

Linda Zagzebski

Virtue Theory and Exemplars

This essay outlines an approach to virtue theory that makes the foundation of the theory direct reference to virtuous exemplars, modeled on the famous theory of direct reference, devised in the seventies by Hilary Putnam and Saul Kripke. The basic idea is that exemplars are persons like that, just as water is liquid like that, and humans are members of the same species as that, and so on. In this theory exemplars are picked out directly through the emotion of admiration rather than through the satisfaction of a description. We discover the virtues empirically by investigating the qualities of exemplars in a way that parallels the discovery that water is H₂O. It is also possible that although the virtues are discovered empirically, the connection between being admirable and having certain traits is necessary, just as Kripke claims that “water is H₂O” is necessary, but known a posteriori.

In recent decades, the topic of the virtues has regained the prominence it once had. Everyone agrees that virtues are good traits of a human person, but there is controversy over the way in which a virtue is good. Many writers focus on the goodness of virtues in the sense of the *desirable*. Virtues are allegedly good *for* their possessor. They benefit their possessors, contributing to their living a life of well-being, either as a means to the end of well-being, or as a constituent of well-being¹. This approach is modeled on that of Aristotle. In addition, writers on virtue often follow Aquinas in arguing that virtues are desirable for other people as well as for the possessor of the virtues. They are desirable because they benefit both virtuous persons themselves and other people.

On the other hand, some writers stress that virtues are good in the sense of the *admirable*. That is the view I endorse. What we admire is not the same as what we desire, and similarly, what is worth admiring is not identical with what is worth desiring, I believe it can be shown that admirable traits are desirable to have, but many things are desirable that are not admirable – e.g., long life, health, friendships, financial means, enjoyment. One of the foundational questions for a virtue theory of ethics is the way virtue connects with a life that is good in the sense of de-

¹ See, for example, Rosalind Hursthouse’s book, *On Virtue Ethics*, Oxford University Press, Oxford 2002.

sirable, and with such important moral concepts as a right act and a good outcome. In this brief essay I will outline a type of theory I call exemplarist virtue theory, which makes the foundational move in the theory direct reference to exemplars of admirability. All other basic moral concepts are defined by reference to exemplars – a virtue, a desirable life, a right act, and a good outcome.

In the 1970s Hilary Putnam and Saul Kripke devised a theory of the reference of natural kind terms that produced a revolution in semantics². Leaving aside variations in the 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 has the advantage of explaining how “gold” referred to the same thing before and after the discovery of the atomic structure of gold.

One of the exciting features of this proposal was that it meant that competent speakers of the language can use 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.

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 can then discover the nature of water by observation. Under the assumption that the molecular

2 S. Kripke, *Naming and Necessity*, Blackwell, Oxford 1980, and H. Putnam, “The Meaning of ‘Meaning’”, in *Mind, Language, and Reality*, Philosophical Papers vol. 2, Cambridge University Press, Cambridge 1975; first published in *Language, Mind, and Knowledge*, edited by K. Gunderson, University of Minnesota Press, Minneapolis 1975.

structure of water is essential to it, it follows that certain necessary truths such as “Water is H₂O” are discovered *a posteriori*.

This idea can be used in the construction of a moral theory³. 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 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 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 both 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 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 theoretical structure to simplify and explain 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 do not 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 conceptual connection between good persons and those we identify as good – good persons are persons like that. And the parallel point holds for water.

3 I first proposed a theory of this type in *Divine Motivation Theory*, Cambridge University Press, Cambridge 2004, and outlined it in “Exemplarist Virtue Theory,” *Metaphilosophy* 41: 1-2 (2010), pp. 41-57.

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 what 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 be settled at the outset. That is, we need not start by assuming that certain traits are the virtues or that certain acts are right. 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. 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. For instance, it might be necessary that a good person is generous, that a good person is compassionate, and that a good person is fair. It also follows that empirical research on virtuous exemplars may reveal interesting features of their attitudes and behavior. The theory therefore has a place for both stories and empirical research within its structure.

To summarize, 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 like our other emotions, we should not trust it unless it survives conscientious reflection, including reflection upon the emotional responses of other people, but when we do trust it, 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 specific acts and specific persons who are admirable. But the feeling of admiration is a kind of attraction that carries the impetus to imitate with it. 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 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* 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 concept to be defined (virtue, good state of affairs, right act, etc) 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 desirable 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.

Exemplarist virtue theory has a number of advantages. It is simple, comprehensive, and I think it gives us understanding of the moral life, but that is for others to judge. In addition, since admiration is motivating, the motive to be formed morally in the likeness of exemplars is at the foundation of the theory. The theory includes an emotion that leads to the practical aim of moral improvement within the structure of the theory, as well as having the theoretical advantages of aiding understanding of the moral life. Another advantage is that it links the traditional *a priori* side of ethics to empirical psychology. There is a small but growing body of research in psychology and neuroscience on the features of moral exemplars⁴, on imitation or mimesis⁵, and on moral emotions. These empirical studies have not previously been connected as far as I know, nor have they been linked to a theory of ethics. The theoretical structure I am proposing here gives plenty of room for empirical investigation of exemplars in a way that is analogous to the role of scientific observation in the Putnam-Kripke theory of natural kinds, and it permits both secular versions and versions for faith communities, such as a Christian version of ethics based on the imitation of Christ. From a strictly philosophical viewpoint, this type of theory has the important advantage that substantive issues about what a good life is, what the virtues are, and which acts are right and wrong do not have to be settled at the outset, although common agreement on these issues is part of the pre-theoretical moral practice the theory seeks to explain and justify.

Theoretical ethics is enjoying a renaissance in virtue ethics, but there is a need for new ways to structure a virtue theory. I think there can be many good ways to structure a moral theory, and the way we evaluate one over another depends upon which of the many aims of theory we consider most important at any given time. Some theories are more theoretically illuminating than others. Some theories are more practically useful than others. Some do a better job than others of integrating into the theory components of our moral practices other than moral judgments and acts – for instance, moral emotions. Some theories do a better job than others of integrating empirical research into the way the theory is constructed. Exemplarist virtue theory does a better job than most other theories in all these respects, but

4 See, for instance, L.J. Walker and K.H. Henning, "Differing Conceptions of Moral Exemplarity: Just, Brave, and Caring." *Journal of Personality and Social Psychology* 86:4 (2004), pp. 629-647.

5 See J.A. Herdt, *Putting On Virtue: the Legacy of the Splendid Vices*, University of Chicago Press, Chicago 2008, chap. 1. Another example is M. Iacoboni, *Mirroring People: the New Science of How We Connect with Others*, Farrar, Straus, and Giroux, N.Y. 2008.

I would not deny that there can be more than one good moral theory, and I do not intend this theory to be a competitor to neo-Aristotelian theories. It seems to me that we are better off as philosophers and as moral educators if we try out different approaches to the project of constructing a moral theory that gives virtue a central place, and test them for both their theoretical and their practical benefits.

Linda Zagzebski
University of Oklahoma
lzagzebski@ou.edu

Linda Zagzebski è George Lynn Cross Research Professor di Filosofia e Kingfisher College Chair di Filosofia della religione ed Etica presso la University of Oklahoma. È stata Presidente della Society of Christian Philosophers e della American Catholic Philosophical Association. Ha tenuto numerosi corsi e conferenze. Si ricordano qui: le Romanell Lectures presso la Phi Beta Kappa (2005), le McCarthy Lectures presso la Pontificia Università Gregoriana a Roma (2006), le Wilde Lectures in Natural and Revealed Religion a Oxford (2010), e le Kaminski Lectures presso l'Università Cattolica Giovanni Paolo II di Lublino, Polonia (2012). I suoi libri includono *The Dilemma of Freedom and Foreknowledge* (Oxford University Press, 1991), *Virtues of the Mind* (Cambridge University Press, 1996), *Divine Motivation Theory* (Cambridge University Press, 2004), *Philosophy of Religion: An Historical Introduction* (Blackwell, 2007) e *On Epistemology* (Wadsworth, 2008). Molti i libri e gli articoli dedicati alla epistemologia delle virtù, alla filosofia della religione e alle virtù etiche. Il suo libro più recente è *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief*, in corso di pubblicazione presso la Oxford University Press.