



Article

# Moral exemplars in theory and practice

Theory and Research in Education

11(2) 193–206

© The Author(s) 2013

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1477878513485177

tre.sagepub.com



**Linda Zagzebski**

University of Oklahoma, USA

## Abstract

In this article I outline an original form of ethical theory that I call exemplarist virtue theory. The theory is intended to serve the philosophical purposes of a comprehensive moral theory, but it is also intended to serve the practical purpose of moral education by structuring the theory around a motivating emotion – the emotion of admiration. In this theory, basic moral concepts are defined via direct reference to exemplars of moral goodness, picked out through reflective admiration. The theory gives narratives a critical function, and it connects empirical studies with the a priori side of ethics.

## Keywords

Admiration, direct reference, exemplars, imitation, moral theory, narratives

## 1. Introduction

For most of philosophical history, moral philosophers have not done a very good job of creating theories that serve philosophical purposes, but which can also be used to make people moral. Narratives are useful for the purposes of moral education and improvement because they engage our motives much more than abstract theories, and narratives are crucial to shaping our vision of a good life, but it is hard to see how narratives can be woven into a theoretical framework. Modern life makes another demand on us that is very difficult to satisfy: we need to find a public moral discourse that allows us to talk across cultures, but which is sensitive to the beliefs of individual communities, including faith communities. In addition, a number of philosophers have argued recently that moral philosophy should integrate empirical work, or at least it should be responsive to such work. My position is that even though a significant part of moral theorizing is a priori, it is an advantage of a theory if it has a natural relation to empirical studies.

---

## Corresponding author:

Linda Zagzebski, Department of Philosophy, University of Oklahoma, Norman, OK 73019-2006, USA.

Email: lzagzebski@ou.edu

I have been working on a form of virtue theory that I call exemplarist virtue theory.<sup>1</sup> The theory has four aims:

1. to serve the philosophical purpose of producing a comprehensive ethical theory,
2. to serve the practical purpose of moral education by structuring the theory around a motivating emotion,
3. to link the a priori side of ethics with empirical research in psychology and neuroscience,
4. to contribute to the perennial effort to find a moral discourse that has thinner versions for cross-cultural discourse, as well as thicker versions for use within particular communities.

Given the topic of this symposium, I will pay particular attention to the first two aims. The idea is that there is a way to construct an abstract moral theory that is adequate for purely theoretical purposes, but which contains within the structure of the theory a moral motivator: the emotion of admiration. The foundation of the theory is something that attracts a person to moral improvement, and which permits a variety of forms for communities that arise out of different moral traditions. So my answer to the question ‘Can morality be taught?’ will be affirmative, but my focus will be on a way of conceptualizing the connection between moral theory and moral learning that makes the theoretical side of ethics derivative from a part of our nature that drives us to be moral. Because I will be proposing a theory that contains something non-conceptual as an intrinsic part, I will need to start by presenting my theory of theory – what I think a theory is supposed to do.

## **2. My theory of moral theory**

I think of a moral theory as an abstract structure that aims to simplify, systematize, and justify our moral beliefs and practices. Creating theories is part of the practice of morality. Since one of the aims of a moral theory is to simplify, it will leave out many subtleties and complexities in the practice of morality. There is nothing wrong with that as long as we are not under the illusion that the features of moral practice left out of the theory disappear. We are simply not attending to them when we are engaged in theory building and discussion. They will reappear when we engage in some other part of the practice. But we would not construct theories unless we thought that there is something to be gained by simplifying. Given the limitations of the human mind, we are not able to understand a domain taken as a whole unless we ignore part of the domain we want to understand. The bigger and more complex the domain, the more we have to leave out if we want to understand it. Morality is an enormous domain that involves almost every aspect of human life and, to some extent, non-human life. It is not surprising that we cannot get our minds around it without mentally stripping away much of interest in the practice of morality.

I think this is a general point about understanding that applies even to the understanding of something as simple as the layout of a city. If every feature of the city was on the city map, the map would be as complex in its layout as the city is, and the map would not help us understand it. So the map leaves out many things, and it may also distort some

things, such as the shape and relative size of Canada or Russia on a two-dimensional map. The map can be misleading, but a two-dimensional map is often more useful than a globe, even with the distortion. The distortion does no harm as long as we are aware of it.

Similarly, it is more useful to conceptualize moral reality without certain things in it, but it is helpful to keep in mind that we made a choice to leave those things out and the result might be a distortion. For instance, most moral theorists believe that a good moral theory leaves out the identities of the persons in the practice. The reason is that we get impartiality in our way of understanding morality if we do not mention who is who. Since I am going to make the unusual move of proposing a kind of moral theory that identifies certain individuals, I think it is worth thinking about the fact that we usually make the choice to leave out the identities of persons in a moral theory, and the choice is made for a reason. But as long as we are aware of the reason for the choice, we might decide that it is not always an advantage to make that choice.

I think it is important that a moral theory is not primarily a manual for decision-making, and it is not constructed to be a manual. Again, a moral theory can be compared to a map. A detailed street map will help us get around a city, but many maps do not have a practical purpose. When we look at maps in books about foreign countries, or when we see maps of battlefields on television or in a book of history, our aim is not to get anywhere. A map of the world is not detailed enough to help us get from place to place, and it is not intended to do so. I think of a comprehensive moral theory as like a map of the world. Theories of parts of morality are closer to street maps. A theory of just war, or a theory of friendship, or a theory of duties to future generations, is closer to a street map. But I think that even a street map is not constructed with the sole purpose of guiding a person from place to place. If your primary purpose was to get from one point in a city to another, you might not use a map at all. A GPS navigation system might be a more efficient tool for getting around. But a navigation system cannot give you the understanding of the layout of the city that you get from a map. Similarly, if our main purpose was to get guidance in moral decision-making, we would want a manual, not a theory.<sup>2</sup> But the manual would not give us understanding of the domain of morality as a whole.

I think, then, that moral theory aims primarily at explaining and justifying moral beliefs and practices, and only secondarily at telling us what to do in any given situation. There are many elements in our moral practices that pre-exist theory and which we seek to understand. These include reactive emotions such as admiration, praise, blame, remorse, indignation, and horror; practices of punishment; rules such as the Golden Rule or the Ten Commandments; and values such as freedom, fulfillment, and social cohesiveness. Some of these elements are reflected in narratives that are cherished by a community and passed down from generation to generation. Of course, there are many other elements of our moral practices that pre-exist theory. My point is that a moral theory is about something that already exists, and we seek to understand it through a moral theory.

What we seek to understand can be altered by the process of seeking to understand it. In this respect a moral theory is unlike the map of a city. I suppose we can imagine a map that we liked so much that when the map and the layout of the streets did not coincide, we changed the streets, not the map. But assuming that we do not want to move the streets around, the point of a map is to give us understanding of the physical layout of a

city that is already there and that will change for reasons that have nothing to do with the map. Moral theory is different because even though there are moral practices that are already there in advance of the creation of theory, since one of the purposes of the theory is to justify the practices, we might find out that some element of the practice is not easily justified if it is related to other elements of the practice as given in the map. That could lead us to change the practice in response to the theory. In contrast, a city map does not seek to justify the layout of a city, only to depict it.

I have said that a theory is not a manual, but it is an advantage if a theory can connect to central features of our moral practices in a plausible way. It is particularly advantageous if it can link up with narratives since narratives are the primary vehicle for the moral education of the young, and the primary way humans of any age develop and alter their moral sensibilities. Narratives capture the imagination, and elicit emotions that motivate action. It might seem impossible that a theory can include an emotion – not the concept of an emotion, but an emotion itself, but that is what I am going to propose. The motivating element is contained in the theory.

Given my theory of theory, it is not clear whether moral theories are in competition with one another. Presumably, some theories are better than others, but it is not obvious that there cannot be two equally good theories that are dramatically different. Most of us would strongly hesitate to allow the possibility of two equally good moral manuals that give conflicting moral directions. But as I've said, a theory is not the same as a manual, and I am leaving open the possibility that there can be more than one equally good moral theory. If a theory aims to be comprehensive, it would be naive to think that it will not be compared with other theories that aim to be comprehensive, but if we return to the map analogy, I think we can see that, since maps always leave something out, good maps can leave out different things. A topographical map often leaves out roads and political borders, whereas maps that indicate roads and borders typically lack indications of elevation and elevation contours. A map that has both is not necessarily a better map because too much information can be confusing. The map may be hard to read. But when a map is designed to get you somewhere, multiple maps should get you to the same place. That is why different moral theories yield mostly the same moral verdicts about particular cases. When they do not, we know that one of them is defective.

### 3. Features of a good moral theory

From what I have said in the last section, we can expand the list of features of a good moral theory from which we began:

1. A moral theory should simplify and systematize our pre-theoretical moral beliefs and practices, aiming at giving us understanding of the practices of morality, and sometimes resulting in a revision of those practices. This may lead to some distortion, but that can be tolerated if something is gained from the distortion and the distortion is not forgotten.

Although a moral theory may leave out some things, I think there are some very general concepts that ought to have a place in the theory. These concepts are *a good life*,

*a good state of affairs, a virtue, and a right act or a duty.* I do not mean to insist that every good moral theory must have a place for every concept. For instance, Elizabeth Anscombe (1958) remarked over a half century ago that Aristotle's ethical theory shows that it is possible to have a comprehensive theory without the concept of duty. But I mention this series of concepts because they play an important role in modern moral discourse, and the way moral theorists relate them reveals deep differences in the ways they understand moral practice. Usually one concept is fundamental and the others are defined by reference to the fundamental concept. For instance, Aristotle makes *eudaimonia*, or a good life, basic, with virtue defined in terms of a good life, and a right act defined in terms of what a virtuous person does. Kant makes a right act, in the strong sense of duty, basic. A virtue is defined in terms of the will to do one's duty. Mill makes a good state of affairs basic and defines a right act as an act that leads to a good state of affairs. The way a theory relates these concepts indicates their importance to the theorist. For instance, virtue gets no attention in Mill and, until recently, only a small amount of attention in contemporary forms of consequentialism. Plato and Aristotle talk about virtuous acts, but give little attention to a right act and arguably none at all to duty. So I am trying to illustrate both a difference in patterns of understanding these fundamental moral concepts, and the difference in the importance these concepts have in the respective theories. There are things that one theory considers important that another theory leaves out entirely or mentions only in passing. My purpose in making this remark is to point out that anybody constructing a moral theory ought to be aware of what is left out, and to do so only for good reason. That gives us another feature that we want in a good moral theory:

2. A moral theory should be as comprehensive as possible, compatible with simplicity. That should include giving a way to connect the concept of a *good life*, the concept of a *good state of affairs*, the concept of a *virtue*, and concepts of the moral properties of acts such as a *right act*, a *virtuous act*, and a *duty*.

I said that a moral theory is not a manual. It primarily aims at understanding, not directions for acting. But if a theory can give us directions while satisfying the other desiderata in a moral theory, so much the better. But to do that, there has to be some way that a user of the theory can connect the theory to what the theory is about. When we are using a city map, it is only helpful if we can find something in the city that hooks it to the map – *that* intersection over there is *this* one on the map. A stationary map will sometimes have a mark that says 'You are here' in order to orient the user. It seems to me that a moral theory needs something that serves that purpose – something that tells us that *this* element of moral belief or practice is *that* element in the theory. Unlike a map of an imaginary city, a moral theory is like a map of an actual city, and a user needs to connect the map with actual moral practice in order to negotiate the practice. I suggest, then, a third feature of a good moral theory:

3. A moral theory needs a hook to connect it to the domain of moral practices of which it is a theory. Just as a city map is useless unless we can identify something on the map by reference to something in our experience, a moral theory is useless

unless we can find a place where the theory connects to a part of the moral domain we can identify independent of the theory.

Finally, I want to turn to the problem with which I began this article. Very few moral theories have any connection with moral motivation. To the extent that they succeed at being abstract structures that *explain* our moral practices, they fail at motivating us to properly *engage* in moral practices. Moral theory is a map, but you will not get to the destination if you do not want to use the map. The problem is not that the theory does not have sufficient detail. The problem is that it is not the sort of thing that makes you want to get to the destination, no matter how much detail is added. If I am right that the primary function of a moral theory is to map the domain of morality, it does not count against a theory if it does not motivate. But, given that a moral theory justifies our moral practices by providing reasons for adopting certain courses of action and attempting to have certain character traits, there ought to be a connection between the moral reasons given by the theory and what actually motivates us. Otherwise, moral theory is like a map that tells us how to get to the destination we have, but we rarely use it for that purpose. The map might still be explanatory in some sense, but it does not explain what we actually do and why we do it. I suggest, then a fourth feature of a good moral theory:

4. Ideally, a moral theory can be used to actually motivate people to be moral, and to lead to moral improvement.

#### 4. Exemplarism

A stumbling block to any moral theory is the justification of the starting point. The theory I want to propose is foundational in structure, and although a foundational structure has advantages of simplicity and elegance, it is particularly vulnerable to skepticism about the proposed foundation. It is no accident that neo-Aristotelian theories that start from the concept of happiness or well-being generate more discussion about the vagueness of the foundation and the difficulty in justifying it than about the very rich and interesting derivative parts of the theory. The same point applies to the Kantian idea of the good will and social contract theories. I am not objecting to any of these theories, but when a theory begins with something that requires that important substantive issues have already been decided, it is bound to raise lots of questions before the theory gets going. I want to make the foundation of my theory a state that most people trust – their emotion of admiration. It is foundationalist, but the foundation is not conceptual. Instead, the construction of the theory begins with direct reference to exemplars of moral goodness, picked out by the emotion of admiration. My model for the foundational move in constructing a theory of this kind is the influential theory of direct reference, which was developed by Hilary Putnam (1975) and Saul Kripke (1980) in the 1970s, and which produced a revolution in semantics.

Leaving aside differences in different versions of the theory, the basic idea is that a natural kind term like ‘water’ or ‘gold’ or ‘human’ refers to whatever is the same kind of thing or stuff as some indexically identified instance. For example, gold is, roughly, whatever is the same element as *that*, water is whatever is the same liquid as *that*, a human is whatever is a member of the same species as *that*, and so on, where in each

case the demonstrative term 'that' refers directly – in the simplest case, by pointing. One of the main reasons for proposing this account of reference was that Kripke and Putnam believed that often we do not know the nature of the referent, and yet we know how to construct a definition that links up with its nature. We may not know the nature of gold – its deep structure – and for millennia nobody did, but that did not prevent people from defining 'gold' in a way that fixed the reference of the term and continued to do so after it was discovered what distinguishes gold from other elements. In fact, we would not say that modern humans 'discovered' the nature of gold unless we thought that modern speakers know the nature of the same stuff of which people used to be ignorant. The theory of direct reference explains how 'gold' referred to the same thing before and after the discovery of the atomic structure of gold, and hence how we can say that nobody knew what makes gold 'gold' until the modern era.

One of the exciting features of this proposal was that it meant that competent speakers of the language can use certain kinds of terms to successfully refer to the right things without going through a descriptive meaning. Compare a term like 'hammer'. When you say 'hammer', you refer to whatever satisfies a description given in advance. Presumably you cannot talk about hammers unless you grasp that description. In contrast, speakers need not associate descriptions with terms like 'water' and 'gold' in order to successfully refer to the right kinds. In fact, they can succeed in referring to water and gold even when they associate the wrong descriptions with terms like 'water' and 'gold'. It is not even necessary that every speaker be able to identify water and gold reliably themselves as long as some speakers in the community can do so and the other speakers rely upon the judgment of the experts.<sup>3</sup>

An interesting consequence of the theory of direct reference is that there are necessary truths discovered empirically. Kripke thought that once the reference of a natural kind term like 'water' is fixed by ostension, scientists then discovered the nature of water by observation. Under the assumption that the molecular structure of water is essential to it, it follows that certain necessary truths such as 'Water is H<sub>2</sub>O' are discovered a posteriori.

This idea can be used in the construction of a moral theory. I suggest that basic moral concepts are anchored in exemplars of moral goodness, direct reference to which are foundational in the theory. Good persons are persons *like that*, just as gold is stuff *like that*. Picking out exemplars can fix the reference of the term 'good person' without the use of descriptive concepts. It is not necessary for ordinary people engaged in moral practice to know the nature of good persons – what makes them good. In fact, it is not necessary that anybody knows what makes a good person good in order to successfully refer to good persons, any more than it was necessary that anybody knew what makes water 'water' to successfully refer to water before the advent of molecular theory. We need not associate any descriptive meaning with 'good persons', and users of our language can successfully refer to good persons even when they associate the wrong descriptions with the term 'good person'. As with natural kinds like gold and water, people can succeed in referring to good persons as long as they, or some people in their community, can pick out exemplars.

Practices of picking out such persons are already embedded in our moral practices. We learn through narratives of fictional and non-fictional persons that some people are admirable and worth imitating, and the identification of these persons is one of the

pre-theoretical aspects of our moral practices that theory must explain. Moral learning, like most other forms of learning, is principally done by imitation. Exemplars are those persons who are *most imitable*, and they are most imitable because they are most admirable. We identify admirable persons by the emotion of admiration, and that emotion is subject to education through the example of the emotional reactions of other persons. I am proposing, then, that the process of creating a highly abstract structure to simplify and justify our moral practices is rooted in one of the most important features of the pre-theoretical practices we want to explain, the practice of identifying exemplars, and in a kind of experience that most of us trust very much – the experience of admiration, shaped by narratives that are part of a common tradition.

This theory is compatible with the view that our identification of exemplars is revisable. Just as we can be mistaken in our judgment that some portion of water we identify is really water, we can also be mistaken in our judgment that some person we identify as paradigmatically good is really good. However, I don't think that we could be mistaken about most exemplars for the same reason that we cannot be mistaken that most of what we take to be water is water. That is because there is a referential connection between good persons and the individuals we identify as good – good persons are persons like that, and the parallel point holds for water.<sup>4</sup>

One of the most interesting features of the Kripkean account of natural kinds is the way empirical investigation can reveal natures, and I think this also is a feature of exemplarist virtue theory. If the concepts in a formal ethical theory are rooted in a person, then narratives and descriptions of that person are morally revealing. It is an open question what it is about the person that makes him or her good. For the same reason, when we say that a good person is a person like that, and we directly refer to St Francis of Assisi, or to Confucius, or to Jesus Christ, we are implicitly leaving open the question of which properties of Francis, Confucius, or Christ are essential to their goodness. The exemplarist approach has the advantage that substantive matters about what makes a person good need not be settled at the outset. We need not start by assuming that certain traits are the virtues or that certain acts are right. But we do think in advance of observation that what makes a good person good is their psychological structure, just as we think in advance of observation that what makes water 'water' is its molecular structure. Observation tells us what the psychological structure of a person is, just as observation tells us what the molecular structure of water is. If we also think that being H<sub>2</sub>O is an essential property of water, we might also think that having certain properties – e.g. compassion, justice, and wisdom – are essential to being a good person. I think, then, that Kripke is right that deep and important, perhaps even necessary, properties of the object class can be determined by empirical observation, although the determination of what counts as deep and important is not itself empirical. Since narratives are a form of detailed observations of persons, exemplarism gives narrative a crucial place within the theory analogous to scientific investigation in the theory of natural kinds. I am leaving open the possibility that narratives even reveal necessary features of value by uncovering the deep properties of a good person. If so, there would be necessary a posteriori truths in ethics that can be discovered in a way that parallels the discovery of the nature of water.

Perhaps we will find out that exemplars do not have a common deep psychological structure. Perhaps it will turn out that 'good person' is not like a natural kind term,



referring to something that has an essential nature. Maybe there is nothing interesting and important that the Dalai Lama, Jesus Christ, Confucius, the Stoic sage, and the Greek heroes have in common. I do not want to prejudge the results of close observation of exemplars. After all, it could have turned out that the stuff in the lakes and streams and falling from the sky when it rains does not all have a common chemical structure. So 'good person' might turn out to be more like 'tasty drink' than like 'water'. A more plausible outcome is that 'good person' is multiply realizable.<sup>5</sup> There is not a single essence of good personhood, but there is a set of interesting, yet different, ways in which a person can be good. It is also possible that the set of exemplars gradually changes over time. After a moral revolution such as the revolution in attitudes towards persons of different races or ethnic backgrounds, some, but by no means all, features of the persons we recognize as exemplars changed. I am suggesting that these are all testable hypotheses.

If we can identify virtuous exemplars, empirical research may reveal interesting features of their attitudes and behavior. I know of at least three topics of recent empirical research that directly bear on the theory I am proposing. First, there is a small but growing body of research in neuroscience on the features of exemplars (e.g. Van Slyke et al., 2012; Walker and Henning 2004). There is also relevant research on mimesis or imitation (e.g. Garrels, 2006; Herdt, 2008: Ch. 1; Iacoboni, 2008). Third, there is research on admiration, or what Jonathan Haidt calls "elevation" (e.g. Algoe and Haidt, 2009; Viancello et al., 2010). It would be very interesting to connect these pieces of empirical work, and find out whether there is empirical validation of a connection between the emotion of admiration and imitation, as I am suggesting here, and how the identification of exemplars by other methods connects with the identification of exemplars through the emotion of admiration. I know of work on L'Arche caregivers, and the neuroscience of exceptional persons as they play economics games.<sup>6</sup> There are many other potential empirical studies on exemplars that would be useful for identifying their moral traits and behavior dispositions. The theory I am proposing shows how empirical research of this kind, as well as narratives about exemplars, can be integrated into an abstract structure that I believe illuminates the moral life.

Let me review. What I mean by an exemplar is a paradigmatically good person. An exemplar is a person who is most admirable. We identify the admirable by the emotion of admiration. I do not assume that we always trust our emotion of admiration, and the emotion is shaped by the emotional responses of others, but when our emotion survives reflection and we trust it, we take the object of admiration to be admirable. A person who is admirable in some respect is imitable in that respect. The feeling of admiration is a kind of attraction that carries the impetus to imitate with it. This is rough because there are many reasons why we do not or cannot imitate those we admire. For instance, I admire Robert Falcon Scott, who made an expedition to the South Pole in 1912 under very arduous circumstances, resulting in the death of the entire party, but I have no inclination to go to the South Pole myself, even if I was assured I would not die. But there is a way in which my admiration *does* make me want to imitate Scott – not in any particular way, but in wanting to be the kind of person who could do such a thing. So I do think that the urge to imitate is included in admiration.<sup>7</sup> The function of admiration is very important in the theory I am proposing because the ways in which exemplars are admirable, and hence imitable, can be used to give us both a way of understanding significant moral

concepts and a way of using those concepts to make ourselves and our lives conform to the admirable.

My view of a moral theory is that it is an abstract structure, and I said that, to be comprehensive, it should include definitions of all the most important moral concepts. So far I have proposed that the theory begin with direct reference to paradigmatically good persons, picked out through the emotion of admiration. I suggest that basic moral concepts be defined roughly as follows:

A *virtue* is a trait we admire in an admirable person. It is a trait that makes the person paradigmatically good in a certain respect.

A *right act* in some set of circumstances C is what the admirable person would take to be most favored by the balance of reasons in circumstances C.

A *good outcome* is a state of affairs at which admirable persons aim.

A *good life* (a desirable life, a life of well-being) is a life desired by admirable persons.

In each case, the definiendum is defined via indexical reference to a paradigmatically good person. So a virtue is a trait we admire in *that* person and in persons like that. A good state of affairs is a state of affairs at which persons like that aim. A good life is a life desired by persons like that. A right act is an act a person like that would take to be favored by the balance of reasons.

Notice that these definitions are not intended to reveal the ‘deep’ nature of virtue, right action, or a good life. They are like defining water as stuff like that, where the determination of the deep nature of water is left for investigation. The purpose of the definition is to identify the reference of the term to make investigation of it possible. Similarly, the purpose of the definitions I have given is to permit us to identify the reference of moral terms in such a way that we know what to investigate to find out what virtue, right action, and a good life are.

Aristotle uses a definition similar to this pattern in the *Nicomachean Ethics* when he defines virtue as a mean between extremes ‘as a person with practical wisdom (*phronesis*) would determine it’ (1106b35). A well-known objection to this style of definition for right action is that what is right for the exemplar might be wrong for me because my defects change what is right for me to do. For instance, if I followed the exemplar’s course of action, that might lead to serious temptation for me, but not for the exemplar. Or the exemplar’s behavior might not be right for me because it is too difficult for me to do. I should aspire to something within my reach. Furthermore, there might not be such a thing as what the exemplar would do in circumstances like mine because the exemplar would never be in circumstances like mine. I might be in a fix that no exemplar would get into. For example, maybe I made conflicting promises, then lied to avoid keeping one of the promises, then discovered that if I try to keep the other promise, I must lie again, and so my misdeeds are spiraling out of control and sinking me into further wrongdoing. What should I do in such a situation? It cannot be what the exemplar would do because the exemplar would not be in such a situation.

I think that we can answer this objection by taking another hint from Aristotle. Although Aristotle defines virtue as a mean between extremes, the determination of

which is given by the *phronimos*, he says that a person with a vice should aim for the opposite vice. A cowardly person should aim for foolhardiness, an intemperate person should aim for insensibility, and so on. I am sure we can think of ways this advice would not always work, but the important point here is that moral improvement may come in stages, and direct imitation of the exemplar may come only after a person has reached a certain level of moral development. Before that, the person does better at imitating persons who are better than he is, but not so much better that he cannot see clearly the path to becoming like the exemplar.

The problem of imitating a person too much better than oneself illustrates another feature of the relationship between virtue and right action: the importance of focusing first on acquiring the traits of the exemplars (the virtues), and secondarily on the behavior of the exemplars as models for one's own behavior. An act performed by a foolish person may differ in its moral value from the same act performed by a wise person. We do better by aiming to imitate the character of exemplars, both because their specific acts may be inappropriate for us, and because we cannot fully succeed in creating moral value through our acts without virtuous motives, understanding of the meaning of our acts, and knowledge of their consequences. I should first try to become a person like that, and then I will be on a firmer ground when I imitate their acts.

## 5. Advantages of exemplarist virtue theory

Exemplarist virtue theory has a number of desirable features in a theory:

1. It is simple, comprehensive, and I think it gives us understanding of the moral life, although that is for others to judge.
2. It has a hook connecting the theoretical structure of the theory with the actual practice of the moral life. As I mentioned, we cannot use a street map unless we can tell where we are by observation, which permits us to say that *that* intersection is *this* place on the map. Similarly, our experience of admirable persons permits us to say a virtue is a trait of *this* person, the right thing to do is what this person says it is, and so on. Our experience of admirable persons provides the hook that ties our theoretical moral discussions with our experience.
3. Since the foundational move in the theory is not conceptual, controversial substantive issues about the virtues, the nature of a good life, and the catalogue of right and wrong acts do not have to be decided in advance. Since the theory makes these issues a matter for observation of admirable persons, it can be used both to justify and to test our moral judgments.
4. The theory can be used for the practical purpose of moral improvement as well as the theoretical purposes of a theory because the theoretical structure is built upon a motivating emotion. If I am right about the psychology of admiration, admiration is the link between the desire to be moral and a conceptual network of moral concepts. Admiration is elicited through personal experience and narratives. This gives it another advantage:
5. Narratives have a place within the structure of the theory analogous to scientific observation in the Putnam/Kripke theory of natural kinds. Narratives are the

primary vehicle through which moral education is accomplished, but it has always been difficult to connect narratives with a theoretical structure. For instance, Christian ethics is supposed to be focused on the imitation of Christ, but there is rarely any attempt at a connection between the person described in the Gospels and systematic moral theology. My approach can be used to make the imitation of Christ the foundation of a comprehensive moral theory that serves the purposes of a philosophical theory, but which integrates into the theory Gospel narratives, lives of the saints, and reflections on the experience of admirable Christians. What I am providing is the structure; what Gospel narratives and other Christian stories provide is the substance.

I know of one case in which the exemplarist structure I have proposed may already exist in a full ethical theory. My colleague Amy Olberding argues in her book on the ethics of Confucius (2011) that this theory offers the key to interpreting the ethics of the *Analects*. Confucius treats the Duke of Zhou as an exemplar, and Confucius' students treat him as an exemplar. In both cases, the ideas of a good human trait and the proper way to behave are determined by reference to the exemplar. Olberding's book interprets the text in detail using this theoretical structure. I am not professionally qualified to say whether Confucian ethics is exemplarist, but I find it heartening that the exemplarist framework offers possibilities of interpretation that have not previously been noticed.

6. The theory connects with empirical studies, and suggests new avenues of cross-disciplinary research on admiration, the features of exemplars, and the place of imitation in moral learning. It might seem that persons cannot be investigated scientifically the way water was investigated in the seventeenth century, but the neuroscience of virtuous exemplars is already underway, and it may yield interesting results that are not noticed in ordinary human observation and narratives. The theory may also lead to new empirical studies on the identification of exemplars and the connection between the emotion of admiration and emulation of admired persons.
7. The theory provides a framework for cross-cultural dialogue on morality. It permits us to identify both cross-cultural similarities and cross-cultural differences in moral beliefs by linking the traits a group of people call virtuous, the acts they call right, and the lives they think are most worth living to the people they admire as exemplars. People can usually understand why the exemplars of a different culture or religious community are admirable. So the Western Christian can admire Confucius, non-Christians admire St Francis, and people who otherwise have very different ideas of a good life may admire secular saints like Martin Luther King Jr. But people outside a community also often notice features they think are missing or underdeveloped in the exemplar, and the differences can be as important as the similarities because they give us a way to engage in cross-cultural critique. I am suggesting a vocabulary for cross-cultural discourse that focuses our attention on the objects of admiration. The objects differ in some ways, but the emotion itself is universal.

Earlier, when I gave my theory of moral theory, I said that I think there can be more than one good moral theory. My aim here has not been to critique other theories, but since exemplarist virtue theory provides a link between moral education and the theoretical side of ethics, it has particular importance for the interests of this symposium.

### Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

### Notes

1. I described a theory of this kind for the purposes of Christian moral theology in Zagzebski (2004). The general form of the theory is outlined in Zagzebski (2010), and this article is based on that one. I am currently working on a book on exemplarist virtue theory that develops the theory for the purposes I am describing here.
2. What I have in mind here is the sort of manual that Catholic priests used to use for the purpose of counseling sinners in Confession. They included a large catalogue of sins, classified under general moral principles. But priests were trained to understand the moral theology underwriting the manual.
3. In Kripke's initial version of the theory, a natural kind term refers to something that is causally linked to an historical dubbing. This theory was plausible as a theory of the reference of proper names. I am not using this causal theory of reference in what follows.
4. As Putnam famously argued in his example of water and Twin Earth, if the Twin Earthians point to XYZ when they say 'water', while we Earthians point to H<sub>2</sub>O, we are not disagreeing about the nature of water. We mean something different by 'water' than they do. Similarly, I would say that if one group of people routinely points to people who are brutal, greedy, and envious when they say 'good person', whereas others point to people who are compassionate, generous, and sympathetic, their disagreement is more radical than a disagreement about which traits are good traits. They disagree about what 'good' means.
5. I thank Catherine Elgin for this suggestion.
6. These are ongoing studies conducted by the research team of Michael Spezio, Gregory Peterson, Warren Brown, James Van Slyke, and Kevin Reimer, editors of Van Slyke et al. (2012). They discuss their work on L'Arche caregivers in Reimer et al. (2012).
7. David Velleman (2007: 523) has proposed that admiration leads to emulation by motivating wishful picturing of oneself in the image of the admired person. There are other models of the connection between admiration and emulation or imitation. I am not endorsing a particular model.

### References

- Algoe SB and Haidt J (2009) Witnessing excellence in action: The 'other praising' emotions of elevation, gratitude, and admiration. *Journal of Positive Psychology* 4: 105–127.
- Anscombe GEM (1958) Modern moral philosophy. *Philosophy* 33(1): 1–19.
- Garrels SR (2006) Imitation, mirror neurons, and mimetic desire: Convergences between the mimetic theory of René Girard and the empirical research on imitation. *Contagion: Journal of Violence, Mimesis, and Culture* 12: 47–86.
- Herdjt JA (2008) *Putting on Virtue: The Legacy of the Splendid Vices*. Chicago, IL: University of Chicago Press.
- Iacoboni M (2008) *Mirroring People: The New Science of How We Connect with Others*. New York: Farrar, Straus, and Giroux.

- Kripke S (1980) *Naming and Necessity*. Oxford: Blackwell.
- Olberding A (2011) *Moral Exemplars in the Analects: The Good Person is That*. New York: Routledge.
- Putnam H (1975) The meaning of 'meaning'. In: *Mind, Language, and Reality (Philosophical Papers vol. 2)*, pp. 215–271. Cambridge: Cambridge University Press.
- Reimer K, et al. (2012) Maturity id explicit: Self-importance of traits in humanitarian moral identity. *Journal of Positive Psychology* 7(1): 36–44.
- Van Slyke J, Peterson G, Brown WS, Reimer KS and Spezio ML, eds (2012) *Theology and the Science of Moral Action: Virtue Ethics, Exemplarity, and Cognitive Neuroscience*. Abingdon: Routledge.
- Velleman JD (2007) Motivation by ideal. In: R Curren (ed.) *Philosophy of Education: An Anthology*, pp. 517–526. Oxford: Blackwell.
- Viancello M, Gallani EM and Haidt J (2010) Elevation at work: The effects of leaders' moral excellence. *Journal of Positive Psychology* 5: 390–411.
- Walker LJ and Henning KH (2004) Differing conceptions of moral exemplarity: Just, caring, brave. *Journal of Personality and Social Psychology* 86: 629–47.
- Zagzebski L (2004) *Divine Motivation Theory*. Cambridge: Cambridge University Press.
- Zagzebski L (2010) Exemplarist virtue theory. *Metaphilosophy* 41(1–2): 41–57. Reprinted in H Battaly (ed.) (2010) *Virtue and Vice: Moral and Epistemic*, pp. 39–56. Oxford: Wiley-Blackwell.

### Author biography

Linda Zagzebski is George Lynn Cross Research Professor, and Kingfisher College Chair of the Philosophy of Religion and Ethics, at the University of Oklahoma. She writes in the areas of epistemology, philosophy of religion, and virtue theory. Her most recent book, *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief* (Oxford University Press, 2012) was based on her Wilde Lectures at Oxford University (2010), Kaminski Lectures at the John Paul II Catholic University of Lublin (2011), and Olaus Petri Lectures at the University of Uppsala (2011). The book was supported by a Guggenheim Fellowship. Some of her previous books include *Virtues of the Mind* (1996), *Divine Motivation Theory* (2004), and *On Epistemology* (2008). She is past president of the American Catholic Philosophical Association, and past president of the Society of Christian Philosophers.