EXEMPLARIST VIRTUE ETHICS

An Undergraduate Research Scholars Thesis

by

ADAM GOIN

Submitted to the Undergraduate Research Scholars program
Texas A&M University
in partial fulfillment of the requirements for the designation as an

UNDERGRADUATE RESEARCH SCHOLAR

Approved by
Research Advisor: Dr. Linda Radzik

May 2016

Majors: Philosophy
Economics

TABLE OF CONTENTS

ABSTRACT

Exemplarist Virtue Ethics

Adam Goin
Department of Philosophy
Texas A&M University

Research Advisor: Dr. Linda Radzik
Department of Philosophy

In his *Nicomachean Ethics*, Aristotle introduces the concept of virtue, states of character that

allow an agent to perform her function well, and gives a practical account of how someone can

become virtuous. I will argue that Aristotle manages to be vague with respect to two epistemic

questions: First, how can we identify virtuous people? Second, how can we know *which* states of

character are virtuous? Recently, Linda Zagzebski has introduced a moral theory called

exemplarism, which answers that we may identify virtuous people via the emotion of admiration,

and that by studying virtuous people, we may come to know which states of character are

virtuous. But, Zagzebski's exemplarism is unmediated; there is no difference between a moral

concept and what an exemplar would do or feel in certain circumstances. Problematically, on this

account it appears that the more moral experience we have, the murkier our moral concepts

become. In what follows, I propose what I call a mediated exemplarism, an account that answers

our questions about virtue in the same way as Zagzebski, but on which an exemplar's actions or

feelings do not constitute moral concepts. Rather, on my account, exemplars serve the

pedagogical purpose of indicating virtues to an agent, who may then construct a theory of virtue

from which moral concepts may be inferred that are divorced from the actions and feelings of a

particular exemplar.

CHAPTER I

ARISTOTLE

Aristotle gives us an account of eudaimonia and virtue, as well as a practical means by which to achieve both, but his account leaves something to be desired in that some theoretical questions are left open. In this chapter, I will give a rough sketch of Aristotle's account of virtue, and ask Aristotle two epistemic questions: First, how can we identify a virtuous person? Second, how do we know *which* character states are virtues and which are not? I will try to answer these questions from Aristotle's perspective, but I think that the best answers available to him will be somewhat unsatisfying.

1.1 Eudaimonia and virtue

To describe Aristotle's theory of virtue, I find it easiest to start at the top. Aristotle has a teleological metaphysics, one in which everything has a function. As he puts it, "every action and decision, seems to seek some good" (NE I.1 1094a1-2). Aristotle traces a regress of intermediate ends and arrives at one ultimate end — eudaimonia. Eudaimonia is complete without qualification; "we always choose it because of itself, never because of something else" (NE I.7 1097b2). Thus, eudaimonia is the true final end; it is never sought as an intermediate end to something else, but instead is sought always for its own sake in every lower end that is chosen. Further, in its completeness, eudaimonia is self-sufficient: "all by itself it makes a life choiceworthy and lacking nothing; and that is what we think happiness does" (NE I.7 1097b15-17). So, eudaimonia is the complete and self-sufficient end for which all other ends are sought. But how do we become eudaimon?

First, Aristotle tells us that certain external goods are necessary for eudaimonia, which include certain instrumental goods (e.g. wealth and political power), and certain goods without which we "do not altogether have the character of happiness [e.g.] if we look utterly repulsive or are ill born, solitary, or childless" (NE I.8 1099b1-7). Eudaimonia also requires internal goods of the soul, particularly reason and virtue. For Aristotle, there exist two types of virtue: virtues of thought (e.g. wisdom) and virtues of character (e.g. temperance). While virtues of thought arise and grow mostly from teaching, virtues of character, which will be the focus of the following discussion, result from habit (NE I.13 1103a15-20). Further, Aristotle tells us that "the human good proves to be activity of the soul in accord with virtue" (NE I.7 1098a17-18). So, we might say that external goods are necessary, but not sufficient for eudaimonia, and that the good that is aimed at is activity of the soul in accord with virtue.

Now, in order to understand eudaimonia, we need an account of virtue. According to Aristotle, a virtue is "the state that makes a human being good and makes him perform his function well" (NE II.6 1106a23-25). That is, a virtue is *not* right action or a disposition to behave in a certain way, but a state.[1] Further, virtues are ruined by an excess or a deficiency, and are preserved by the mean. For example, a deficiency of abstinence makes one intemperate, while an excess of abstinence makes one a teetotaler; the virtue of temperance exists at the mean (NE II.2 1104a13-27). Virtues of character, Aristotle tells us, are built by habit, but may also be ruined by it, in the same way that "building well makes a good builder, and building badly makes bad ones (NE II.2 1103b11). Virtues of character, then, are states which promote excellence. They are separate from individual excellent acts, but these excellent acts are the means of habituating a person to

---

[1] I take a 'state' to be both more stable and active than a mere disposition, although being in a state of virtue probably requires a disposition to act virtuously. I think that feelings and activity distinguish virtuous states from 'mere' dispositions to act virtuously; a continent person probably possesses a disposition to perform virtuous actions, but is not virtuous because of her feelings toward her actions. That is, she does not do them in the right way, for the right reasons, etc. (NE II.6, 1106b 19-24).

reach a state of virtue. We should also keep in mind that excellence and goodness here are not 'moral' terms in the same way that good and evil are; they are instead functional terms. In the same way that what makes a watchmaker good is that she makes good watches, and what makes a watch good is that it tells time well, good and bad here express functional concepts. So, we might say that a particular watch is bad insofar as it is designed to tell time but it does not do so accurately, or that a watchmaker is bad insofar as she makes bad watches. But from what has been said we can see that Aristotle thinks of humans as having functions as well; excellent human functioning is flourishing. So, we should expect virtues of character for Aristotle to be states of character that are built by habitually performing virtuous (excellent) actions, and we should see that these virtues, states of character, are a necessary component of flourishing, although they are not sufficient for it on their own.

Importantly, virtues of character are related not only to fine actions for Aristotle, but also to feelings. Aristotle takes someone's pleasures or pains to be a sign of his state: "For if someone who abstains from bodily pleasures enjoys the abstinence itself, he is temperate; if he is grieved by it, he is intemperate" (NE II.3 1104b6-8). So, regarding virtues of character, performing the action relevant to the virtue, which is in this case abstinence, is not enough. The agent must also *feel* the right way about his action. Aristotle summarizes: "Virtue is about pleasures and pains; the actions that are its sources also increase it or, if they are done badly, ruin it; and its activity is about the same actions as those that are its sources" (NE II.3 1105a15-17). Aristotle explains further:

> We can be afraid, for instance, or be confident, or have appetites, or get angry, or feel pity, and in general have pleasure or pain, both too much and too little, and in both ways not well. But *having these feelings at the right times, about the right things, toward the right people, for the right end, and in the right way, is the intermediate and best condition*, and this is proper to virtue. Similarly, *actions also admit of* excess, deficiency, and *an intermediate condition*. (NE II.6, 1106b19-24, emphasis mine)

Now we are in a position to see how Aristotle manages to be crystal clear and frustratingly vague at the same time, for he has drawn for us a very neat circle. On his account, humans have a function; that function is to flourish. Of the goods required for flourishing, one of the most interesting and important is virtue. Virtue is the state which makes a person function well; the function of a person is to flourish, and we have come full circle. This becomes particularly vexing if we want to ask Aristotle direct questions about virtue. If we ask "What is the actual state of temperance," Aristotle replies that "actions are called just or temperate when they are the sort that a just or temperate person would do. But the just and temperate person is not the one who [merely] does these actions, but the one who also does them in the way in which just or temperate people do them" (NE II.4 1105b7-9). So we have been given a second circular answer. The temperate person is the one who performs temperate actions in the way that temperate people do. Although this answer is thoroughly frustrating to the contemporary philosopher, Aristotle does not seem to be concerned, saying that "in some cases it is enough to prove rightly that [something is true, without also explaining why it is true]" (NE I.71098b2-3). So, it appears that if we are interested in knowing *why* Aristotle's account is true, we might be on our own.

What I hope to have done so far is to have given enough of a sketch of eudaimonia to understand the role that virtue plays in Aristotle's system, and to have given a rough sketch of Aristotle's view about virtues of character. I also hope to have set these concepts up in a way that illuminates the value of these concepts while showing that Aristotle leaves something to be desired in the way of explanation; his account, which he admits to be a sketch, does not seem to be sufficiently filled out.

1.2 Questions about virtue and Aristotle's responses

Given the above account of virtue, I find two epistemic questions to be particularly pressing for Aristotle. First, how can we identify virtuous people? Second, how can we know which character traits are virtuous? Call the first question the identification question, and the second the differentiation question.

The identification question, which asks how we can identify virtuous people, seems very pressing if we are interested in actually becoming virtuous. If what it is to be temperate is to act temperately as the temperate person does, then it is crucial that we be able to identify the temperate person to enter the circle at some point. If we read into Aristotle a bit, we find two answers to the identification question, but both prove to be unsatisfying. His first answer is that our moral education allows us to identify virtuous people: "[V]irtue of character is about pleasures and pains. . . . That is why we need to have had the appropriate upbringing—right from early youth, as Plato says—to makes us find enjoyment or pain in the right things; for this is correct education" (NE II.3 1104b9-11). So, presumably, if we have been well-raised, we know for ourselves what things to take pleasure in and what things to take pain in, and so can identify the person who is feeling pleasures and pains correctly from that. But this account does a poor job of answering the identification question for three reasons: First, even if we are temperate and know the feelings proper to temperance, it seems difficult to identify those feelings in others. Second, if Aristotle points to the temperate person to look toward if we want to become temperate, appealing to moral education and stipulating that the agent already knows temperance to begin with does nothing to help us come to know temperance. Third, it seems very difficult for a person to know whether or not she was well-raised; she may believe that she is well-raised and be in error about the nature of temperance.

The second answer that Aristotle gives us is that virtuous people are praiseworthy:

> For such [mixed] actions people are sometimes actually praised, whenever they endure something shameful or painful as the price of great and fine results. If they do the reverse, they are blamed; for it is a base person who endures what is most shameful for nothing fine or for only some moderately fine result. (NE III.3 1110a20-24)

From this discussion, it seems that Aristotle thinks that virtuous people are praised, while non-virtuous people are blamed. So, if the virtuous person is praiseworthy, we may look to the person who is praised for his temperance to identify the temperate person. There is something to this answer, and it is certainly more satisfying than the first, but it is not without its problems. If Aristotle indeed thought this, then at best we know that all virtuous people are praised. However, it does not follow that all praised people are virtuous; praiseworthiness is not sufficient for virtue. So this second answer is left wanting, since we do not know how we might distinguish between people who are praised wrongly and people who are praised because they are truly virtuous.

The second question, the differentiation question, which asks how we can know which states are virtuous, seems equally pressing, since we want to pursue virtues that are components of eudaimonia. Aristotle gives a rough account of identifying virtues; he tells us that they exist at a mean:

> Virtue, then, is a state that decides, consisting in a mean, the mean relative to us, which is defined by reference to reason, that is to say, to the reason by reference to which the prudent person would define it. It is a mean between two vices, one of excess and one of deficiency. (NE II.7 1107a1-4)

So, it seems we can say that any feeling or action admits of excess and deficiency, and the mean is always virtue. But this is not right either, for some actions and feelings "automatically include baseness—for instance, spite, shamelessness, envy among feelings, and adultery, theft, murder, among actions" (NE II.7 1107a10-13). But why do some actions and feelings admit of a mean

while others do not? This opens up a wide range of questions we might ask of Aristotle's account. Can we expand Aristotle's account of virtue to include things such as forgiveness as virtues, or are we limited to the virtues that Aristotle discusses? Are the things that Aristotle discusses actually virtues?[2] What we are searching for is an *epistemic justification* for virtue, beyond identifying something that exists at a mean between excesses; that is what is meant when we ask how we can know which states are virtuous, and Aristotle does not seem to have any answer at all.

In any case, Aristotle's answers to these two epistemic questions are unsatisfactory. As it is, this poses a problem if someone is interested in becoming virtuous, but cannot identify virtuous people to learn from, or if someone is unsure of whether the virtues that they strive toward are actually virtues at all. For his part, Aristotle does not seem interested in answering these sorts of questions claiming that "the purpose of our examination is not to know what virtue is, but to become good" (NE II.2, 1103b26-27).

---

[2] In NE I.7 1098a16-18, Aristotle himself seems to express doubt as to whether more than one virtue actually exists, which could indicate that Aristotle may not be entirely certain which states of character are virtues. Since *Nicomachean Ethics* is a collection of lecture notes, it may be that Aristotle was building up toward a more robust account of virtue. But in any case, it seems reasonable to doubt at least whether Aristotle's list of virtues is exhaustive, so it is a good idea to search for a justification for virtue.

CHAPTER II

EXEMPLARISM

With our two questions in hand, it seems that exemplarism can provide a simple, powerful answer. A virtuous person can be identified through the emotion of admiration; the character traits for which a virtuous person is admired are virtues. In this chapter, I give an account of Zagzebski's exemplarist theory and use it to answer the questions we posed to Aristotle in Chapter I. However, it looks like Zagzebski's account has some problems lurking in the background *if* she is interested in bridging a gap between people that individuals admire and people that are universally admired. More seriously, it seems that on her account, the more moral experience we have, the murkier our moral concepts become.

2.1 Zagzebski's view

Whereas other moral theories begin with a conceptual foundation, like reason in Kant or the principle of utility in Mill, an exemplarist moral theory is one that begins with a direct reference to exemplars of moral goodness (Zagzebski, 49). Linda Zagzebski's exemplarism has its roots in the Putnam-Kripke theory of direct reference, so I will take the time to briefly explain the latter.

Zagzebski focuses on the Putnam-Kripke theory as it refers to natural kind terms, which refer to the kinds of things found in nature, like gold and water. What is interesting about the theory is that without understanding the nature of the referent, we can construct a working definition that fixes the referent and then sharpen that definition until it links up with its nature. Zagzebski explains:

> gold is, roughly, whatever is the same element as *that*, water is whatever is the same
> liquid as *that* . . . . 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. . . . 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. (Zagzebski, 50)

Zagzebski's exemplarism runs parallel to the Putnam-Kripke theory. On her account, "[g]ood persons are persons *like that*, just as gold is stuff *like that*" (Zagzebski, 51). Presumably, then, after we have picked out the good persons, we can posit accounts of the good that track whatever it is that makes our exemplars good. An advantage of this sort of model is that we can still successfully refer to good persons even if we are unsure what it is that makes them good. Further, in the same way that we might discover that we were mistaken in referring to pyrite as gold, we might also discover that we were mistaken in referring to some persons as good (Zagzebski, 51).

Zagzebski's exemplarism is less a moral theory and more a pretheoretical framework in which to ground a theory. It begins with direct reference to an exemplar, who is identified through the emotion of admiration, which carries with it "an impetus to imitate."

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. . . . 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. (Zagzebski, 54)

Zagzebski then suggests translating some standard moral concepts into exemplarist terms, giving us a means of accessing them that only requires a capacity to experience the emotion of

admiration instead of the bulky conceptual frameworks that are usually employed to get the

concepts off the ground:

> 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. (Zagzebski, 55)

So, Zagzebski's exemplarism can be summarized in the following way:
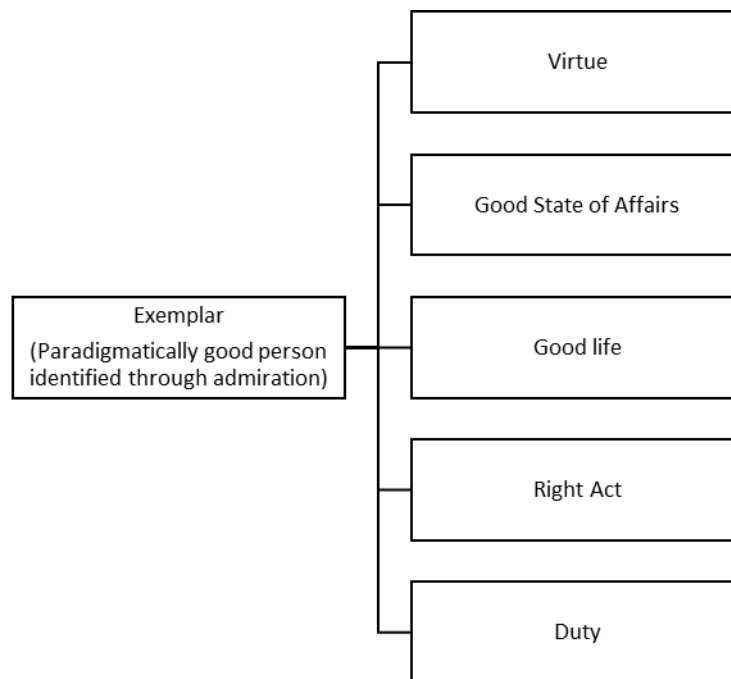
(Figure 1: The Structure of Zagzebski's Exemplarism)

The emotion of admiration indicates imitability; those who are the most admirable and therefore

the most imitable are exemplars. On this account, being the most imitable is sufficient to be

paradigmatically good. With this paradigmatically good person in mind, we can discover the

traits we admire in our exemplar (virtues), and by imagining how our exemplar would act in

various situations, Zagzebski thinks that we can infer a good state of affairs, good life, right

action, and duty.

Importantly, Zagzebski has given us a refreshingly simple way of answering the two questions we posed to Aristotle earlier. If we are interested in how to identify virtuous people, Zagzebski tells us explicitly that we can identify virtuous persons via the emotion of admiration. With respect to the differentiation question, which asks how we can know *which* states are virtuous, Zagzebski answers that "[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" (Zagzebski, 54). The statement "I admire George Washington because he is honest" gives an answer to both the identification question and the differentiation question simultaneously. It picks out an exemplar, a person who is virtuous in some respect, via the emotion of admiration, which gives an answer to the identification question. Then, the agent is in a position to reflect and answer the further question "Why do I admire this person" with "Because he is honest," pointing to honesty as a virtue and providing an answer to the differentiation question. We might see this as an extension of Aristotle's reply that the temperate person is the person who does temperate actions in the way that temperate people do (NE II.4 1105b7-9). Now the emotion of admiration indicates someone who is admirable, further reflection shows that the person is admirable because she is temperate; what exactly temperance is and why it is admirable is something that can be learned by studying the exemplar. Aristotle himself also seems comfortable with exemplarism, saying that "we also praise the wise person for his state, and the states that are praiseworthy are the ones we call virtues" (NE I.13 1103a9-11).

2.2 Potential problems

Zagzebski's view seems to provide some new insights into Aristotle's system, however if we press her a little harder on what she really believes exemplars to be, I think that some important problems come to the surface. First, Zagzebski is (perhaps intentionally) vague with respect to

what she means by an exemplar, and depending on how we read her, this may mean that some problems are lurking in the background for her theory.[3] I have two general objections to Zagzebski's theory. The first is that, depending on whether Zagzebski intends for exemplars to be universal or not, she may encounter some problems. I do not think that this objection is serious so long as we take care to avoid stepping where the problems lurk. The second objection I call the problem of multiple exemplars, and it seems to cause serious problems for Zagzebski if we are using her theory for moral guidance.

First, it is unclear whether Zagzebski is talking about universal exemplars (i.e. the most admirable/imitable for everyone) or personal ones (i.e. the most admirable/imitable for me). I think that from what has been said about Zagzebski's account, we can certainly say that she has a justified account when it comes to personal exemplars. But if she is intending the *most imitable* or *most admirable* person to be universal instead of personal, it is unclear how she can bridge the gap between personal feelings of admiration and what the "right" feelings of admiration are. She has a framework by which to do so, telling us that "[w]e 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" (Zagzebski, 52). Perhaps, if everyone has an educated sense of admiration, then our judgments will coincide and there will be an overlap between who is the most admirable for me and who is the most admirable for everyone. However, this seems unlikely. Perhaps Zagzebski thinks that some people simply *are* the most imitable, but in this case it is unclear what this person would look like, since it is not apparent that anyone is universally admired, let alone the *most* admired. In any case, if Zagzebski is working toward a

---

[3] I do not mean that she is intentionally vague in a derogatory sense; I take her to be advancing a framework for a host of potential exemplarist theories without giving any one particular account. If this is her objective, it is natural for her to be intentionally vague with respect to some concepts.

universal account (and I am not sure that she is), then she has some work to do both in showing that the gap can be closed and in picking out a strategy for closing it.

I take this potential problem to be pressing for Zagzebski, because it means that as her theory stands, she does not have a justified account of an exemplar who is universally the most admirable, and she is limited to personal judgments. This is not a problem if she is only interested in virtue concepts, but Zagzebski is interested in using her exemplarism to access other moral concepts, like duty, right action, and a good state of affairs.[4] All of these ideas seem to lose their grip if they are not universal; a right action that is only a right action for me seems to fall somewhat short of what it is typically meant by right action. I do not think that this threatens exemplarist virtue ethics, but it is a very real problem if someone is advancing a comprehensive exemplarist moral theory, like Zagzebski seems to be doing, or if someone is advancing a deontic or consequentialist theory that is exemplarist in its foundation. Because I am interested only in exemplarist virtue ethics, I am not interested in solving this particular problem, and am happy to set it aside as a problem for other types of exemplarist theories.

The second problem is what I call the multiple exemplar problem, and it stems from the fact that Zagzebski's theory is founded in unmediated direct reference to an exemplar. For Zagzebski, there is no difference between the moral concept that we infer and what an exemplar would feel or do in certain circumstances. A duty, for example, is what an exemplar would feel guilty for not doing in a certain situation, Zagzebski tells us (Zagzebski, 55). Zagzebski talks about

---

[4] Virtue concepts, i.e. virtue and the good life, do not seem to be strictly universal in the way that right action is. In the same way that Milo the wrestler needs to eat more than the novice athlete, which virtues people need to develop, and how these virtues manifest themselves, and what constitutes a good life seems to have a fair amount of variance from person to person (NE II.6 1106b1-7). That is, even if courage winds up being a virtue for everyone, the political activist needs courage and in different ways and amounts than the soldier or the philosopher. So virtue concepts seem to do just fine in the realm of the personal, while the consequentialist and deontic concepts that Zagzebski is also interested in seem to flounder outside of the realm of the universal.

circumstances possibly being different for exemplars and agents, presumably trying to pre-empt objections where what an exemplar would feel guilty for not doing is clearly not a duty, but I think a far more serious problem is lurking here.

It seems like an obvious feature of standard moral practice that we can have multiple exemplars. Zagzebski herself gives a list of three: Jesus, Francis of Assisi, and Gandhi (Zagzebski, 52). So perhaps there is a class of people who are the most admirable, or perhaps each is the most admirable in a certain respect. Regardless, if we can have multiple exemplars, and there is no difference between what an exemplar would feel or do in certain circumstances and the moral concept itself, then as we gain more exemplars, our idea of the moral concepts that we are interested in potentially becomes more and more incoherent. That is, if there is nothing to mediate exemplars and moral concepts, then a duty is what *all of my exemplars* would feel guilty for not doing in certain circumstances. Problematically, our idea of duty shrinks very quickly as our pool of exemplars grows, and the concept of duty threatens to evaporate completely when exemplars disagree. I want to separate the effects of this on deontic and consequentialist concepts (i.e. right action, duty, good state of affairs) and on the virtue concepts (i.e. virtue, good life), because I think that they are very different.

Regarding deontic and consequentialist concepts, this objection seems to carry more weight, because the moral concepts with which we are concerned are absolute; there is no room for two duties or right actions to contradict one another. Consider, for example, a small set of exemplars containing Gandhi, Malcolm X, and Nelson Mandela. It seems plausible that someone could think that these three men are the most admirable in some respect, and if these are our exemplars, we should be able to infer right action from what these men would do given a set of circumstances. But on the question of protest, it appears that Gandhi would advocate nonviolent

protest, Malcolm X would advocate violent protest, and Nelson Mandela might advocate either, depending on whether we admire Nelson Mandela earlier or later in his life. But certainly violent protest and nonviolent protest cannot *both* be right action in the same situation. Regarding these non-virtue concepts, incoherence is a big problem because these concepts are thought to be universal.

Regarding virtue concepts, this objection is still serious, but less so. Consider again our set of three exemplars containing Gandhi, Malcolm X, and Nelson Mandela. If we are interested in virtue, each of the three may be said to be admirable in a different respect. If all three are exemplars for the same virtue, e.g. courage, they might show different ways of acting on the same virtue, and may each be virtuous in their own way. Recall that Zagzebski's account is unmediated; a virtue is just what makes an exemplar admirable in a certain respect. But if all three are said to be courageous, then the more exemplars, and thus the more moral experience we have, the murkier our moral concepts become. This less serious objection is not unique to virtue concepts; it applies equally well if we try to infer deontic or consequentialist concepts from an exemplarist theory. The problem, at bottom, is that because Zagzebski's account is unmediated, we are limited in inferring our moral concepts to what all of our exemplars would agree on. Thus, as our pool of exemplars grows, what is agreed on by all exemplars shrinks and threatens to disappear completely. It seems to be a large flaw for a theory that it gives less coherent accounts of moral concepts and less guidance the more moral experience we have.

So, Zagzebski's theory is valuable in that it is simple and powerful, and answers the questions that we earlier posed to Aristotle well, but it runs into some problems because it is unmediated; there is nothing between what an exemplar would do or feel in certain circumstances and a moral concept. So, in what follows, I will try to build an exemplarist theory of virtue that is mediated

16

and manages to answer the questions that we raised about Aristotle's theory while avoiding the

potential problems we have found in Zagzebski's.

CHAPTER III

AN EXEMPLARIST THEORY OF VIRTUE

Grounding a theory in direct reference to an exemplar is useful to get a theory going, but Zagzebski's exemplars seem to do too much work for her. Her theory begins and ends with direct reference to an exemplar, and so problems arise when we have two or more exemplars. If the theory is never separated from direct reference to the exemplar herself, it cannot move beyond what an exemplar would have do in a particular situation. Further, it is intuitive that more than one exemplar exists. But if we are limited to what a particular exemplar would do in a particular situation, then the more exemplars we have, the less our exemplars can agree upon. Problematically, on this sort of view, the more moral education we have, the less we are able to say. In this chapter, I advance a revised account of what an exemplar is, one that I think manages to answer both questions that we posed to Aristotle in Chapter I while avoiding the multiple exemplar problem. Then, I continue to give an account of a theory of error that might help *epistemically justify* our answers to the differentiation question.

3.1 Exemplars

Zagzebski's idea of an exemplar does the work that she needs it to do, but her account seems needlessly restrictive. In this section, I suggest some changes in our understanding of the term "exemplar" to include a wider range of emotions in picking out our exemplars, namely a distinction between positive and negative exemplars, a distinction between strong and weak exemplars, and the possibility that exemplars might be fictional. These improvements are not motivated by any particularly strong objections, but I think it is clear that the theory is more

informative after our idea of an exemplar has been revised in these ways, and I do not believe that these moves open us up to any serious new objections.

For Zagzebski, exemplars are identified through the emotion of admiration. But it appears that we do not always admire the entire person. Rather, we identify character traits that are admirable, and say that an exemplar is admirable insofar as she possesses one or more of these character traits. This is highlighted in the way that we implicitly exclude certain features of exemplars. Someone who asks "What would Jesus do?" is not suggesting that you grow a beard, wear robes, or perform miracles. That is, they are not referring to the agent in his entirety; they expect you to identify the relevant virtuous character trait and act accordingly. If George Washington is our exemplar for honesty, the questions "What would George Washington do?" and "What would an honest person do?" seem to be asking the same thing. What is important for picking out exemplars, then, is not that the whole agent is admirable or the most admirable, but that the agent possesses one or more character traits that are admirable. The agent is admirable *only in virtue of* his possession of virtuous character traits. If an exemplar were to be admirable for something other than virtue (e.g. beauty), it seems he would not be an exemplar even though he is admired.

Since we have realized that it is the character trait of the exemplar that is admired and not the exemplar in his entirety, we might wonder whether admiration is the *only* emotion that can indicate an exemplar to us. Zagzebski has a somewhat technical understanding of admiration, saying that it carries with it the "impetus to imitate" (Zagzebski, 54), but it appears that insofar as we are interested in becoming virtuous, and insofar as we have positive sentiments toward the character of someone else, we can be said to have an impetus to imitate that positive character trait because we are interested in developing ourselves and have found something that we take to

be good. We can treat anyone toward whom we have positive sentiments as an exemplar and learn something from him. Thus, I think that we ought to expand our account of an exemplar to include any positive sentiment that is found to have its origins in a character trait that an exemplar possesses, although admiration seems to be the easiest of these sentiments to work with.

It seems clear that the primary role that exemplars play is pedagogical; they teach virtue by example. So far, we have talked about what I will call *positive exemplars*. These are exemplars in a traditional sense; they show a sort of excellence in an area that we approve of and may wish to imitate. But we also learn from people who show deficiencies that we do *not* wish to imitate. I will call these *negative exemplars*, and where we may find virtues by reflecting on our admiration of positive exemplars, we infer vices by reflecting on our disapprobation for negative exemplars. I take both to be equally important to moral education: If Jim is our positive exemplar for honesty, and Tom is our negative exemplar for dishonesty, then imitating Jim and avoiding imitating Tom can both give important guidance. We needn't privilege the learning gained from one over the other. But the inclusion of negative exemplars somewhat stretches the meaning of the term 'exemplar', which has historically been charged with aretaic overtones. I do not think this is problematic as long as the distinction is clear.

Another implication of our understanding of an exemplar as playing a primarily pedagogical role is the inclusion of fictional exemplars to our account. If the role of an exemplar is a pedagogical one, that is, if they are the means of moral learning, from whom we may learn both good and bad states of character, then there is no reason to exclude fictional exemplars from our theory. We may learn just as easily from thought experiments, parables, fairy tales, folklore, literature, and other fictional accounts as we can by observing a real person's actions.

This leads me to one last distinction between strong exemplars and weak exemplars. Surely if we include fictional characters in our account of exemplars, we shouldn't exclude everyday people from our account, since we appear to be capable of learning from them as well. Traditional exemplars are both strong and positive. Lists of these people might include Jesus and Socrates; they exhibit many virtues and few (if any) vices. If taken as exemplars for only one virtue, strongly positive exemplars seem to actually *be* virtuous and not simply act in accordance with virtue.[5] Strongly negative exemplars, on the other hand, might include individuals such as Hitler and Iago, who exhibit many vices and few virtues, or actually *be* vicious instead of acting in accordance with vice.[6] In the middle are weak exemplars, people who may be a mixed bag of virtues and vices, or may perform virtuous actions without actually being virtuous (i.e. continent individuals). I take it that most of our moral learning comes from people in this bracket, simply because we have more exposure to these people than we do to strong exemplars. Surely if my friend does something honest, I may learn from her example as easily as I might learn about honesty by reading about George Washington and the parable of the cherry tree. I am exposed to actions in accordance with virtue, which indicate virtues to me, by everyday people who are not virtuous far more often than I am exposed to virtuous actions by someone who is an exemplar in a traditional sense.

So, to recap, whether an exemplar is positive or negative depends on whether what is shown to the observer is a virtue or a vice, and whether an exemplar is strong or weak depends on the observer's *judgement* of how strongly the exemplar actually exhibits the relevant virtue or vice.

---

[5] Aristotle thinks that acting in accordance with virtue is necessary, but not sufficient for being virtuous; *being* virtuous means acting in the right way, toward the right end, at the right time, with the right feelings, etc. (NE II.6 1106b19-24).

[6] Here the parallel between acting in accordance with virtue and being virtuous probably does not hold with acting in accordance with vice and being vicious; perhaps we can say that strongly negative exemplars intentionally act viciously, as opposed to being akratic or trying to be virtuous and failing.

Importantly, being shown that honesty is a virtue from a weak exemplar (e.g. from a friend who accidentally stole a library book and confessed) does not have any bearing on the value or importance of the virtue itself; the virtue that is revealed is the same whether it comes from a strong exemplar or a weak one. Further, we also learn as much from our exemplars if they are fictitious as we do if they are real. The parable of the cherry tree, for example, is likely fictitious, but still serves to illuminate honesty as a virtue. Shakespeare's Iago was not real, but we may still learn from his example as if he were, and we may easily dream up a myriad of thought experiments that show vices to us. So, *an exemplar*, as I will use the term, *is the means by which we learn about virtue and vice*. Specifically, exemplars exhibit character traits that, in being admired or disapproved of, reveal themselves as virtues or vices.

Now we have a clearer answer to both the identification question and the differentiation question. How can we *identify* virtuous people? We identify them when we feel positive moral sentiment. How can we know *which* states of character are virtues? When we reflect on *why* we feel sentiments toward exemplars and find their origins to be in the character of an exemplar, we are also shown a virtue or a vice.

3.2 Virtues

In section 2.2, I discussed the problem of multiple exemplars. I said that the problem stemmed from Zagzebski's account being an unmediated one; because there is no difference for her between what an exemplar would do or feel and moral concepts, the moral concepts shrink when our pool of exemplars grows. Since we now have a more robust account of exemplars, we are in a position to understand a mediated exemplarist theory of virtue. In short, my account of how virtues are found is as follows: After an exemplar has been identified by our sentiments toward him, if we want to find the source of our sentiments, we might ask ourselves *why* we feel

admiration or disapprobation toward an exemplar and *reflect* on our feelings. When reflection

illuminates a character trait that is hypothesized to be the source of admiration or

dissappribation; this is posited as a virtue or a vice and is added to the agent's personal theory of

virtue. It is this step, in which an agent chooses to add a virtue or vice to their personal theory

that makes this account a mediated one.

Talking about a personal theory of virtue might sound like a far cry from the Aristotelian theory

that we are interested in supporting, so I ought to start by clarifying a few features of personal

theories of virtue as I understand them. First, I do not take Aristotle to be giving a *justification*

for virtue, or a *theory* of virtue at all. I take him instead to be giving a practical account of how

persons can become good, and I take him to give advice regarding virtue insofar as it helps a

person to become good, and I take him to understand himself in this way as well.[7] How we can

identify virtuous people and how we can know *which* states of character are virtues are the

questions about the epistemic justifications for Aristotle's theory that I take our exemplarist

theory to be answering, and it seems obvious from Aristotle's answers to these questions that a

sort of realist stance toward virtue does a poor job justifying itself, even if it gets the answer

right. Instead, if we are to do some work toward giving an epistemic justification for virtue by

answering the identification and differentiation questions, it must be an open question which

virtues and vices (if any) exist; we must take the differentiation question seriously. So, how do

we get from it being an open question whether there are any virtues and vices at all to an account

of which states of character are virtues? As we have seen, our understanding of what an

exemplar is tells us how to identify virtuous people if there are any, and gives us a rough account

---

[7] Aristotle says that "[o]ur present discussion does not aim, as our others do, at study; for the purpose of our examination is not to know what virtue is, but to become good" (NE II.2 1103b 28-30). Likewise, he tells us that virtues are concerned with actions and feelings and an intermediate condition, and then goes on to give a laundry list of feelings, actions, and their means. To me this seems to be a practical account; Aristotle does not seem interested in giving a justification for virtue in *Nicomachean Ethics*.

of which character states are virtues—we theorize that they are the character states that our exemplars have, and update our theories constantly as we engage in moral practice and gain more moral experience.

It seems that this creates a problem, however, because when Aristotle gave us a laundry list of virtues as part of his practical account, he only offered us virtues that were consistent with one another. If we are interested in answering how we can identify virtuous people, and which states of character are virtues, we no longer have a guarantee that our account of virtue is consistent, so we run the risk of making ourselves worse off. We might have three general worries about how this account might make us worse off: (1) In following our exemplars, we might come up with a theory of virtue that is internally inconsistent, e.g. we might simultaneously believe that courage is a virtue that is best exemplified charging into battle and killing an enemy, and also believe that compassion is a virtue that prohibits taking human life. (2) We might come up with an internally consistent, but wayward theory of virtue. If someone has an internally consistent account of virtue that features cruelty as a virtue, we might be interested in changing her mind if we disagree. (3) We might be concerned that an agent is insincere with respect to the virtues that she has adopted, and so has not gained anything from our system.

To hedge against these worries, I want to introduce a coherentist framework to provide guidelines for creating personal theories of virtue:

The Three-Part Coherence Test:

1) From the perspective of each of the other accepted virtues and vices, the proposed virtue is at best approved of and at least not in conflict with other virtues and vices.
2) The agent's set of virtues and vices, her personal theory of virtue, ought to be at least not incompatible with the personal theories of other agents whose moral opinions the agent respects.

3) From the perspective of the agent reflecting on her accepting a trait as a virtue, she does not feel negative moral sentiment toward herself for accepting it.

The first point amounts to a constraint that an agent's theory of virtue is internally consistent. We have already seen the example of an agent who thinks that courage entails killing people in battle, but who also wants to be compassionate, as one example of incoherence. If a virtue fails this constraint, that is, if the proposed virtue is in conflict with the theory of virtue and vice that the agent already has, the agent can resolve the conflict in two ways: First, she can revise one or more of her concepts of virtue to render them consistent. Second, she can reject one or more virtue concepts to resolve the tension.

The second point accounts for external corroboration. If we are concerned that an agent might commit an error in choosing her virtues, we need to appeal to the mechanisms by which our judgments are informed by the judgments of others; we can argue and change one another's mind if we are in disagreement. It seems that everyone has someone whose opinion they respect regarding moral matters, whether it is a friend, family member, an exemplar, or someone else, and if our personal theories of virtue are inconsistent with the theories of those whom we respect, it constitutes another form of dissonance. The agent can resolve this second kind of dissonance by revising her evaluation of her peer as someone whose moral opinion she respects, by revising her own theory to be more consistent with her peers, or by convincing her peers to revise their theories to be more consistent with hers. I take this to be a fairly uncontroversial account of the sorts of conflicts between people that happen in everyday situations.

Of course, we might still share Zagzebski's concern that a whole community of persons could become totally wrong in their conception of the good. I think that the second constraint of the coherence test allows for avenues to remediate such situations (e.g. social reform), but it does not

stop such problems from occurring in the first place. I do not think that this is problematic. After all, cases like this *do* occur (e.g. The People's Temple cult); a theory that cannot explain how these sorts of occurrences arise cannot remediate them.

The third point is less powerful than the first two; it boils down to a constraint that the agent be sincere in her revisions to her theory. If, for example, an agent is convinced to renounce compassion as a virtue by the opinions of peers whom she respects, but feels guilty for doing so, it appears that she is not fully convinced and so needs to reflect further on whether to take compassion to be a virtue or whether her peers are actually worthy of respect in this case.

To review, Zagzebski's account gave satisfying answers to epistemic questions about virtue that Aristotle could not. However, because Zagzebski's account is unmediated, having multiple exemplars for one virtue potentially muddies the notion of the virtue that we are trying to discover. Thus, we needed something between the traits for which we admire our exemplars and the traits that we take to be virtues. To achieve this, we have broadened our account of what an exemplar is and now understand them as a means of *showing* the agent a *candidate* for virtue, whereas on Zagzebski's account they were what *makes* a virtue. Now, the agent has to reflect and decide for herself whether the thing that she has identified by her sentiments is really a virtue or not; this is the way in which our account is a mediated one. However, we might be concerned that an agent could go wrong in her deliberations about virtue; it seems that at minimum her theory of virtue, if it is a good one, ought to be at least coherent. Introducing a coherentist framework to give guidelines when deliberating about virtue provides one means of resolving conflicts and errors that come up in the deliberative process.

Notice that on this account, the multiple exemplar problem is not a problem at all, because exemplars play a much smaller role in this theory than they did in Zagzebski's. Whereas on

Zagzebski's account, recognizing that a person was highly admirable *constituted* and identification of a virtue. Here, the role of an exemplar is merely to *show* a *candidate* virtue to an agent, who then must decide whether they accept this virtue or not. Thus, on this account, having multiple exemplars for one virtue means being having lots of data points with which to inform an agent's idea of a particular virtue.

In this essay, we have identified two epistemic questions that seem to be important for Aristotle, and found answers for them that seem to line up with Aristotle's thinking in an exemplarist theory. However, we saw that unmediated exemplarism runs into problems when faced with multiple exemplars for one virtue, and so have changed our understanding of an exemplar to be someone who shows candidate virtues to an agent, who then constructs her own personal theory of virtue from the data points that she finds in her sentiments toward exemplars. Finally, we have given some guidelines for resolving conflicts in creating these personal theories.

REFERENCES

Aristotle. 1999. *Nicomachean Ethics.* (Trans. T. Irwin). Indianapolis, IN: Hackett
     Publishing Company.

Zagzebski, Linda. 2010. "Exemplarist Virtue Theory." *Metaphilosophy* 41:41-57.