EXEMPLARIST VIRTUE THEORY

LINDA ZAGZEBSKI

Abstract: In this essay I outline a radical kind of virtue theory I call exemplarism, which is foundational in structure but which is grounded in exemplars of moral goodness, direct reference to which anchors all the moral concepts in the theory. I compare several different kinds of moral theory by the way they relate the concepts of the good, a right act, and a virtue. In the theory I propose, these concepts, along with the concepts of a duty and of a good life, are defined by reference to exemplars, identified directly through the emotion of admiration, not through a description. It is an advantage of the theory that what makes a good person good is not given *a priori* but is determined by empirical investigation. The same point applies to what good persons do and what states of affairs they aim at. The theory gives an important place to empirical investigation and narratives about exemplars analogous to the scientific investigation of natural kinds in the theory of direct reference.

Keywords: virtue, exemplars, moral theory, moral foundationalism, admiration.

1. Introduction

In this essay I outline a radical kind of virtue theory I call "exemplarism," which is foundational in structure but which is grounded in exemplars of moral goodness, direct reference to which anchors all the moral concepts in the theory. I compare several different kinds of moral theory by the way they relate the concepts of the good, a right act, and a virtue. In the theory I propose, these concepts, as well as the concepts of a duty and a good life, are defined by reference to exemplars, which are identified directly through the emotion of admiration. It is an advantage of the theory that what makes a good person good is not given *a priori* but is determined empirically. The same point applies to what good persons do and the states of affairs at which they aim. The theory gives an important place to empirical investigation and narratives about exemplars, analogous to the scientific investigation of natural kinds in the theory of direct reference.

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. Constructing a moral theory is part of moral practice. Moral practice includes the construction of theories about the practice.

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 do not think 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 wouldn't construct theories unless we thought that there is something to be gained by attending to certain features of our moral practices and their relations at the expense of others. 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, nonhuman 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 as the city is, and the map would not help us understand the city's layout. So the map leaves out many things, and it may also distort some things. Think of the shape of a country like 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. Most moral theorists believe that a good moral theory leaves out the identity of the persons in the practice and there is no first person pronoun in the theory. Others believe it would be better if we identified certain persons, or at least put thicker descriptions of persons into a theory. It seems to me that just as a two-dimensional map of the world distorts the shape and relative size of countries, a moral theory without personal identity distorts the moral relations among persons. However, it does not follow that we should put identities into our theories. Leaving out identity gives us an understanding of general moral relations between persons that would be very hard to grasp with the identities specified. If you want to highlight the fact that a person is a member of the community of moral agents, you

don't want to add that the person is the same one your grandfather despised until his death because of an injury to your grandmother. But that information would be relevant to other parts of moral practice, such as those involving loyalty. 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 make a choice to leave out personal identity in a moral theory, and the choice is made for a reason. 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.

You might think that a theory should include reference to oneself and the identity of some other persons because you think a good theory should include all the conditions relevant to moral decisions. The identity of some persons is relevant to many decisions, particularly those arising from loyalty, friendship, and familial relations. But even though I think that we hope to get moral guidance from a good theory, 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 a map of the world is not detailed enough to do that and it is not intended to do so. I think of a moral theory as more like a map of the world than a street map. Theories of parts of morality may be closer to street maps, and it is good that we have them, but 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 place to another, you might not use a map at all. A Global Positioning System would 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. 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 correlatively, showing us which beliefs and practices are unjustified. The aim of telling us what to do in any given situation is secondary. The different elements to be explained include reactive emotions such as admiration, blame, praise, and remorse, practices of punishing some but not all acts of wrongdoing, rules such as the Golden Rule or the Ten Commandments, and values such as freedom, fulfillment, and social cohesiveness, which are often revealed in narratives that are cherished by a particular community. These are only some of the elements of our moral practices that preexist our theory. My point is that there is already something there that we seek to understand 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 a street map. 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 in certain ways. 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, it seeks only to depict it.

If a moral theory is intended to explain our moral beliefs and practices in a way that can lead us to revise those practices, it is natural to wonder how a theory can do that. For whom is a theory intended? Would we want to promulgate it for the whole society? Is it instead for moral leaders? Or is there a class of people whose business it is to produce moral theories—the moral philosophers? I assume it is the latter. We produce moral theories first for other philosophers, and secondarily for students in philosophy classes. But we think that theoretical discussions can ultimately influence practice. In this essay I compare a variety of moral theories at the most abstract level of theoretical structure. It is pretty obvious that theory at that level does not influence practice, but one of the issues I am interested in is the path from abstract theory to revisions of practice. I suspect that the path goes through disciplines other than philosophy, publications aimed at the general educated public, the arts and the media, and sometimes the law, and most of the time the path withers before ordinary people are affected, but theory can influence practice. I think that it is an advantage if a theory can link up with moral practice in a plausible way, particularly if it can link up with narratives that capture the imaginations of ordinary people.

Since we are going to compare moral theories, an obvious question to ask is whether the 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. If we also think a manual is generated from a theory, that can explain our resistance to the idea that there can be two equally good moral theories. We don't want to be committed to allowing two equally good but incompatible manuals. But as I've said, the connection between a theory and a manual is not straightforward. Most moral theories generate most of the same moral directions, so the theories differ more than the manuals they generate. In fact, the manuals would be mostly the same, assuming the moral practices the theories explain and justify preexist the theories.

But a comparison of theories often becomes most interesting in precisely those places in which they lead to conflicting directions—differences in the manuals they produce. I am going to propose a theory that does not produce a manual by itself but produces it in conjunction with other components of our moral practices, particularly narratives. The manual we get from such a theory is no doubt vague, but I prefer a vague manual to the wrong manual. In any case, I am not going to say much about moral manuals and how we get them.
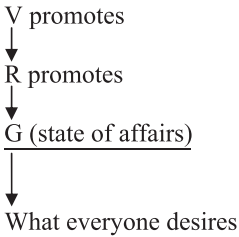
3. The Structure of Some Moral Theories

If we want a comprehensive yet simple moral theory, I think we should start by looking at three very general concepts of positive moral evaluation. (Theories are almost always built around positive rather than negative evaluation). Deep differences between theories can be revealed by comparing the ways they relate these three concepts: *the good* (G), *a virtue* (V), and *a right act* (R). The good applies to different kinds of things in different theories, and it might seem misguided to compare the good in hedonistic utilitarianism, Aristotle's idea of *eudaimonia*, Kant's notion of a good will, and Plato's Form of the Good. Clearly, there is a sense in which Plato, Aristotle, Kant, and Mill are not talking about the same thing, but there is a sense in which they *are* because each has selected a form of the good that is allegedly pivotal in understanding moral practice, and that has important relations to the evaluation of acts and persons.

There is another complication that I want to mention before we compare the theories. There are more concepts of act evaluation than of anything else. In addition to the concept of a right act, there is the concept of a virtuous act and the concept of a duty. Furthermore, the concept of a right act is ambiguous because sometimes it is treated as the complement of a wrong act—an act that is not wrong, in which case it includes the evaluatively neutral as well as the evaluatively positive. Sometimes, instead, it is treated as the equivalent of what one *should* do—the act that is favored by the balance of moral reasons. And sometimes it is treated as the equivalent of the even stronger notion of duty—what one *must* do. If something is a duty, it is wrong not to do it. So a right act can be understood as (a) an act it is not wrong to do, or (b) an act it is wrong not to do, or (c) an act that one has most moral reason to do. A virtuous act is typically treated as an act that expresses a virtue and is hence good, but it is not a duty, nor is it conceptually equivalent to what one should do in the sense of an act that is favored by the balance of moral reasons.
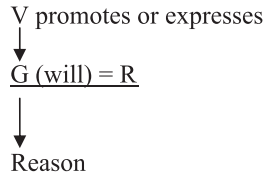
The purpose of the diagrams below is to help us in comparing theory structure and to reveal what a theory leaves out, as well as the different ways the theories organize these three moral concepts. In each diagram the foundational concept is at the bottom, and the concepts above are

defined in terms of their relations to the concepts below. For instance, in utilitarianism, a right act is defined in terms of a good state of affairs, and the relation is one of promotion. A right act is one that promotes a good state of affairs. In some cases the relation in question is one of constituency. (I use "<" to mean "is a constituent of".) In some theories the real foundation is something outside ethics, such as reason or human nature or what everyone desires, and I indicate this below the line separating the moral concepts from the nonmoral foundational concept.
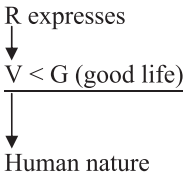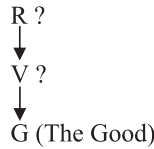
Classical utilitarianism   **Kantian deontology**
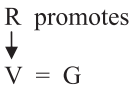
V promotes   V promotes or expresses
↓   ↓
R promotes   G (will) = R
↓   │
G (state of affairs)   ↓
│   Reason
↓
What everyone desires

Aristotelian virtue theory   **Platonistic virtue theory**

R expresses   R ?
↓   ↓
V < G (good life)   V ?
│   ↓
↓   G (The Good)
Human nature

Perfectionism   **Exemplarism**

R promotes   V   R   G (outcome)   G (life)
↓
V = G

                                               G (that person)
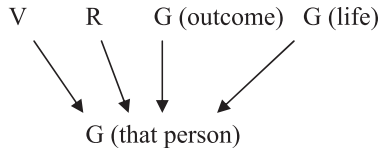
The diagrams reveal some differences among the theories. Whatever is at the bottom of the diagram is most fundamental in the theory, and I think also most important. What is derivative is less important and typically gets less attention. In fact, virtue gets no attention in Mill and only a small amount 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 three 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. This might seem obvious, yet it is common for a philosopher to critique the account of his favorite moral concept by another philosopher who really is not interested in the concept. That sometimes happens in critiques of the way virtue theorists use the concept of a right act.

A comparison of the diagrams also shows us some features that these theories have in common. Perhaps the most obvious one is that they are all foundational in structure. Making one concept foundational has the advantage of theoretical elegance, but I think it should be acknowledged that if the aim of moral theory was just to simplify and systematize our moral practices, there would be no special advantage in a foundational structure. Any clever person could make up structures using the concepts of good, right act, and virtue that are just as simple as the theories in the diagrams, but that do not have a hierarchical form with a single foundational concept. The attraction to foundationalism, I believe, is due to the fact that in the modern era a moral theory is not only expected to justify our individual moral practices, it is assumed that the entire practice of morality itself is in need of justification, and it is assumed further that a secure foundation is the best way to justify the practice of morality. Neither of these assumptions is obvious. I have already said that I accept the fact that justification is a purpose of moral theory, but I do so as a concession to modern moral philosophy. This aim would not have been recognized in the premodern era, and I imagine that that is the reason it is so difficult for modern philosophers to recognize anything in the premodern period that constitutes moral theory at all. It is only at the risk of severe artificiality that the moral philosophies of Plato and Aristotle can be squeezed into the structures I have labeled ''Platonistic'' and ''Aristotelian,'' nor can they be easily aligned with any alternative foundationalist structure. Readers may have their own view on whether any moral philosopher before the modern period advocated a theory with a foundationalist structure. I suspect there is none, and that is because nobody before the modern era thought that moral theory needed to justify moral practice. Moral philosophers would have thought that moral theory is a part of philosophy that simplifies and systematizes a complex practice, but there would have been no thought that the practice itself had to be justified. Granted, they sometimes used moral theory to adjudicate disputes about issues such as justice in war, but they surely did not think that morality itself is in need of justification. That assumption is probably due to the naturalistic tenor of our times. The thought now is that we can be confident of the existence of nature in the sense of nature investigated by empirical science, but morality is not part of that, and we

cannot be confident of its credentials. Morality needs something outside it to justify it, something of which those who question the credentials of morality can be confident.[1]

This is why it is desirable that moral theory be foundational in structure. It doesn't actually have to be, but it is much easier if it is. Presumably it is hard enough to tie even one of the fundamental concepts of moral evaluation to something outside morality; to tie more than one would be more challenging than philosophers can tolerate. But we could declare victory if there is a single concept foundational to morality that is justified by reference to something outside morality on which everyone can agree, and which in turn can support all other moral concepts. And that is the aspiration of at least three of the theories in the diagram. The foundational concept is tied to something else that is not moral and that is accepted by anybody who might be inclined to question the justifiability of moral practice.

The three I have in mind are Kantian moral theory, neo-Aristotelian moral theory, and utilitarianism. Kant grounds the concept of a right act in reason understood formally or, at least, in some uncontroversial way. Neo-Aristotelian virtue theory grounds the good for human beings in human nature, understood in a way that makes minimal claims about the substance of human nature. Hedonistic utilitarianism grounds the good of human beings in something every human, indeed, every animal, naturally desires: to get pleasure and to avoid pain. In each case, morality derives from something that is allegedly less in need of justification, and it does so in a simple and elegant way. I think this is also the reason that Platonistic theories seem nonexplanatory to most modern philosophers.

Now I would like to propose another feature that a good theory should have, whether or not it is foundationalist in structure. Since a theory is a theory *of* something, there has to be some way that a user of the theory can connect the theory to what the theory is about. If a theory can be compared to a map of some domain, the user of the theory should be able to superimpose the map on the domain. She should be able to say, "*This* element in the theory refers to *that* element in the domain." Sometimes the elements of the theory are objects of the user's background experience, and this feature is easy to satisfy. When the economist refers to prices and interest rates, he assumes the users of the theory can identify those elements in the world of economic exchange. Sometimes the theorist gives directions for finding the elements of the theory, as when the botanist draws pictures of the plants she is classifying. The same point applies to a city map. It is useful only if the user can find something in the city that

---

[1] It is interesting that the motive for Divine Command Theory is different. The Divine Command theorist rejects the independence of moral authority from divine authority, not because our moral practices are in need of justification, but because divine authority is threatened if moral authority has a source independent of God.

hooks it to the map—that intersection over there is this one on the map. A stationary map will sometimes say "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 him 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 moral practice in order to negotiate the practice.

Let me now review the desiderata of a moral theory as I understand it.

1. A moral theory should simplify and systematize our pretheoretical moral beliefs and practices, aiming at giving us understanding of the practices of morality, and sometimes resulting in a revision of those practices. It is possible that simplifying and systematizing results in some distortion, but that can be tolerated if something is gained from the distortion and the distortion is not forgotten.

2. A moral theory is not a manual and its main purpose is not to give directions in decision making, but it is an advantage if a theory can help us in our practical lives. Many directions for making moral decisions already exist in our pretheoretical practices, including narratives, parables, and practical rules.

3. A moral theory should also justify our moral beliefs and practices. I am not convinced that the entire domain of moral practice needs to be justified by something outside the practice, but I am willing to accept both the aim of producing a foundationalist structure, and the need to make the foundation something relatively uncontroversial.

4. A moral theory needs a hook to connect it to the domain of moral practices of which it is a theory. Just as a 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.

4. Exemplarism

The theory I want to propose is foundational in structure, but the foundation is not conceptual. Instead, the construction of the theory begins with direct reference to exemplars of moral goodness. My model for the foundational move in constructing a theory of this kind is the Putnam-Kripke theory of direct reference, particularly in the form in which it was used to define natural kind terms.[2] Leaving aside differences

---

[2] This theory originated with Saul Kripke's *Naming and Necessity* (Kripke 1980) and Hilary Putnam's "The Meaning of 'Meaning'" (Putnam 1979, first published in Gunderson 1975).

in the versions of the theory, the basic idea is that a natural kind term such as "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 has the advantage of explaining how "gold" referred to the same thing before and after the discovery of the atomic structure of gold.

This proposal began a revolution in semantics because it meant that competent speakers of the language can use terms to successfully refer to the right things without going through a descriptive meaning.[3] Unlike a term such as "hammer," the referent of natural kind terms like "water" and "gold" is not whatever satisfies a description given in advance. Because speakers need not associate descriptions with natural kind terms in order to successfully refer to the right kinds, an important consequence of this theory is that it is possible that speakers succeed in referring to water and gold even when they associate the wrong descriptions with terms like "water" and "gold."[4] What is required instead is that they be related by a chain of communication to the actual stuff water and gold.[5] It is not even necessary that every speaker be able to identify water and gold reliably herself as long as some speakers in the community can do so and the other speakers rely upon the judgment of the experts.

An interesting feature of this theory is that a definition through direct reference is only a contingent truth. It is not a necessary truth that what I

---

[3] Initial discussion focused on natural kind terms and proper names, but later the theory was applied to a broader class of terms. The extent of the class of terms which can refer directly is not important for my point in this essay.

[4] On one version of the theory, natural kind terms have no meaning; they are purely denotative (like Mill's theory of proper names). On another version of the theory, natural kind terms have a meaning, but meanings are not in the head. That is, they are not something a speaker grasps and through which he finds the referent. See Putnam 1979.

[5] In some later versions of the theory the chain is thought to be causal, hence the term "causal theory of reference," but the idea that the use of a term by many speakers is causally connected is not a necessary part of the theory.

am pointing to right now is gold. Of course, *if* I am pointing to something gold and if it is essential to anything gold that it is gold, then it is essential to the thing I am in fact pointing to that it is gold. However, it is not a necessary truth that I am pointing to *this* thing, and so it is not necessary that I am pointing to gold. Hence, we must accept either that some definitions are not necessary truths or that the way we connect words with objects in the type of "definition" we have been considering is not actually a definition. For my purposes, it does not matter which option we take.

One other interesting consequence of the theory of direct reference is that there are necessary *a posteriori* truths. Kripke thought that once the reference of a natural kind term like "water" is fixed by ostension, scientists can then discover the nature of water empirically. Under the assumption that the nature of water is essential to it, it follows that certain necessary truths such as "Water is H_2O" 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 that ordinary people engaged in moral practice know the nature of good persons—what makes them good. In fact, it is not necessary that anybody know 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 at least some people in their community, can pick out exemplars.[6]

Practices of picking out such persons are already embedded in our moral practices. We learn through narratives of both fictional and nonfictional persons that some people are admirable and worth imitating, and the identification of these persons is one of the pretheoretical aspects of our moral practices that theory must explain. Moral learning, like most

---

[6] It is an important part of the theory of direct reference that a person can successfully refer when she is not good at identifying the referent herself, and even when she has never had any experience of the referent. So we all can refer to uranium, and Putnam says he can refer to elm trees even though he is not good at recognizing them. Nonetheless, those of us who are only distantly related to uranium or elm trees are epistemically disadvantaged relative to the experts who are good at identifying them. We lack the understanding that the experts have. But given the importance of moral understanding by as many people as possible in a moral community, it is important that the ability to identify exemplars is spread as widely as possible. This is one of the functions of narrative, mentioned below.

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 itself 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 pretheoretical 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.

I am assuming that the emotion of admiration is generally trustworthy when we have it after reflection and when it withstands critique by others. We have no guarantee that what we admire upon reflection is admirable, but then we do not have any guarantee that our vision or memory is trustworthy if it withstands reflection either. All we can do is the best we can do by using our faculties as conscientiously as we can, and our disposition to admiration is one of those faculties.

This theory is compatible with the possibility that paradigmatically good individuals are only contingently good,[7] and it is also compatible with the theory that our identification of exemplars is revisable. Just as we can be mistaken in our judgment that some portion of a substance we identify as water is really water, we can also be mistaken in our judgment that some person we identify as paradigmatically good is really good. However, given that there is a conceptual connection between water and stuff "like that," we cannot be mistaken in thinking that most of what we take to be water is water. Similarly, there is a conceptual connection between good persons and persons "like that," but unlike the case of natural kinds, this conceptual connection guarantees that we are usually right only if we can generally trust our disposition to admiration. It is possible that a community of persons is so radically wrong in its identification of exemplars that even its concept of the good is mistaken. I don't think we need worry about anything analogous in the case of natural kinds, since there is nothing to get right when we point to a material substance or element.

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 good. For the same reason, when we say that a good person is a person like that, and we directly refer to Saint Francis of Assisi, or to Confucius, or to Jesus Christ, we are implicitly leaving open the

---

[7] Unlike water, which is probably water essentially.

question of what properties of Francis, Confucius, or Christ are essential to their goodness. Perhaps there are nonevaluative descriptions of these persons that are sufficient to determine their moral goodness; perhaps not. Perhaps their goodness is not determined by any descriptive properties we know how to apply. The exemplarist approach has the advantage that neither these metaphysical matters nor substantive matters about what makes a person good need be settled at the outset. I am assuming 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.[8] Since narratives are a form of detailed observations of persons, exemplarism gives narrative an important place within the theory analogous to scientific investigation in the theory of natural kinds. Narratives might 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.[9] Furthermore, new empirical research on virtuous exemplars may reveal interesting features of their attitudes and behavior.[10] The theory therefore has a place for both stories and empirical research within its abstract structure.

Are there any historical examples of exemplarist virtue theory? I have suggested that Aristotle's definition of *phronesis* involves an essential demonstrative reference, but there is no indication that persons with *phronesis* play a foundational role in his theory. However, Amy Olberding (2008) argues that Confucius's *Analects* can be read as rooting moral concepts in the experience of paradigmatically good individuals like the Duke of Zhou, heroes of Chinese history, and Confucius himself.[11] If she is right, the conceptual schemata of the *Analects,* including its account of specific virtues, human flourishing, and the path of self cultivation,

---

[8] The parallel point applies to the discovery of the necessary truth, "Water is H_2O." Empirical observation yields the conclusion that water is H_2O, but the judgment that the molecular structure of water is essential to it is *a priori*.

[9] I have not said anything about the difference between fictional narratives and biography. The place of fiction in philosophy is an interesting one, but I am leaving it aside for this essay.

[10] For example, research by Lawrence Walker and Karl Hennig (2004) suggest that there are three distinct types of moral exemplarity: just, brave, and caring. Exemplars have been the subject of other psychological studies, for instance, Kevin Reimer and David Wade-Stein's (2004) work on adolescent exemplars, and research on the participants in the L'Arche communities. Currently, a research group headed by Michael Spezio at Cal Tech is studying the way exemplars play economics games with neuroimaging during the playing of the games.

[11] Olberding is not suggesting that Confucius treated himself as an exemplar. *The Analects* is a compilation of teachings of Confucius, conversations with his students, observations of his behavior by his students, and teachings of his students. The work has several strata in its composition, and Olberding argues that the function of exemplars operates both in Confucius's own teachings and in the way Confucius is treated by his followers.

originate in the experiences of admiration for the figures the text vividly describes. As I said above, there is reason to think that the search for a foundationalist moral theory is a feature of modern Western philosophy, so it would be very surprising if Confucius aspired to such a theory, but I find it interesting that Confucius and his followers may have used an exemplarist approach to explain how the abstract concepts of most significance in Confucian moral practice arise.

5. A Comprehensive Exemplarist Virtue Theory

An exemplarist moral theory does not have to be a virtue theory. There can be exemplary acts and exemplary lives, and possibly exemplary states of affairs, as well as exemplary persons, and any of these could be defined by direct reference. Whether all of these concepts have the potential to be the foundation of a plausible moral theory is another question. If the exemplary is the most imitable, I doubt that there are exemplary states of affairs. There can be exemplary acts, but reference to a few exemplary acts will not be helpful in constructing a comprehensive theory. The idea of beginning the construction of a theory with direct reference to exemplary lives is more promising, but a difficulty with that approach is that only some of the features of a life good as a whole can be imitated. In contrast, admirable persons can be imitated insofar as they are admirable, and if all important moral concepts could be defined by reference to these persons, that would not only give us a comprehensive moral theory, it would allow a smooth connection between moral theory and moral training.

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 assume that our emotion of admiration is generally trustworthy, but I do not assume that we always trust it. When we do, we take the object of admiration to be admirable. A person who is admirable in some respect is imitable in that respect. This is rough because there are many reasons why we do not or cannot imitate the admirable. But the feeling of admiration is a kind of attraction that carries with it the impetus to imitate. The ways in which the exemplar 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 as a way of making ourselves and our lives conform to the admirable.

Here is a suggestion for defining a series of basic moral concepts in terms of a paradigmatically good or admirable person.

> 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* (an act that a person would have most moral reason to do) 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 *duty* (an act it would be wrong not to do) in some set of circumstances C is what the admirable person would feel compelled to do in C in the sense that if he did not do it, he would feel guilty for not doing it.

A *good state of affairs* (more precisely, that subset of states of affairs that can be the outcome of human acts) 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.[12]

In each case, the concept to be defined (virtue, good state of affairs, right act, and so on) 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. A duty is an act a person like that would feel compelled to do, and so on.

Is the theory I have described a type of virtue theory? That depends, of course, upon what it takes to be a virtue theory. Exemplarism does not make the virtues primary, although it does make virtuous persons primary. Perhaps a comparison of the theory with other theories in the diagrams near the beginning of the essay suggest that it is a distinct class of theory, neither act-based nor virtue-based. I don't want to insist that it is closer to traditional virtue theories than it is to act-based theories, but it seems to me that it is more easily combined with work in virtue ethics than with work in consequentialist and deontological theories. However, it probably does not much matter how it is ultimately classified.

In talking about my theory of theory, I mentioned that sometimes a theory distorts what it explains, and moral theory distorts moral practice by not putting the first person pronoun and other descriptive features into the theory. What is right or a duty or a good life *simpliciter* is determined by an act or attitude of a certain kind of person, the paradigmatically admirable person. But it is possible that the exemplar differs from me in ways that affect the way the evaluative concepts we have considered apply to me. Maybe the most desirable life for me is a life no exemplar has ever desired, nor would desire it if she thought about it because its desirability for me is due in part to my idiosyncrasies. Maybe acts that for me are right, even duties, are not acts that any exemplar would do or feel compelled to do because no exemplar is exactly like me and perhaps no

---

[12] I propose variations of these definitions of virtue, right act (permissible act), and a duty in Zagzebski 2004, and the above definition of a good life in Zagzebski 2006.

exemplar has ever been in my circumstances. And if an exemplar were in my circumstances, then perhaps the exemplar would not have most reason to do what I have most reason to do precisely because being an exemplar gives a person reasons for acting that do not apply to the ordinary person, and conversely, the ordinary person may have reasons for acting that do not apply to the exemplar. This is one of the ways that features of persons and situations left out of theory can sometimes be relevant to practice. A theory should give us moral guidance, but it is not primarily a manual. Nonetheless, even when a theory does not generate guidance, it should help us construct a manual that guides us in particular situations.

Let me end by summarizing some advantages of exemplarism. Exemplarist virtue theory has the theoretical simplicity and power of foundationalism without the problems of a conceptual foundation. Perhaps grounding moral theory in the concept of human nature or reason or uncontroversial objects of human desire can succeed, but I doubt it. In any case, the success of those approaches is not so convincing as to make it unnecessary to look for an alternative approach. Exemplarism puts at the foundation of the theory a crucial element of moral practice and, indeed, of moral experience: the identification of persons we admire and whose admirability is something of which we are confident.

Direct reference to exemplars serves another desideratum for an adequate theory mentioned in section 3. It gives the theory a hook that links it to the real world of moral practice—something comparable to a map that says "You are here."

Exemplarism provides a theoretical structure within which the empirical side of ethics can be linked with the traditional *a priori* side of ethics. Lately, the issue of how moral philosophers ought to use empirical research has attracted a lot of attention (see, e.g., Appiah 2008). This theory is one way to do that. Similarly, exemplarism gives an important place to narrative ethics within the structure of the theory. In my opinion this is critically important. Only a tiny percentage of people in the world care about moral theory in the sense I have been discussing, whereas 100 percent of the people in the world like stories. Most moral insights come from stories, but it is the special virtue of the philosopher to organize those insights. As I said at the beginning of the essay, I do not see any reason why there cannot be more than one equally good moral theory, but I think that exemplarist virtue theory has some notable advantages over other types of theory.

Department of Philosophy
University of Oklahoma
455 W. Lindsey Room 605
Norman, OK 73019
USA
lzagzebski@ou.edu

References

Appiah, Kwame Anthony. 2008. *Experiments in Ethics*. Cambridge, Mass.: Harvard University Press.

Gunderson, Keith, ed. 1975. *Language, Mind, and Knowledge*. Minneapolis: University of Minnesota Press.

Kripke, Saul. 1980. *Naming and Necessity*. Oxford: Blackwell.

Olberding, Amy. 2008. "Dreaming of the Duke of Zhou: Exemplarism and the *Analects*." *Journal of Chinese Philosophy* 35:625–39.

Putnam, Hilary. 1979. "The Meaning of 'Meaning'." In *Mind, Language, and Reality*, volume 2 of *Philosophical Papers*. Cambridge: Cambridge University Press.

Reimer, Kevin, and David Wade-Stein. 2004. "Moral Identity in Adolescence: Self and Other in Semantic Space." *Identity* 4:229–49.

Walker, Lawrence, and Karl Hennig. 2004. "Differing Conceptions of Moral Exemplarity: Just, Brave, and Caring." *Journal of Personality and Social Psychology* 93:845–60.

Zagzebski, Linda. 2004. *Divine Motivation Theory*. Cambridge: Cambridge University Press.

———. 2006. "The Admirable Life and the Desirable Life." In *Values and Virtues*, edited by Timothy Chappell, 53–66. Oxford: Oxford University Press.