4 Model Exemplars

Eric Yang

Introduction

Many of us want to become morally better people (or so I would hope). We may also think of ourselves as being fairly decent at picking out morally good people, i.e., exemplars that we admire or want to imitate. If asked to provide some candidate exemplars, several of us would likely name real people, such as Josephine Bhakita or Confucius. Some may name fictional characters, such as Lucy Pevensie or Samwise Gamgee. We also typically include people who are not well known, such as family members or close friends.

According to an exemplarist moral theory, good people are the foundation of our moral understanding. We come to know what good character traits are or what right conduct (in a particular circumstance) is based on our knowledge of exemplars. As evident from the examples given, we often take real individuals to play the role of such exemplars.

There are, however, a couple of problems with including real people as exemplars. First, while we may know quite a bit about some of our exemplars, we may not know enough about them to know how to imitate them in certain circumstances, especially in situations that are quite unlike the ones that the exemplar may have faced. Let us call this the "*Ignorance Problem*." Secondly, there have been many occasions where someone has selected an exemplar only later to find out that this person is morally vicious or has engaged in morally reprehensible behavior. It was not too long ago that many people would have identified Mahatma Gandhi or Jean Vanier as exemplars. In light of recent information and evidence about their behavior, their status as exemplars have been revoked or at least rendered suspect by many people. Let us call this the "*Bad Exemplars Problem*."

In this chapter, I begin by laying out a version of exemplarism, and I develop the *Ignorance Problem* and the *Bad Exemplars Problem* in greater detail. Rather than giving exemplarism up, I offer a solution to these problems by presenting a view that advocates construing exemplars as models of real (or fictional) people. One understanding of models, as employed in some domains such as the sciences, is to take them as intentionally distorted representations of some portion of reality. Models are simplifications and

DOI: 10.4324/9781032648392-6

idealizations, and so models are different from the things they represent. According to my view, exemplars are (or should be) models of people in this sense. My aim is to motivate this view by showing that a model-theoretic approach to exemplarism avoids both the *Ignorance Problem* and the *Bad Exemplars Problem*.

Exemplarism and Admiration for Real People

Exemplarism treats good people as fundamental in moral theorizing. Right or wrong actions, good or bad motives, and virtues or vices are analyzed in terms of exemplars. For example, an action is morally right in a particular circumstance just in case a good person would take that action to be most favored by the balance of reasons in that circumstance, and a virtuous trait is one we admire in a good person (Zagzebski 2010, 54).

An exemplar can be identified without providing any necessary or sufficient conditions for what makes someone a good person, or even without providing any particular description that a good person must satisfy. To defend this claim, Linda Zagzebski utilizes the theory of direct reference as developed by Putnam (1979) and Kripke (1980), according to which natural kinds can be denoted through some initial dubbing or some other reference-fixing act such as pointing and using a demonstrative indexical. We can pick out water by pointing to it and saying, "that is water." And we do not have to know that the thing we are pointing to is comprised of H_2O or that it is colorless, odorless, potable, etc.

Similarly, Zagzebski avers that we can do the same with exemplars. We can point to someone and say, "Good persons are persons like that" (Zagzebski 2017, 15). Moreover, we do not randomly point at just anyone when we say that. What enables us to pick out exemplars is the emotion of admiration. We admire good people because they are "imitably attractive" (ibid., 43). Admiration is a fallible guide, and so we can be mistaken about who we pick out as being a good person. So our identification of exemplars is revisable (ibid., 16). However, we do not have to take our emotions of admiration at face value. After careful reflection, we may discover that someone we admire is not worthy of being so. Even though we may make mistakes about those we identify as exemplars, our admiration may be justified when it survives conscientious reflection (ibid., 50).

Merely listing facts about someone may neither lead to admiration of that individual nor help us realize they are a good person. We may admire people that we observe, but we can also admire someone when we learn about them through stories. We can acquire relevant data concerning characters in narratives, by which we may come to admire that figure. While I cannot now observe Confucius, I can glean data about Confucius when reading the stories or mini-vignettes in which he is involved in the *Analects*, whereby he is presented as "the most complete and compelling exemplar" (Olberding 2012, 105). Since we may identify exemplars through narratives, it may not

matter to our moral development whether the story is fictional or nonfictional, and hence we can take real or fictional characters as exemplars since we can admire them and find them worth imitating (Zagzebski 2010, 51). By reading stories about Josephine Bhakita, I find myself admiring her and wanting to imitate her in order to inculcate her character traits. It is thus easy to suppose that my exemplar is Bhakita. However, when we take real people as our exemplars, we run into some trouble.

Ignorance and Bad Exemplars

The *Ignorance Problem* for exemplarism may be construed as a specific version of the action-guidance problem for virtue ethics. As the latter problem goes, some normative theories, such as deontological ethics and utilitarianism, are able to offer specific prescriptions of moral behavior in particular situations. Is the act non-universalizable or does it treat someone else as a mere means? If yes, then don't do it (says the Kantian). Would the action lead to a consequence that would increase the total aggregate of happiness? If yes, then do it (says the utilitarian). However, the virtue ethicist encourages us to cultivate virtuous character traits such as courage, prudence, temperance, and the like. And we are supposed to act in some circumstance in the way that a virtuous agent would act in that very circumstance. But we do not always know how a virtuous agent would act in some circumstances, and sometimes virtuous agents can act in competing or conflicting ways in a particular circumstance.

Exemplarism appears to have an advantage over bare-bones virtue ethics insofar as we can identify specific exemplars and specific qualities and traits of those exemplars. Whereas I may not know whether the generic "virtuous agent" would or would not push a person of sufficient mass and density off the bridge to stop an oncoming train that would kill several people, I can say with a lot more confidence that Martin Luther King Jr. would not have done so. However, my knowledge of many exemplars will be considerably limited, and so situations may arise where one does not know exactly how to imitate that person.

One way of imitating someone is to observe directly what they are doing and to copy it as closely as one can, yet many exemplars are not directly observable. Another way is to learn about their personality and character from narratives, especially for exemplars that are no longer living or are not within the observable vicinity. But if we are going to imitate an exemplar, then we need to be able to situate them in our context, facing a circumstance that is pressing to us. To do so, we need to be able to ascertain the truth-value of various counterfactuals relevant to understanding the character of the exemplar.

So begin with a subjunctive conditional schema:

1 If S were in circumstance C, then S would perform action A in C.

Since I regard Bhakita as my exemplar, I should plug "Bhakita" in for S, and I should substitute a particular circumstance I face in for C. One regular

situation I face is being in a classroom teaching a course entitled "Philosophy and Science Fiction" to students at Santa Clara University. So I can fill in the schema as follows:

> 2 If Bhakita were teaching a course entitled "Philosophy and Science Fiction" to students at Santa Clara University, then Bhakita would perform_____when teaching a course entitled "Philosophy and Science Fiction" to students at Santa Clara University.

I still cannot assess the truth value of (2) because the schema in (1) has not been completely filled. But how do I fill in the blank in (2) given what I know of Bhakita? Despite the several narratives I have read of her life, I have no idea how to fill it in, let alone for many other potential candidates for the substitution of S.

Now perhaps the blank in (2) cannot be filled in if the action is too narrowly described (e.g., "Bhakita would teach a section on time travel…"), but we may be able to fill it out at a more coarse-grained level. We can say that she would be compassionate and empathetic (especially for struggling students) in that environment. The problem, however, is that real people are complex. Many times, our expectations or predictions of how someone will behave are mistaken, and this may be the case not only when someone is acting out of character but also because we may not understand or appreciate the full depth of their character. People's personalities and character traits may be more nuanced and subtle than we realize, and so people may be acting in character and yet in unpredictable or unexpected ways due to our limited understanding of their character. The lack of knowledge of that nuance may lead us to fill in the blank incorrectly. Moreover, if we fill in the blank in a way that is too coarse-grained, then the counterfactual may be trivial or unhelpful for imitation (e.g., "Bhakita would act in a way that a good person does"). Our ignorance of real people, then, leads us either to fail to fill in the blank, to fill in the blank incorrectly, or to fill in the blank in a trivial way that is unhelpful for the practice of imitation. So goes the *Ignorance Problem*.

Yet there is another worry for exemplarism: the *Bad Exemplars Problem*. Good people are picked out by reflective admiration, i.e., admiring someone who is regarded as admirable after conscientious reflection. But as noted above, this method is fallible, and so we may still select a bad exemplar through this procedure. We need not completely reject a procedural method merely because it is fallible. However, there are some reasons to be worried about admiration. We are naturally drawn to some individuals, even at a young age, where certain types of people become attractive to us. The admiration of someone may be so strong that an open-minded or critical stance will not be psychologically possible for some people. Some cult leaders have been extremely charismatic, to the point that even after their misdeeds or secret lives have been revealed, their followers continue to defend them and seek to be like them or follow in their paths. To be sure, this need not give us reason to reject reflective admiration entirely as a helpful heuristic in

selecting exemplars. But the problem remains that such a heuristic may permit too many bad people to be regarded as exemplars.

Bad people may be regarded as exemplars because of hidden or unknown facts about them. Some people lead secret lives, as has been revealed about some people who would have been regarded as exemplars prior to the discovery of their secrets. Now exemplarism takes exemplars as that which grounds other moral concepts. And we use those exemplars to plug in for the subject of the relevant counterfactuals. But if we are plugging in the actual person in the counterfactual judgment, then it may turn out that many of our counterfactual judgments about that person are false if we have selected a bad person as an exemplar.

For example, suppose I take someone named "Pretendsy" to be an exemplar because he appears to be generous and compassionate even though he leads a double life and engages in morally reprehensible behavior in secret, including embezzling. In aiming to imitate Pretendsy, suppose I form the judgment that if Pretendsy were in my shoes, then Pretendsy would give a sizeable amount of my income away to charitable organizations. However, taking a possible world semantics for counterfactuals, the nearest possible world (or nearby possible worlds) is one where Pretendsy is only pretending to give money away while being engaged in embezzlement and profiteering. If Pretendsy is my exemplar, then my counterfactual judgment is false, since I would not thereby be imitating Pretendsy, since giving my money away is not what Pretendsy would do if he were in my circumstance.

Reflective admiration as a tool to pick out exemplars cannot account for those who lead secret lives engaging in morally reprehensible behavior. This prevents genuine imitation, at least of the sort that seeks to be more than mere copying behavior but also aims at cultivating relevantly similar motives or character traits of the exemplar. So if we take a bad person as our exemplar, many of our counterfactual judgments involving them turn out to be false (since we are mistaken about what that person would do in the circumstance we are considering). But we might instead take the construct of who we thought the bad person was as our exemplar. In that case, our counterfactual judgments turn out to be true. However, the real person no longer counts as the exemplar. Perhaps this is the direction we should be pursuing.

Model Exemplars

Exemplars are, in a sense, models or paradigmatic examples of a good person. But there is another sense of "model" that I want to bring to bear on this issue, in particular an understanding of models that come from the sciences. In this sense, models are simplified and typically idealized representations of some target object (Potochnik 2017). For example, a map is a model of some geographical location. Some models may be abstractions of the target domain, as a map may leave out many of the features of some

particular region, depending on what the purpose of the model is. If the map is trying to show how drivers should navigate various freeways in some cities, many of the topological boundaries and ecological features may be left out, focusing solely on the structures of the freeway. Moreover, some models may not be abstractions but rather deliberate distortions of the represented target, where such distortions are ineliminable or beneficial given the purpose of the model (Batterman 2002; Rice 2019).

In the sciences, models are not typically measured by their accuracy as a representation. Rather, models are evaluated based on whether they are *adequate for a purpose*. A two-dimensional map may not be as accurate of a representation as a three-dimensional model of some geographical location, but a two-dimensional map may be adequate for the purpose of helping a driver navigate around the city. Some models may be deliberately distorted such that we might even regard them as false models of some target reality, and yet such models may be useful in arriving at more plausible theories or explanations about some object or target domain (cf. Bokulich 2009, 2011, 2016; Elgin 2009).

Epistemologically, distorted models, while involving some falsehoods, may yield epistemic goods such as understanding. One way of analyzing understanding is to construe it as a non-propositional cognitive grasp of the structure of some domain (Zagzebski 2001). For example, someone can understand how the scenes of a narrative fit together or how a musical composition thematically hangs together. Someone may read a book many times, grasping each individual sentence, but perhaps after reading it for the umpteenth time, they declare, "Now I get it!" when they comprehend the structural framework that enhances their appreciation of the narrative as an aesthetically sophisticated piece of literature (this happened to me after reading Jane Austen's works over and over again). In the sciences, the falsity of posited physical laws (e.g., the ideal gas law) can make understanding the behavior of some physical phenomena more tractable, allowing for greater explanatory and even predictive power (Elgin 2006). Much more can be said about models and the epistemic and scientific benefits of distorted models, including false models. But what has been said about models should be adequate to offer a modification to exemplarism.

The view I propose is as follows:

Model Exemplarism: all of our immediate exemplars are (or should be construed as) models of real or fictional people.

A few remarks on *Model Exemplarism*. Many people ordinarily state a real person as an exemplar, e.g., Jesus, Confucius, Bhakita, etc. A view may be regarded as counterintuitive if it takes people as being mistaken about whom they identify as their exemplars. So I do want to grant that Confucius can be regarded as someone's exemplar, but only indirectly. What *Model Exemplarism* denies is that Confucius (the real person) is someone's

immediate exemplar. To be an immediate exemplar is to be a substitution candidate in counterfactual judgments such as (1) when one employs such judgments for the aim of imitation. So Confucius can be an indirect exemplar for someone by virtue of someone having a model that represents Confucius as their immediate exemplar. Moreover, a model of a real person can help attain an understanding of that person's character, even if that model involves distortions or some false propositions concerning the real person (just as a biography of a person can capture the authentic personality of a historical person even if there are hyperbolic or false statements made about that person).

Now Xunzi may believe that the real Confucius is his immediate exemplar. However, when he is trying to imitate Confucius, he will have to fill in the blank of (1). But the subject of (1) should be Xunzi's conceptual model of Confucius. After all, Xunzi does not grasp the entirety of Confucius' psychology let alone the nuance and complexity of anybody's psychological profile. Models are distinct from the target reality they represent. So Xunzi's model of Confucius is distinct from the real Confucius. Now Xunzi may think his concept of Confucius is identical to Confucius, but this would be mistaken, for he may believe that Confucius would act in a certain way in a certain circumstance when the real Confucius would have perhaps acted in a different way. In virtue of Xunzi's model of Confucius being a representation of the real Confucius, we can admit that Confucius is one of Xunzi's (indirect) exemplars. However, the immediate exemplar for Xunzi is the model of Confucius that Xunzi has in mind, for it is that model that Xunzi will use to plug in for the subject in the relevant counterfactual judgments.

Someone may resist and insist that the real person is their immediate exemplar. If they do so, then the *Ignorance Problem* and the *Bad Exemplars Problem* can be raised again to show why that will not work. And so for such individuals, we can take *Model Exemplarism* to be a normative thesis such that we ought to take our models of (real or fictional) people as our immediate exemplars.[1]

As has already been stated, the primary advantage of *Model Exemplarism* is that it provides a way of answering the *Ignorance Problem* and the *Bad Exemplars Problem*. The *Ignorance Problem* arose because we often know too little to make reliable counterfactual judgments involving real people. But a model of a real person is a simplified conceptual representation. My simplified understanding of Bhakita and her character allows me to plug that conceptual model into a counterfactual schema such as (1). Given the model of Bhakita, I may take it that in my SCU class, she would ensure that I allot time for those students who are likely reticent to speak up given their social location to be able to give voice in the discussion. Would the real Bhakita do that? I believe so, but I'm not sure. But I don't have to be sure. My model of Bhakita would do that, and so now I have some guidance regarding how to act in a particular circumstance. Of course, my conceptual model of Bhakita can be refined as I learn more about the real

Bhakita through biographical narratives, just as models can undergo revision given their relationship to the extant data as well as to the acquisition of new data.[2]

Model Exemplarism also has the resources to avoid the *Bad Exemplars Problem*. We might and sometimes do make false judgments about who a good person is, especially since they may be living a secret life that involves morally reprehensible behavior, or they may have begun as morally good people but later became morally vicious. But if real people need not be the exemplars we use as the immediate object of our imitation, then we can avoid imitating morally bad people as well as ensure that our counterfactual judgments regarding our exemplar are true. For we may be imitating a model of a morally vicious person, where the model itself would count as morally good or virtuous given the qualities and traits that belong to the model. The vicious or bad traits of the real person need not be attributed to the model—indeed, they will not be attributed to the model in those cases where we do not know that the person is living a secret life. Models are not identical to the target reality they represent; the model of a bad person is not identical to that bad person. But it is possible for the model not to be bad even when the person the model is representing is in fact bad since it is an abstraction or distortion of the target object. Models do not include all the features of the target reality. The model may also be picked out in the way that exemplarism standardly prescribes, viz., reflective admiration. We may find the conceptual model of some individual admirable even under conscientious reflection.

Now suppose the real person that the conceptual model is representing is hiding morally pernicious behavior. That does not prevent the model from counting as admirable, even if the real person is not admirable. So even if Pretendsy is morally bad or morally vicious, the model of Pretendsy I have developed based on my knowledge of him may be a legitimate immediate exemplar worthy of imitation, for that model is not identical to the real Pretendsy. As long as we do not conflate or confuse the model of the person with the real person (e.g., assuming that the real person perfectly satisfies the description of the model), the *Bad Exemplars Problem* is avoided.

Model Exemplarism can even explain why it is that people who are trying to imitate the same exemplar sometimes make different judgments about what they should do or what their exemplar would do in a particular circumstance. Two people may both take Martin Luther King Jr. as an exemplar, but both may prescribe different courses of action in a particular situation, where one person believes King would act in one way and the other person believes that King would act in an incompatible way. What would explain this deviation is that both have different models of the same person, and their distinct conceptual models account for their differences in judgment and prescription. Both models may count as immediate exemplars, and depending on how one handles moral dilemmas, one may suppose that both actions are morally permissible (cf. Hursthouse 1998).

Conclusion

In order to avoid the *Ignorance Problem* and the *Bad Exemplars Problem*, I recommend that exemplarists embrace *Model Exemplarism*. It may seem surprising to treat all of our immediate exemplars as models, but it seems that many of us already do so without explicitly acknowledging that the person we are actually imitating is the conceptual model we possess of some person (whether real or fictional). There is a sense in which we can still say we are imitating a real person, but we imitate them indirectly by imitating our model of that person. Yet for those still insisting that they are directly imitating a real person, the *Ignorance Problem* and the *Bad Exemplars Problem* should have them reconsider who their immediate exemplar really is. Exemplars are models in the imitative sense. I hope to have shown that exemplarists should also regard exemplars as models in the representational sense.[3]

Notes

1 Our model of a fictional person is also not identical to the fictional person it represents (granting that fictional objects exist), since there may be a mismatch between what the model would do and what the fictional person would do, which perhaps is determined by what the author(s) would claim the fictional person would do in such a circumstance (and if fictional objects do not exist, there can still be a mismatch between the model and what the author claims the fictional person would do in such a circumstance).

2 There is a complicated relationship between models and data, and naïve approaches that take there to be unfiltered data from which models are to be construed should be rejected in favor of approaches that take seriously a model-data symbiosis (cf. Bokulich 2020, 2021; Parker 2020).

3 Many thanks to Meilin Chinn, Kimberly Dill, Erick Ramirez, Pilar Lopez Cantero, and Brother John Baptist Santa Ana for helpful feedback. I am also grateful to the SET Foundations and its director, Meghan Page, for their support of my research on the use of models in the sciences.

References

Batterman, Robert. 2002. *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*. Oxford: Oxford University Press.

Bokulich, Alisa. 2009. "Explanatory fictions." In *Fictions in Science: Philosophical Essays on Modeling and Idealization*, edited by M. Saurez, New York, NY: Routledge.

Bokulich, Alisa. 2011. "How scientific models can explain." *Synthese* 180: 33–45.

Bokulich, Alisa. 2016. "Fiction as a vehicle for truth: Moving beyond the notice conception." *The Monist* 99: 260–279.

Bokulich, Alisa. 2020. "Towards a taxonomy of the model-ladenness of data." *Philosophy of Science* 87: 793–806.

Bokulich, Alisa. 2021. "Using models to correct data: Paleodiversity and the fossil record." *Synthese* 198: 5919–5940.

Elgin, Catherine. 2006. "From knowledge to understanding." In *Epistemology Futures*, edited by S. Hetherington, Oxford: Oxford University Press.

Elgin, Catherine. 2009. "Exemplification, idealization, and understanding." In *Fictions in Science: Philosophical Essays on Modeling and Idealization*, edited by M. Saurez, New York, NY: Routledge.
Hursthouse, Rosalind. 1998. "Normative virtue ethics." In *How Should One Live? Essays on the Virtues*, edited by R. Crisp, Oxford: Oxford University Press.
Kripke, Saul. 1980. *Naming and Necessity*. Oxford: Blackwell.
Olberding, Amy. 2012. *Moral Exemplars in the Analects: The Good Person Is That*. New York, NY: Routledge.
Parker, Wendy. 2020. "Local model-data symbiosis in meteorology and climate science." *Philosophy of Science* 87: 807–818.
Potochnik, Angela. 2017. *Idealization and the Aims of Science*. Chicago, IL: University of Chicago Press.
Putnam, Hilary. 1979. "The meaning of 'meaning'." In *Mind, Language, and Reality*. Cambridge: Cambridge University Press.
Rice, Collin. 2019. "Models don't decompose that way: A holistic view of idealized models." *The British Journal for the Philosophy of Science* 70: 179–208.
Zagzebski, Linda. 2001. "Recovering understanding." In *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, edited by M. Steup, Oxford: Oxford University Press.
Zagzebski, Linda. 2010. "Exemplarist virtue theory." *Metaphilosophy* 41: 41–57.
Zagzebski, Linda. 2017. *Exemplarist Moral Theory*. Oxford: Oxford University Press.