical guide for moral development. Despite this, we maintain that EMT points the way for significant future theoretical and empirical research into some of the most well-established questions in ethical theory.

Keywords: Exemplarism, Exemplarist Moral Theory, Zagzebski, admiration, virtue ethics, emulation, character education.
Introduction

Linda Zagzebski’s *Exemplarist Moral Theory (EMT)* has impressive scope. It offers much that will be of interest to ethicists of all stripes, as well as epistemologists, philosophers of language, and philosophers of religion. Moreover, *EMT* has plenty to offer for experimental psychologists in search of new directions for empirical research into character and virtue, as well as education theorists interested in moral and/or intellectual character education. This broad reach is no small achievement and constitutes a noteworthy feature of the book. In this respect, and several others, *EMT* provides a rich resource for theorising about some of the most well-established questions in ethical theory, whilst drawing on contemporary methods for the investigation of ethical practice. In this review essay, we offer a brief summary of the book followed by critical discussion of its merits and of the areas of future research that it is likely to inspire.

I. Summary

In Chapter 1 of *EMT*, Zagzebski sets out the key ideas behind the exemplarist approach. Her aim is to provide “a comprehensive ethical theory that serves the same purposes as deontological, consequentialist, and virtue theories” (p. 3). What sets exemplarism apart from the alternatives is its central claim that the foundational “point of origin” in ethics is found in exemplars of moral goodness (p. 10). All moral terms (including ‘right act’, ‘virtue’, and ‘good life’) are to be defined with reference to the exemplars that we admire. The rest of the book seeks to develop this central claim, and to explore and evaluate the definitions of other moral terms that it suggests.

The second and third chapters provide more detail on exemplars and how they are identified. An important role is assigned here to the emotion of admiration. It is through admiration that Zagzebski believes we can identify exemplars and be inspired to emulate them. Exemplars are those people who we admire upon reflection. Our experiences of admiration allow us to identify these exemplars, given that admiration tends to be a reliable emotion (unlike other emotions, such as “envy, jealousy, and feelings of vengeance” (pp. 45-46)). Zagzebski then provides some examples of the exemplars that she has in mind, dividing them into the categories of the hero, the saint, and the sage. Heroes, such as Leopold Socha and others who rescued people during the Holocaust, are most associated with the virtue of courage. Saints,
such as Jean Vanier and those who work within L’Arche communities, are most associated with the virtue of charity. And sages, such as Confucius, are most associated with wisdom.

Chapter 4 begins the process of demonstrating how moral terms can be defined with reference to the exemplars that we admire. One of Zagzebski’s “guiding principles” here is that “we are more certain of the identities of certain exemplars than we are of the terms to be defined” (p. 103). This is why it makes sense to begin with the exemplars themselves. This leads us to the definition of virtue as “a trait we admire in an exemplar. It is what makes persons like that admirable in a certain respect” (p. 105). And Zagzebski argues that reflecting on our exemplars can reveal more than this initial, basic, definition. Future empirical work focusing on exemplars could reveal deeper features of the virtues (such as whether they must be reliably successful character traits) and could perhaps even reveal additional virtues that have not typically been recognised by virtue theorists.

Zagzebski then turns to the practical issue of moral development. Chapter 5 summarises some empirical work used to show that experiencing admiration for another can make us more likely to perform altruistic actions ourselves. Zagzebski proposes that our moral development is encouraged by our admiration for moral exemplars. Admiration prompts us to imagine being like the exemplar, and over time this leads us to actually develop the imagined virtuous dispositions. The chapter concludes with a discussion (pp. 150-153) of whether emulating others is compatible with being an autonomous agent.

Chapters 6 and 7 return to the task of defining key moral terms. Chapter 6 focuses on a ‘good life’, arguing that the desirable ought to be defined with reference to the admirable. A desirable life is defined as “a life desired by an admirable person. It is a life desired by persons like that” (p. 159). Zagzebski argues that one positive feature of exemplarism is that reflection on the variety of possible exemplars highlights the variety of ways in which it is possible to lead a flourishing life. Chapter 7 moves on to deontic terms, focusing on the definitions of right and wrong acts. At this stage, we no longer refer to exemplars in general, but instead focus on those exemplars who possess phronesis. A wrong act is “an act that is blamed by persons of practical wisdom” (p. 196). The exemplarist definition of a right act is set out as: “A right act for A in some set of circumstances C is what the person with phronesis (persons like that) would characteristically take to be the most favored by the balance of reasons for A in circumstances C.” (p. 201)
The final chapter provides more information on the theories of direct reference and externalist semantics that serve as an inspiration for Zagzebski’s approach throughout. Focusing on work from Kripke, Putnam and Donnellan, Zagzebski explores the extent to which exemplarism is compatible with a realist meta-ethical viewpoint. One key question here is whether the extension of ‘good person’ (or exemplar) will change over time, given that the people who are admired plausibly do change over time. There is then a brief conclusion, recounting some key points and highlighting some possible avenues for future research. This includes the interesting suggestion that we could empirically investigate “anti-exemplars;” those who are appropriate targets of our contempt (pp. 232-233).

II. Critical Discussion

II. I. Exemplarism as a Comprehensive Ethical Theory

A key aim of EMT is to present exemplarism as a comprehensive ethical theory, capable of standing alongside deontological, consequentialist and virtue theories. Given this aim, it is worth asking whether exemplarism offers a distinct ethical theory and, if it does, whether that distinct theory is a promising alternative to the other, perhaps more familiar, theories that Zagzebski mentions.

There are two reasons for accepting exemplarism as a distinct ethical theory, both of which relate to the foundation of the theory. Firstly, exemplarism differs from other theories by selecting good people as the foundational “point of origin” for ethics, as opposed to alternatives such as positive consequences, a good will, human flourishing, or fundamentally virtuous character traits. Secondly, exemplarism also differs from other theories by opting for a foundation that is non-conceptual. The theory is not based on the concept of a good person, but rather on actual people – those people we admire upon reflection. As Zagzebski explains (pp. 9-10):

> I want to make the foundation of my theory something that most of us trust – the people we admire upon reflection. The structure is foundationalist, but instead of starting with a concept, the theory begins with exemplars of moral goodness identified directly by the emotion of admiration.

If we view exemplarism as an ethical theory with a distinct foundational structure, then it is understandable for Zagzebski to suggest that “it might be regarded as a kind of theory that
does not fit any traditional category” (p. 231). It is certainly possible to view EMT as offering a distinct form of ethical theory. For this reason alone, the book is worthy of attention from those interested in philosophical ethics.

However, we think that exemplarism, understood in this way, faces potential challenges. An initial worry for the approach is the fallibility of admiration. It is possible for admiration to be misplaced, even among those who take the time to reflect on the people they admire. This possibility creates problems for an approach on which ‘good person’ just means ‘person like that’, and where the scope of persons ‘like that’ is determined by those who we admire (p. 15). More generally, it creates problems for an approach that would define all moral terms by reference to the people we admire. If our admiration can be misplaced, and if our theory is built upon the people we admire, then our theory has the potential to go seriously awry.

The issue of whether admiration is reliable is one of the central questions that Zagzebski tackles in Chapter 2. In this chapter, Zagzebski acknowledges the possibility of misplaced admiration, and states that a process of conscientious reflection is sometimes necessary to uncover instances where we have gone wrong. That is why the theory is based only on those who we admire upon reflection. The process of conscientious reflection is described as follows (p. 45):

The best we can do to be confident an emotion is fitting is the same as the best we can do to be confident that a belief is true: we find that it survives reflection over time on our total set of psychic states when we are using them the best we can to make them fit their objects. That is what I mean by conscientious self-reflection.

Perhaps this will be sufficient to defend the idea that it is rational to trust our admiration upon reflection, or even that admiration upon reflection will be generally reliable. However, it does not show that such admiration will be infallible. Even in the most obvious of cases, Zagzebski is less than fully confident that conscientious reflection will uncover mistaken admiration: “I cannot say with certainty that self-reflective Nazis could have figured out that there was something wrong with Hitler... but that is my hypothesis.” (p. 47) If the process of conscientious reflection is not certain to rule out admiration for Hitler, then it seems unlikely to rule out admiration for people who, while less extreme than Hitler, still ought not to be accepted as moral exemplars.
This looks to be a problem for exemplarism, given the decision to base the theory on those people we admire upon reflection, as opposed to on the concept of admirability or on the concept of a good person. If the theory was based on the concept of admirability, then it could be accepted that it is possible for those who we admire (upon reflection) and those who are truly admirable to come apart. The possibility of admiration going astray would show only that those with mistaken admiration are liable to misinterpret or misapply the theory. Instead, exemplarism rejects this approach and bases the theory on the actual people that we admire upon reflection. If admiration upon reflection can pick out Hitlers or racists (as discussed in Chapter 8), or simply those who happen to be petty or vain, then exemplarism will be stuck with the implications of that admiration. This includes the implications for other moral terms, including the understandings of an admirable life and of a desirable life that the theory generates.

This brief discussion is not, of course, sufficient to show that exemplarism needs to be altered. Zagzebski accepts that the theory is not presented “in sufficient detail to be a fully developed alternative” (p. 3), and it is possible that further detail and development will allay these concerns. However, it will be important for those who would develop exemplarism to say more on this issue, and on the possibility of different communities (and different individuals) having distinct sets of people who they admire upon reflection. One potential area of development would be to build upon a comment that Zagzebski makes when discussing the possibility of a group of people who admire those who are “brutal, greedy, and envious” (p. 17). We are told that, “Fortunately, I think that such a situation cannot happen for beings like us who share the same nature, the same emotional dispositions, and at least roughly the same physical, psychological, and social needs.” (p. 17) Building on the idea that our shared nature rules out a consistent admiration for those who are vicious (and explaining the compatibility of this idea with the possibility of reflective admiration for racists, as discussed in Chapter 8) might be one way of attempting to develop a response to this issue.

One further way of developing the account of exemplarism’s distinct foundational structure would be to provide some more detailed comparisons between exemplarism and other theoretical approaches to ethics. Given the above comments regarding admiration, one important candidate for comparison with exemplarism would be Michael Slote’s agent-based virtue ethics, as set out in his *Morals from Motives* (2001). When discussing virtue theories in *EMT*, Zagzebski tends to focus on Aristotelian or eudaemonist accounts. And yet, much like
exemplarism, Slote’s form of virtue ethics has a foundational structure that is based on admirability. Slote tends to focus on fundamentally admirable motives rather than on examples on fundamentally admirable people. However, a detailed comparison of the two approaches (and of the extent to which they are vulnerable to the above challenge concerning the use of admiration as a foundation) would be one option for developing exemplarism and drawing out the distinctive features of the approach.

II. II. Exemplarism as a Practical Guide

A second stated aim from Zagzebski in *EMT* is to help make persons moral. Tellingly, she refers to this as a practical aim and comments throughout the book on the power of the theory to impact upon actual moral practice and enable the living of moral lives. In the book’s concluding paragraphs, she writes, “We know that being moral is more important than thinking well about morality, but I hope that exemplarist moral theory helps us with both” (p.234). Chapter 5, on emulation, is where the majority of the discussion regarding the impact of exemplarism on moral practice takes place. Zagzebski notes at the outset that the chapter is “devoted to the way exemplars can serve as a guide for moral training” (p.129) with a focus on “the learning of virtue by imitation” (p.129). She goes on to build an argument in support of the claim that the emulation of exemplars can lead a person to act virtuously. This includes an extended discussion of moral reasons, addresses the worry that emulation is incompatible with autonomy, and results in a ‘model of moral learning’. Zagzebski writes, “The model I am proposing starts with admiration of an exemplar, which leads to an imaginative ideal of oneself, which in turn produces emulation of the exemplar’s motives and acts” (p.138). The argument for this model draws on an impressive range of theoretical and empirical resources. It is not, however, made clear how this model can or should be applied, what it would look like as a form of pedagogy, or in what contexts it could effectively impact upon moral practice. In short, the chapter does much to convince the reader that exemplars can serve as a guide for moral training, but little to explain or demonstrate how this can or should be harnessed for the purposes of moral education or guidance.

This observation extends to the book as a whole. Zagzebski says that “a moral theory is not primarily a manual for decision-making” (p.6). However, she does maintain that exemplarism can help make persons moral and this is referred to as a practical aim. By way of elucidation, Zagzebski introduces the analogy of a map in Chapter 1, returning to this at points throughout
the book. Zagzebski uses this analogy in order to account for the idealised nature of any moral theory intended to “simplify, systematize, and justify our moral beliefs and practices” (p.5). A map leaves out many features that exist in the terrain that it covers in order to simplify that terrain and allow us to navigate more easily. Any moral theory as extensive as exemplarism must do the same. Building on this, Zagzebski suggests that, as a comprehensive moral theory, exemplarism should be thought of on the model of a map of the world, as opposed to a city street map. The distinction here is between maps “that help us get around” and those that are “not detailed enough to help us get from place to place, and…not intended to do so” (p.6). Exemplarism is akin to the latter, according to Zagzebski. Indeed, she comments that “many maps do not have a practical purpose” (p.6) and exemplarism should be considered as analogous with these. This makes the stated practical aim of EMT harder to understand. It is not a ‘manual for decision-making’ nor will it help us to navigate the moral terrain that it covers. This raises the question of to what extent EMT succeeds in meeting its practical aim. Zagzebski contends that exemplarism allows us to “connect the map with actual moral practice” (p.8) via exemplars. Perhaps, as with a map of the world, this helps us to better understand the existing terrain, or existing moral practice. It is not yet obvious that it will help us to be more moral.

Further to this, Zagzebski contends that ‘the biggest advantage’ of exemplarism is that “It allows us to map value terms without relying upon a conceptual foundation” (p.128). As noted, this constitutes a novel contribution of EMT to ethical theorizing. Again, however, this raises a question about the extent to which this core advantage of the theory is related to the practical aim of making persons moral. Rather, it seems that the core advantage of exemplarism would lie in the success of the theory in accurately describing moral practice, as opposed to guiding it. This is no small achievement but at times throughout the book it appears as though Zagzebski conflates exemplarism’s ability to describe, explain, or account for moral practice, with an ability to guide or otherwise influence it. It is not made clear how the former is intended to achieve the latter. Zagzebski is optimistic about the prospects of exemplarism to impact upon moral practice but further evidence, both theoretical and empirical, is needed in order to substantiate this optimism. One significant benefit of EMT is that it points the direction for this future theoretical and empirical work, something that we will discuss in more detail below. Nonetheless, as it stands it is not clear how EMT achieves the stated practical aim of making persons moral.
A final and somewhat related worry concerns another of Zagzebski’s aims: to give a central place to empirical work. To be clear, there is little doubt that Zagzebski achieves this aim in the book. She draws on empirical studies in almost every chapter, and highlights a number of key studies as providing empirical support for the theoretical claims of EMT. In Chapter 2, for example, she cites empirical work on admiration in psychology to support the claim that the emotion of admiration can be trusted. In Chapter 3, she discusses a number of empirical studies, including laboratory experiments, with a view to advocating a multi-disciplinary approach to the observation of exemplars. This multi-disciplinary approach, with an emphasis on empirical research, is a distinctive feature of EMT and is in many ways a good one. Zagzebski regards it as such: “The exemplarist approach has the advantage that substantive matters about what makes a person good need not be settled at the outset” (p.18). Rather, she contends, “Careful observation tells us what the psychological structure of a person is, just as careful observation tells us what the physical structure of water is” (p.18).

The worry with this appeal to empirical work in EMT is that it is arguably overly optimistic. The idea that the psychological structure of a person can be discovered in the same way as the physical structure of water requires one to regard the methods and objects of study as relevantly similar in a way that is not immediately obvious. At the very least, we believe that the analogy between the structural make-up of water and that of persons (and the claim that they are similarly amenable to empirical study) warrants further scrutiny, which it does not receive in the book.

Leaving the analogy to one side, our general worry is not so much about the reliance on empirical work per se but about the extent to which EMT relies upon the feasibility of empirical work that has not yet been conducted. Zagzebski is arguably overly optimistic about the prospects for this work, given the notoriously elusive subject-matter of human character and psychology. She writes, for example;

“Maybe admirable persons do not have anything important in common other than being admired by us!...Perhaps there is not a single essence of good personhood, but there is a set of interesting, yet distinct ways in which a person can be admirable. It is also possible that the set of exemplars gradually changes over time” (p.19)

Zagzebski claims that “these are all testable hypotheses” (p.19). Likewise, she suggests that it will be possible to use empirical studies to settle various other questions. This includes
whether admiring natural talents “feels different” from admiring acquired excellences (Chapter 2), and whether being a sage requires above average “natural” intelligence whereas being a saint or a hero does not (Chapter 2). Along similar lines, Zagzebski maintains that the constituents of exemplarity can be discovered by empirical study (Chapter 3), that the role of the success condition for the virtues can be confirmed by empirical study (Chapter 4), and that the motivations for exemplary actions can be discerned by empirical study (Chapter 5).

In each of these cases we would argue that it is unclear whether or not empirical research will be able to provide the answers, or even what the relevant empirical studies would look like. In certain respects, of course, this is itself an empirical question and represents a Catch 22 for Zagzebski. She is keen to advocate for an approach to ethical theorizing that draws on empirical research in an area where limited empirical research has so far been done. This requires her to place a degree of confidence in the results of future empirical work. Such confidence is not in and of itself misplaced. However, Zagzebski is arguably overly optimistic about the extent to which future empirical work will be able to confirm or disconfirm key aspects of the theory. That being said, a key benefit of EMT lies in its pointing us towards future empirical work and clarifications of the theory that will need to be undertaken. In this way, concerns about places where the book is lacking in practical guidance may be outweighed by its potential for inspiring and directing new work within and across a range of disciplines.

III. Future Directions

One point that is worth re-emphasising is the sheer scope and ambition of EMT. In addition to skilfully engaging with ethical theory from across the theoretical spectrum, Zagzebski draws centrally on the theory of direct reference in the philosophy of language, applying its distinctive externalist semantics to the moral domain (see esp. Chapter 8). Alongside this, Zagzebski cites a wide range of empirical work throughout the book and this too plays a central role in the formation of the theory. Lastly, and perhaps most exceptionally, Zagzebski assigns a central role in EMT to biographical narratives, providing a rich resource for the reader and helping to fill out many of the theoretical details (see esp. Chapter 3). Zagzebski herself writes that she intends to “integrate psychological studies, narratives from literature and history, neurological information, and personal experience in a multi-disciplinary approach” (p.70). This successful merging of distinct methods and domains into one
theoretical framework is a rare achievement and speaks to the intended inclusivity of *EMT*.

Moreover, by presenting an inclusive approach to ethical theorising, Zagzebski also goes some way towards achieving another of her stated aims, namely, to provide a theory that crosses cultural boundaries. Ethical theories and practices arising anywhere in the world and throughout history can and should be incorporated into exemplarism, according to this wide-ranging approach. By way of demonstration, Zagzebski selects exemplars from different cultural contexts for the biographical narratives around which much of the theory is built and suggests that different classes of exemplarity may be more or less central for different ages and cultures: “Ancient China is incomprehensible without the sage, and Ancient Greece makes no sense without the hero. Arguably the saint arose within Christianity” (p.97). This inclusivity extends to Zagzebski’s discussion of ‘a (not the) good life’ and she acknowledges that “There is more than one way to flourish,” (p.171) given that ‘good life’ is defined by direct reference to the lives lived by exemplars. Zagzebski also endorses the idea that there are non-trivial differences between the exemplars that different cultures admire. In general, the approach to cultural comparison and engagement in *EMT* is well captured by Zagzebski’s comment in Chapter 7 (p. 211):

> a focussed discussion between the exemplars of different cultures on their areas of agreement and disagreement can lead to an expansion of the areas of agreement… it is a mistake to think that the way to get an agreement is to change our moral vocabulary into an alleged culture-neutral language within which we agree on certain judgments.

The explicit inclusivity of *EMT*, both in terms of its theoretical scope across philosophical sub-disciplines and advocacy of culturally diverse reference points and exemplars provides a potentially valuable model for future ethical theorising.

Zagzebski’s approach is also notable for the important role that is assigned to emotions. In addition to the fundamental role that is assigned to admiration, Zagzebski also discusses other emotions, including contempt, sympathy, compassion, and indignation. The contention that “It is hard to imagine a functional moral life without trust in these emotions” (p.46) places emotions at the heart of the exemplarist picture. This is particularly evident in the discussion of moral development and, crucially, in Zagzebski’s compelling account of how admiration can potentially be corrupted and lead us astray. In Chapter 2, Zagzebski offers an account of the psychological path from admiration, through envy and spite, to the Nietzschean
phenomenon of ‘ressentiment’ which amounts to a “pervasive rejection of the admirable, particularly the morally admirable” (p.58). This account is both intuitive and well-grounded in theoretical work, particularly on envy, which Zagzebski draws on throughout the discussion. It is to be hoped that EMT will provide a basis for future work on the role that emotions can play in our moral development and on the ways in which our emotions may sometimes be exploited so as to hinder that development.

The critical discussion offered above also suggests a number of clear directions for future research. Firstly, there is the future work to be done in developing exemplarism itself as a comprehensive ethical theory. This will involve providing further detail on the foundational structure of the approach, and that project is likely to benefit from detailed comparisons between exemplarism and alternative ethical approaches. Philosophical ethicists may be particularly keen to compare the merits of the exemplarist account of right action with other, more familiar, accounts. And the development of exemplarism as a comprehensive theory may also benefit from further work on the nature of admiration, given the central role that is assigned to that emotion. As Zagzebski suggests, research into the moral emotions represents a promising line of empirical inquiry in moral psychology, and EMT may provide further impetus for such a project.

Secondly, while Zagzebski’s practical aim for EMT is not, we argue, fully realised in the book, the prospects for practical applications of the theory are nonetheless exciting. This is particularly so in the case of moral and/or intellectual character education where EMT provides the basis for a programme of research into the cultivation of moral and/or intellectual character through the emulation of exemplars. In particular, research into the use of narratives for moral learning and the fostering of moral emotions in education are two interesting lines of inquiry suggested by the theory. In addition, empirical studies on the role that imitation plays in learning more generally could be informed and guided by the exemplarist framework. Finally, exemplarism requires us to identify certain people in our lives, communities, and histories that can serve as exemplars. Further theoretical and empirical research into the role of exemplars in our moral lives would therefore help to substantiate the theory and may provide an impetus for the creation of historical, biographical, and literary accounts of exemplars in the public domain.

With these suggestions for future research inspired by the theory in mind, it is, we think,
evident that *EMT* is an important work of ethical theory that will appeal to readers across a broad spectrum of philosophical and multi-disciplinary interests.

Lani Watson, University of Edinburgh, lani.watson@ed.ac.uk

Alan T. Wilson, University of Bristol, alan.wilson@bristol.ac.uk

**Acknowledgements:** Lani Watson’s work on this review essay was supported by a Leverhulme Early Career Fellowship (grant number: R44476). Alan T. Wilson’s work on this review essay was supported by the Templeton Religion Trust. The views expressed here do not necessarily reflect the views of the Templeton Religion Trust.