Exemplarist Moral Theory

Linda Zagzebski

There are people we encounter in history or fiction or in our personal lives whom we find supremely admirable, and who show us the upper reaches of human moral capacity. These people are what I mean by exemplars. They not only reveal what an admirable person is like, but they inspire us to become better persons. Recently, I have been working on a moral theory I call "Exemplarist Moral Theory," or just "Exemplarism," which is a comprehensive ethical theory based on direct reference to exemplars, people we find most admirable. We identify the admirable through our emotion of admiration. Admiration is developed, refined, and altered through experience, including the experience of others whom we trust, and the cumulative experience of admiration in past ages and in past cultures is transmitted to us through stories of exemplars. The set of exemplars forms the basis for a theoretical map in which I define "virtue," "good motive," good life," "duty," and other moral terms by referring directly to exemplars. An advantage of this theory is that it is practically useful as well as theoretically simple and comprehensive. It can be used in moral education, and the body of the theory incorporates empirical studies and narratives, so it does not have the limitations of a wholly a priori project.

I am using the term "direct reference" in the sense that became famous in the 1970s, particularly in the form in which it was used by Saul Kripke and Hilary Putnam to define natural kind terms, or terms for naturally occurring substances or species, like "water," "gold," and "tiger." Briefly, "water" is defined as "stuff like *that*," "tiger" is defined as "creatures like *that*," and so on, where in each case the indexical "that" is used to point to real objects. Direct reference revolutionized semantics because it meant that we succeed in thinking about and talking about objects in the natural

world without needing a descriptive meaning in our heads. People could think about water, ask questions about water, and make assertions about water long before they knew that what makes water *water* is that it is H2O. So the meaning of "water" cannot be "H2O." But the meaning of "water" also cannot be a description that people carry around in their heads like "colorless, odorless liquid that flows in the streams and falls from the sky." We know that because we realize upon reflection that something other than the substance water could have fallen from the sky, could have been in the oceans and streams, could have been the liquid we drink, and so on.

The theory was also revolutionary because of the way it linked empirical science with semantics, and it led to a great deal of work on the social construction of language. The upshot was that we are not connected to the outside world through a mental description. We are (or can be) connected to it directly. What we are talking about when we say "tiger" or "water" or "gold" is determined by observation of something we can pick out through ostension. A meaning is not a description in the head. In fact, a meaning is not a description at all.

This means that the theory of direct reference was *semantically externalist:* the contents of our thoughts and speech when we talk about water, gold, cats, etc. are determined outside of our heads. One way it was externalist is that it maintained that the contents of our thoughts and speech are determined, in part, by the way the world is – what gold and tigers are actually like in a deep way. We find out what makes gold gold and what makes a tiger a tiger by empirical observation. We find out by observation that water is H2O, that gold is the element with atomic number 79, that tigers are animals with a certain biological structure. Kripke argued that there are also superficial, easily observable features of natural kinds that permit users to *fix the reference* of a term, but the experts tell us what the stuff whose reference has been fixed is like in its deep structure, the structure that makes it what it is. It is *because* water is H2O that it has the properties ordinary people use to identify water. It is *because* gold has a certain atomic structure that it has the properties we ordinarily use to identify a piece of gold. The deep physical structure explains the superficial features.The theory was semantically externalist in another way. What we think about and talk about when we use words like "gold" and "tiger" is partly determined by a social network that connects ordinary speakers with the things out there in the world. Putnam proposed a principle he called the Division of Linguistic Labor according to which competent users of a term have different functions in the use of the term. In order to

have properly acquired a term, ordinary people must grasp what Putnam called a "stereotype" of the kind in order to be connected to the network of users of the term for that kind. So to be properly connected to the network of users of the word "gold," people must grasp a description of gold that is vague, but usually roughly accurate. The stereotype of "gold" may be something like "golden-colored metal made into jewelry." The stereotype of "water" might be the description I gave above, probably expanded a bit: "clear, tasteless liquid that flows in the streams and falls from the sky, and which we habitually drink to stay alive." Stereotypes of less common kinds are permissibly shorter and even vaguer. For instance, the stereotype of titanium may not include anything more than being a lightweight metal that the experts refer to by that name. The stereotype is a description, but it does not give necessary and sufficient conditions for being a member of the kind and it is not the meaning of the term. We defer to experts who tell us what the precise conditions are for being a member of the kind, and so the experts have an important semantical function. We ordinary users of natural kind terms intend to refer to whatever the experts are referring to. The experts have the linguistic job of determining the extension of the terms in the domain of their expertise and of determining what the deep properties are that make something a member of the kind. The theory therefore had the attractive feature of smoothly connecting semantics with science. When Putnam extended the theory to many more terms than natural kind terms, it revolutionized the way many philosophers think of the connection between the mind and reality.

There is one other fascinating feature of the Kripke version of the theory of direct reference that I would like to mention for its possible use for moral terms. Kripke argued that there are necessary a posteriori truths. "Water is H2O" is necessary in a strong sense of necessity because it is not possible for anything to be water and not be H2O, and it is not possible for anything to be H2O and not be water. But this truth is discovered empirically, and it certainly was not always known. It is not an analytic truth because someone can understand the word "water" without understanding that it is H2O, yet it is not a mere convention that we do not treat anything as water unless it is H2O. Instead, we think that that is the way the world is. We did not decide to make it that way.

Exemplarism is a theory in which I have borrowed components of direct reference for moral terms. The basic idea is that exemplars are persons *like that*, and we point directly to exemplars of goodness like Confucius, Socrates, Jesus, Gandhi, Nelson Mandela, Holocaust rescuers, Jean

Vanier, or many ordinary people who are known only to a small circle of acquaintances. We find out what makes them admirable by observation, just as we find out what makes water the substance that it is by observation. The observation of admirable persons is obviously a much more complex process than the observation of water since the psychological structure of an admirable person is much more complex than the physical structure of water, and individual exemplars differ from each other much more than individual samples of water. Also, we cannot simply put admirable persons under a microscope (although neuroimaging of exemplars is currently being done). Rather, we observe them through narratives and more recently, through controlled empirical studies. So for exemplarism the place of science in the theory of direct reference is held by narratives as well as ordinary observations and controlled studies. The deep structure of an exemplar is a psychological structure, so psychological structure holds the place of molecular or biological structure in the theory of natural kinds. We find out the motivational structure of exemplars by observation, and that permits us to define basic moral terms like "good trait of character," "good life," "good motive," "right act," and so on by features of exemplars or features of their judgments. We do not need a descriptive meaning for terms like "good person," "good life," "good trait of character," "right act," and the other moral terms, any more than we need a descriptive meaning for natural kind terms.

Direct reference is semantically externalist. I mentioned above that one of its attractive features is that it smoothly connects semantics with science. For natural kinds, the deep physical structure of water or gold, discovered by empirical science, is *what we mean to be referring to* when we say "water" or "gold," and it explains the superficial properties that we use to fix the reference. So, being H20 is both what we mean to be referring to when we say "water," and it explains why water is the colorless, odorless liquid that we drink. Being the element with atomic number 79 is both what we mean to be referring to when we say "gold," and it explains why gold is the golden- colored, malleable metal used to make jewelry.

Similarly, my theory is committed to moral semantic externalism. When we say "good person," the deep psychological structure is what we mean to be referring to when we say "good person," and that structure explains the easily observable behavioral properties that we use to fix the reference of "good person." The easily observable features of a good person are usually overt acts and patterns of acts that we admire upon reflection. We may call them acts of bravery, compassion, generosity, justice, and so on, but the

virtue terms "bravery," "compassion, "generosity," and so on arise from a social history of observation of acts that we collectively admire, with narratives that attempt to identify the deeper psychological features of those persons, so the recognition of admirable acts does not depend upon a prior account of what makes someone admirable. Instead, we can admire an act in advance of knowing what it is we admire about it, and it can take careful observation to uncover the deeper psychological structure of an admirable person. Exemplarism is therefore semantically externalist in the first way I have identified because what we mean when we say "good person" or "admirable person" depends in part on what admirable persons are actually like in their psychology.

Exemplarism is also semantically externalist in the second way I identified above. I mentioned Putnam's Principle of the Division of Linguistic Labor according to which semantic success depends upon being properly connected to a social linguistic network that distinguishes the role of expert in identifying the members of the extension of a term from the role of the ordinary user of the term. Ordinary users are expected to grasp a "stereotype" of the kind in question, but they defer to the experts to identify the term's extension and to give an account of what it takes to be a member of the kind. Similarly, I have proposed a principle I call the Division of Moral Linguistic Labor for moral terms. Ordinary people need to grasp a stereotype of good persons in order to be properly connected to the linguistic network with respect to the term "good person," but they do not need to know what makes a good person good (an admirable person admirable), nor do they need to be able to correctly identify every good person. The stereotype no doubt includes some general descriptions, and is often spread through a linguistic community via narratives. So, for instance, if you ask a person what compassion is, she might give you the story of the Good Samaritan. A difference between Putnam's Division of Linguistic Labor and my Division of Moral Linguistic Labor is that Putnam thinks that ordinary people defer to the scientific experts, and ordinary people succeed in referring to the right thing when they say "elm tree" or "titanium" or "gold" because the experts can reliably pick out elm trees and titanium and gold. In contrast, most people either think they are moral experts, or they think that nobody is an expert. I propose that there is still a division of labor for moral terms, but there are more functions than just ordinary speakers and experts. There are at least four distinct groups which have an important linguistic function in connecting all speakers of moral terms to their extensions. Story-tellers have the function of shaping the stereotypes

of good persons and their virtues and spreading them widely. Philosophers contribute their powers of abstract reasoning to the community, which permits them to analyze stereotypes and to reveal problems of incoherence in them, and to produce arguments explaining and justifying what virtuous persons do, helping to make the network with respect to moral terms clearer. Empirical scientists have the role of finding out how widespread the extension of a virtue term is, how changeable the extension is (whether virtuous persons tend to remain virtuous), and whether there are any connections between one virtue term and another. I agree that there are few if any moral experts recognized by most people in a society, but faith communities have acknowledged authorities, such as the Pope and Bishops of the Catholic Church, and many local religious communities acknowledge their pastor or leader as having a degree of moral authority the exercise of which is a part of that person's function in the community. These different groups of people have a linguistic function in my view because they shape and gradually change the stereotype of a moral term, aid the community in identifying the members of the extension of a term, and can sometimes cause a term to go out of use.

I have said that moral termsdo not have a descriptive meaning, but can be defined by direct reference to exemplars of goodness whom we identify through the emotion of admiration. A list of the main moral terms defined in this way is as follows:

(1) A *virtue* is a trait we admire in an exemplar. It is a trait that makes a person like that admirable in a certain respect.
(2) A *good motive* is a motive we admire in an exemplar. It is a motive of a person like that.
(3) A *good end* is a state of affairs that exemplars aim to bring about. It is the state of affairs at which persons like that aim.
(4) A *virtuous act* is an admirable act, an act we admire in a person like that.
(5) An *admirable life* is a life lived by an exemplar.
(6) A *desirable life* (a life of flourishing) is a life desired by an exemplar.
(7) A *right act* for person A in some set of circumstances C is what the admirable (practically wise) person would take to be most favored by the balance of reasons for A in C.
(8) A *duty* is what persons like that demand of themselves and others.
(9) A *wrong act* is what persons like that demand not be done. It is intolerable.

As I have said, these definitions are not intended to give the content of a series of concepts, but notice also that they are not intended to reveal the "deep" nature of virtue, right action, or a good life. They do not tell us what a virtue, a right act, or a good life *is*, but they give us directions for finding out. They are like defining "water" as "stuff like that," leaving the determination of the deep nature of water for investigation. The purpose of the definition is to identify the reference of the term to make investigation of it possible.

There are different linguistic expectations for the deontic terms like "wrong act" and "duty" and the value terms like "virtue," "good act," "good motive," "good end," and "good life," and the division of linguistic labor differs for the two sets of terms. We have a social obligation to know the members of the extension of the terms "wrong act" and "duty," and the linguistic community is much more demanding of competent users of those terms than of the value terms. The terms "wrong" and "duty" exist because no civil society can survive without agreement about a range of behavior that is critical to the basic functioning of the society. In particular, there are certain acts that we cannot tolerate, and it is crucial that we agree about what those acts are. A speaker who fails to recognize many wrong acts is deemed linguistically incompetent in the use of the word "wrong," and may be called a sociopath. In theoretical ethics, moral terms are associated with concepts that are imbedded in complex and subtle theories, but it is not necessary that individuals have the same *concept* of wrong or duty. All that is necessary is that they agree that acts *like that* should not be done. It does not matter why they think that those acts should not be done. A well-functioning society cannot tolerate theft, but if you ask people why theft is wrong, it does not matter whether they give different answers or no answer. Furthermore, it does not matter whether their behavior is virtuously motivated as long as they refrain from stealing. That means that the stereotype of these terms is exceedingly thin, and the function of the linguistic network with respect to those terms is to make everyone in the community know all of the members of the extension of the terms. Doing one's duty and avoiding wrongful acts does not go very far in giving a person a good life, but it makes society functional. In contrast, the value terms have an intricate connection to the linguistic network and each has a rich and subtle stereotype. The degree of grasp of these terms varies from individual to individual, and the ability of individuals to acquire good motives, good ends, and the virtues depends upon their place in the social network and the individual characteristics of the admirable

persons of their acquaintance. We want as many people as possible in our society to grasp the value terms because emulating the acts and persons in the extension of these terms is the best way for people to lead a life that is both desirable and admirable. But since the deontic terms are critical to avoid social collapse, they tend to get the most attention.

Exemplars help us to acquire higher moral aspirations. I have examined three basic categories of exemplars: the hero, the saint, and the sage. There are many stories about exemplars in these categories, but since some of them are dominant at certain stages of history or in certain cultures, I think we are in danger of losing some of these categories of exemplars in our linguistic networks. Fortunately, there is recent empirical research on all of them. I know of research on Holocaust rescuers, whom I interpret as modern heroes. There is also research on many saintly persons, including people like L'Arche caregivers, who sometimes devote many years of their lives to living in a community with persons who are mentally and sometimes physically disabled. There are also recent empirical studies on wisdom, although only a little of it focuses on particular wise persons, the approach I advocate. However, there is a multitude of narratives on the great wise persons of the past, such as Jesus, the Buddha, and Confucius, as well as some contemporary moral leaders like Chief Plenty Coups, the last great chief of the Crow Nation, who is described as an exemplar of Aristotelian virtue in Jonathan Lear's recent book, *Radical Hope*. One of the important things we learn from exemplars is the variety of good lives. Since we are all different in our talents, personalities, and social situation, we need to spread throughout our communities narratives of many different kinds of exemplars who not only teach us what it means to be moral, but inspire us to emulate them ourselves.

One of the advantages of exemplarism is its connection to a natural method of moral learning through emulation. Much of what we learn is by imitation – how to speak our native language, how to play games and sports, how to cook, how to dance, how to do philosophy. Some imitation is automatic and even subconscious, as when a student picks up a teacher's mannerism, but some of it is conscious and we have some control over it. Emulation is a form of imitation in which the emulated person is perceived as a model in some domain – a model cook, dancer, philosopher. The emulated person might be like James Dean in *Rebel Without a Cause* (1955), a model of the daring teenager. Unfortunately, some teens imitated the game of racing their cars toward the edge of steep cliffs, and some were killed, precipitating classic work on imitation by Albert Bandura and others. Re-

cently, there have been empirical studies of individuals emulating a person rather than an act – someone admired as a whole person rather than a domain-specific role model[1]. What I hope to see is more research on the acquisition of motives and reasons from other people. I have proposed that we can acquire motivating emotions by emulation of admirable people, and these motives can also justify behavior, but we cannot acquire reasons that are propositional beliefs by emulation. We can acquire beliefs from other persons, but by a different process than emulation.

I intend my exemplarist virtue theory to be a philosophical framework for studies in many fields. It has a simple theoretical structure that is philosophically comprehensive. It is designed in a way that gives a place for empirical work and narratives in the structure of the theory. It permits different versions for different communities, including faith communities, but it can also facilitate cross-cultural discourse through investigation of the overlapping sets of admirable persons in different cultures. It is constructed with the purpose of inserting the motive to be moral into the theoretical structure. This is a significant advantage because so often we hear complaints that moral philosophy does nothing to make people moral. I believe that admiration is one of the most significant of the human emotions. The cognitive side of the emotion has the potential to generate a conceptual framework. The affective side of the emotion moves us to emulate the admirable and become better persons. Exemplarism broadens the reach of moral philosophy by creating a structure that encompasses many aspects of our moral practices besides the theoretical. I believe that our societies are morally healthier when moral philosophers integrate their work with the work of scholars in other fields and with the narratives that shape the culture. This theory is my contribution to that effort.

Works cited

Bandura A., Ross D., Ross S. (1963), *Imitation of film-mediated aggressive models*, in «Journal of Abnormal and Social Psychology», 66, pp. 3-11.

Kripke S. (1980), *Naming and Necessity*, Blackwell Pub, Oxford.

[1] See, for example, some of the projects funded through the Moral Beacons project, directed by William Fleeson at Wake Forest University, http:// www.moralbeacons.org. The initiative is funded by a grant from the Templeton Religion Trust.

Lear J. (2008), *Radical Hope: Ethics in the Face of Cultural Devastation*, Harvard University Press, Cambridge (Mass.).

Putnam H. (1975), *The Meaning of 'Meaning'*, in «Minnesota Studies in the Philosophy of Science», 7, pp. 131-193; reprinted in H. Putnam, *Mind, Language, and Reality* (*Philosophical Papers*, vol. 2),Cambridge University Press, Cambridge 1975.

Zagzebski L. (2017), *Exemplarist Moral Theory*, Oxford University Press, New York.

Abstract

This paper summarizes my new moral theory, which is based on direct reference to exemplars of goodness, identified through the emotion of admiration. Since a motivating emotion is at the root of the theory, it is intended to serve both the theoretical purpose of mapping the main moral terms by reference to features of exemplars, and the practical purpose of making us want to act morally and showing us how to do so through emulation of exemplars. The theory links the a priori side of ethics with empirical work in psychology and neuroscience, and it gives narratives a key function in the theory. Since it tracks a natural process of moral development in the emulation of exemplars, it also connects with moral education.

Keywords: moral theory; goodness; admiration; ethics; moral education.

Linda Zagzebski
The University of Oklahoma
lzagzebski@ou.edu