Character and Moral Judgment: Designing Right and Wrong

Seth Robertson

Chapter 1: Introduction

1.1 Introduction

Puzzles about rightness and wrongness have long bedeviled moral philosophers. Consider the following cases, based on an example from Sverklik (2011):

> **Mary's Anxiety:** Mary is at a dinner gathering with friends, and is introduced to several new people, including Arjun. Mary has some social anxiety and feels nervous shaking hands with people, so instead of shaking hands with Arjun, she waves. No one else notices or is bothered.

> **Mary's Prejudice:** Mary is at a dinner gathering with friends, and is introduced to several new people, including Arjun. Mary has an extreme prejudice against people from India, and so she refuses to shake Arjun's hand. Instead, she waves. No one else notices or is bothered.

In both cases, we might imagine that Mary's bodily movements, facial expressions, gestures, and tone of voice are identical, and since in neither case anyone else really noticed or was bothered by Mary's actions, the consequences were also identical. While many moral philosophers have claimed that the motivations driving an act should have no bearing on its rightness or wrongness (e.g. Mill 1979, 17–18), many people will have the intuition that Mary did something wrong in the second case but not in the first *because of* her motivations.[1]

And of course, the question of motivation aside, there are even deeper questions and disagreements about rightness and wrongness. For example, there seem to be at least three ways that rightness is treated most generally:

Desirability: An act is right if it is the most morally desirable act for a (any?) person to perform.[2]

Obligation: An act is right if it is the act that a (any?) person is morally obligated to perform.[3]

---

[1] In the opening pages of *The Right and the Good*, W.D. Ross gives what seems to me to be a very poor argument (in an otherwise very good book) that motivations cannot justifiably be included in a theory of rightness. The argument is that since an act is right just if it is morally obligated, and we cannot be obligated to feel motivations (because this would force us to feel motivations we do not, and thus violate the "ought implies can" principle), then motivations cannot help determine what is right. Whatever one thinks about whether motivations should play a role in a theory of rightness, this argument has problems: even if we grant (and we shouldn't) that this is an acceptable use of "ought implies can" and that we cannot become motivated when we aren't, this argument would show that nearly nothing at all could be right, since if I cannot be motivated to do something, then I cannot do it, and if I cannot do it, then it is not the case that I ought to do it.

[2] This is the sense of rightness we saw at the beginning of the chapter, which, for example Arpaly (2003) emphasizes.

[3] This is the sense of rightness emphasized by, for example Ross in *The Right and the Good* (1930).

Reasons: An act is right if it is the act that is, on the balance of reasons, the act with the best reasons to perform.[4]

This additional disagreement aside, there is still further disagreement about the relative importance of rightness for morality. A great many philosophers have been implicitly or explicitly less focused on rightness and wrongness as central moral concepts than other moral concepts: Slote includes Plato, Aristotle, Abelard, Schopenhauer, Kant, Hutchinson, Nietzsche, Spinoza, and "certain Chinese and Indian thinkers" in this group (Slote 2001, 7–10). Even more strongly, Elizabeth Anscombe famously argued that we should (*at least* temporarily) discard the notions of rightness and wrongness altogether (Anscombe 1958).

There are even further a great number of smaller and more specific disagreements about rightness. For example, philosophers have argued about whether rightness should be construed as a scalar or binary notion: that is, whether rightness or wrongness come in degrees, or whether all acts are either simply right or wrong (e.g. Peterson 2013; Brown 2016). To take a final example, another important disagreement centers around whether rightness should be agent-neutral or not: that is, whether what is right to do in a given situation would be right for *anyone* in that situation. Nel Noddings, in *Caring* (2003), explicitly denies this: some of our relationships reshape what we should do in certain circumstances.

In this dissertation, I am particularly interested in questions about the relation between our judgments of right and wrong and the moral character of the person who performed those

---

[4] This is the sense of rightness utilized, for example, in Pettit's *The Robust Demands of the Good* (Pettit 2015).

acts. Let's begin by examining pairs of cases and asking whether the actions performed in each pair were equally right or wrong, and if not, what made it so.

Bully Bryan: Bryan is a bully. He takes genuine joy in insulting his fellow high school students in the most hurtful ways he can imagine. One day, feeling anger, hatred, disgust, and maliciousness, he insults Ivan in the cruelest way he can imagine. Ivan did not know Bryan, and (for other reasons) transferred schools the following day and never saw him again. Bryan continues to be a bully.
Regular Ryan: Ryan is not a bully – in general he is a typical and friendly high school student. One day, he was in a bad mood, and though it was out of character, he insulted his classmate Ivan. In that moment, Ryan felt the exact same emotions to the exact same degrees as Bryan had, and said the exact same insult as Bryan did. Ivan did not know Ryan, and (for other reasons) transferred schools the following day and never saw him again. Ryan did not bully other people after this incident.

Kind Chris: While waiting in line at the cash register in a grocery store, the person currently paying is making a small purchase (a few pieces of fruit) and their credit card is unexpectedly declined. The cashier apologies, saying the system just crashed. The customer checks their wallet and is mortified to realize they don't have any cash, and thus cannot pay for their purchase. Chris notices all this, and without thinking about it, immediately offers to pay for the other customer.
Calculating Caleb: While waiting in line at the cash register in a grocery store, the person currently paying is making a small purchase (a few pieces of fruit) and their credit card is unexpectedly declined. The cashier apologies, saying the system just crashed. The customer checks their wallet and is mortified to realize they don't have any cash, and thus cannot pay for their purchase. Chris notices all this, and while it is unfolding, deliberates about what he should do. He eventually comes to the conclusion that he should offer to pay, and he does, but only after weighing several moral and prudential considerations (such as that helping would make him feel good later, that it would make others think more highly of him, that he could brag about it). Ultimately, his main motivation was to help the distressed customer, but his action was certainly not immediate.

Honest Amanda: Amanda is a high school student, and yesterday was the annual skip day. She skipped class, and had a fun day with her closest friends. Later that week, Amanda's mother pointedly asks whether Amanda skipped school. Amanda knows that she might get in trouble, but she is an honest person. Though it is not an easy situation, and if she lied, she likely could get away with it, she tells the truth.
Truth-telling Miranda: Miranda is a high school student, and yesterday was the annual skip day. She skipped class, and had a fun day with her closest friends. Later that week, Miranda's mother pointedly asks whether Miranda skipped school. If Miranda thought she *could* get away with it, she would not tell the truth. However, in this circumstance, she suspects that she would not be able to get away with the lie, and so she tells the truth.

> **Caring Kate**: Kate has been accepted both to medical school and veterinary school. She believes (correctly) that if she went to medical school, she would be able to help a great many more people than the number of animals should would be able to help as a veterinarian. But, she is deeply committed to animal welfare. She decides to become a veterinarian. She has a long career as a veterinarian and helps a great many animals.
>
> **Capricious Cat**: Cat has been accepted to both medical school and veterinary school. She believes (correctly) that if she went to medical school, she would be able to help a great many more people than the number of animals should would be able to help as a veterinarian. But, on a whim, she decides to become a veterinarian. She has a long career as a veterinarian and helps a great many animals – in fact, she helps an identical number of animals to an identical degree as Kate.

Now, I expect people will have quite different responses to the question of whether each pair of cases contains acts that are equally right or wrong. I suspect some will say that the acts are obviously equally right or wrong, while others will say that the pairs of acts obviously differ in their rightness and wrongness. Some might offer further analysis. Of course, some might try to find ways in which the consequences of the acts will diverge – and to these readers I will stipulate that the details of all cases should be filled out so that the consequences of the paired acts are, as far as possible, the same. Other readers might say that the acts are in fact equally right or wrong, but the people are not – to think otherwise is to confuse character evaluations with act evaluations. Some might say that the acts themselves are far too different to meaningfully compare, and so it is no surprise that they are morally different. Others might say that the acts are intuitively unequal in terms of rightness or wrongness, but this difference has nothing to do with character - perhaps the moral differences are wholly explainable in terms of occurrent motivation or some other features of the situation that do not amount to character (we will see in the next chapter, the distinction between sporadic moral motivations and those moral motivations inextricably bound up with our moral character will be important).[5] Working

---

[5] In Chapter 2, I'll describe in some detail this important distinction, which has been developed largely in discussions about virtue ethics (see Hursthouse 1999; Hurka 2006; Cokelet 2015). The

through these differing intuitions and figuring out how we ought to treat such cases is a key goal of this dissertation. My general goal is to argue that in *some* cases, a person's moral character *should* influence our moral evaluations of the rightness or wrongness of their acts. In particular, I won't be trying to provide a complete account of rightness here, but I will argue that adequate theories of rightness should simultaneously meet the following two conditions:

The Consequences Condition: for at least some (but not all) cases, the rightness or wrongness of an act should be determined conclusively by the act's consequences (i.e. how well the act promotes well-being: though there may be competing moral considerations, no competing moral considerations could outweigh the consequences in determining the rightness or wrongness of the act.

The Character Condition: in some (but not all) cases, the moral character of a person should influence the determination of the rightness or wrongness of that person's acts.

Before I say more about how I'll be interpreting these conditions or how the argument for them will go, it will be helpful to briefly consider what it would mean for our ethical theorizing if we accepted them.

1.2 Implications of the Consequences and Character Conditions

The main project in this dissertation is developing and defending a hybrid account of rightness and wrongness, according to which the rightness of some acts should be *conclusively* determined by their consequences alone (the Consequences Condition) while the rightness of

---

main idea is that certain motivations stem from engrained character dispositions while others do not. Importantly, the claim will not be that only character-influenced motivations have moral value, but instead that the ways in which certain motivations are bound up in our moral character - our entrenched and habitual psychological patterns of interpreting, understanding, acting in the world – will inform and help us determine the value of certain acts.

some (other) acts should be partially determined by the character of the agent who performed them (the Character Condition). Before I outline the argument that I'll be providing in support of these two conditions, it is worth briefly taking some time to explore what it would mean for normative ethics if both these conditions held, because it would mean a great deal: in short, that nearly every major ethical theory and many of their most important variations would require at least some revision.

Let's begin with consequentialism. While consequentialist theories of rightness might disagree about exactly which consequences are good, and how that goodness should be added up, they will agree that it is the consequences alone, counted up somehow that should determine whether an act is right or wrong. This commitment is, of course, perfectly compatible with the Consequences Condition: if the Consequences Condition holds, it does not falsify consequentialism. However, consequentialism is not at all compatible with the Character Condition (as I'll spell it out), since some of the particular ways in which character should influence moral judgment will not be reducible themselves to purely consequentialist considerations. Now, I should emphasize that there *are* ways of spelling out the Character Condition that *are* potentially compatible with some versions of consequentialism. Consider Philip Pettit's version of consequentialism in *The Robust Demands of the Good* (Pettit 2015), according to which we cannot fully capture what we value or cherish about certain actions without considering some of their counterfactual consequences. Pettit argues that for some actions, such as actions of love or honesty, part of what we value or cherish about them is that the person who performed them would also perform them in certain relevant counterfactual circumstances: that the person has a character disposition to perform them. If our theory of rightness is meant to capture the totality of what we value then, it must be (albeit, slightly

indirectly) informed and influenced, in some cases, by the character of the person who performed the action. However, this isn't a particularly strong version of the Character Condition: the goodness of the counterfactual consequences of an action are doing most of the work. The arguments I'll develop in favor of the Character Condition will go beyond this: the ways in which (I'll argue that) character should influence our moral judgment about certain actions will not all be neatly reducible (or reducible at all) to actual or counterfactual consequences. And so, insofar as we find the case for the Character Condition plausible, we should reject a fully consequentialist theory of rightness.

While deontological views come in a great many varieties, we are most immediately interested here in those that stand in contrast to consequentialist views by denying that the rightness or wrongness of an action should be determined only by way of its consequences but instead by its accordance with certain moral rules, principles, or duties. But, there is tension here with both the Consequences and the Character Conditions. First, and most obviously, the Consequences Condition claims that the rightness or wrongness of some acts should be determined *conclusively* by their consequences alone, and is thus absolutely inconsistent with fully anti-consequentialist deontological theories. But further, on the interpretation of the Character Condition I'll defend, the ways in which character should influence rightness are not straightforwardly representable by moral rules or moral principles: they are not codifiable (and, even if they were, their rightness in these cases does not *derive* from their adherence with moral principles). I should however note here, before moving on, a certain affinity with the view of rightness I will develop and the view that Ross spells out in *The Right and the Good* (1930). Both will treat determinations of rightness as a process of weighing various moral considerations

against each other, many of which are non-consequentialist: the views will differ in the types of considerations they include.

Virtue ethics, which recognizes the importance of moral character most centrally of the major normative theories, might seem at first glance to be most friendly to the overall project of this dissertation, but in fact there is tension coming from both the Consequences and the Character Conditions. Typically, virtue ethical accounts of right action treat an act as right just in case it is the virtuous act, or the act that a virtuous person would characteristically do in that situation (e.g. Hursthouse 1999). It seems likely to me that the acts meant to be covered by Consequences Condition would be coextensive with the virtuous acts in those situations: it is hard to see how, in situations in which the differences in the consequences between options is incredibly vast, that a virtuous person would choose the option that, for example, causes vastly more suffering than happiness. But, co-extension is not identity. Vitally, according to the view I'll develop, in these types of cases, the right options are right (and the wrong options are wrong) *conclusively* because of purely consequentialist considerations: a virtuous person should characteristically perform them *because* of these consequentialist considerations and not the other way around. The tension between virtue ethical theories of rightness and the Character Condition is more slight, and it depends on the degree of implicit agent-neutrality within the virtue ethical theory: that is, the degree to which an act is virtuous regardless of the agent-relative characteristics of the person performing the act. Different virtue ethical theories can vary here, and those that leave less space for differences in behavior, motivation, and composition of character between virtuous agents will clash more strongly with the Character Condition.

1.3 The Argument

The argument for the Consequences and Character Conditions must overcome two initial hurdles: (1) its conclusion is radically revisionary relative to certain views in the literature: to some, it is so counterintuitive that it is simply and obviously false, and (2) in giving such a high level argument, I must ignore a great many important disagreements, omit a great many important discussions, overlook a great many important debates, and elide a great many important distinctions about rightness and wrongness. So, I'll need to find some way to overcome these first two hurdles before even beginning to make the positive case for the Consequences and Character Conditions.

What we will need first and foremost is a good reason for dislodging our certainty about our most strongly-held intuitions about rightness. Doing so will clear the space for and lift a bit of the burden off of the argument for the Consequences and Character Conditions. But, moral reasoning and ethical theorizing run on our moral intuitions, so we also need the argument to not push us into *complete* distrust of them. The strategy I develop for doing all this will unfold in several steps. First, I deploy a modified version of evolutionary debunking arguments (against moral realism) to dislodge our intuitions about rightness that are based on the belief that there is an objective fact of the matter about rightness (which we know) or the belief that we know certain features of rightness because they are self-evident.

But, in deploying this argument, I appear to dig the hole deeper for myself. The Consequences and Character Conditions are normative claims and need justification *somehow*. But, if my version of the debunking argument succeeds, then justification from moral realism in general, or moral intuitionism in particular, is not available. And so, I put myself on the hook for an alternative (and, hopefully more plausible) account of moral justification – one that is not susceptible to the debunking argument I give, but still helps achieve the goal of dislodging our

strong moral-theoretical intuitions without pushing us to discard them altogether. Fortunately for us, however, the metaethical view I will develop actually provides *further* reason to dislodge our strongest moral-theoretical intuitions without needing to throw away all of them. Also, fortunately, in developing this metaethical view, I won't need to start from scratch: several philosophers have already developed anti-realist and constructivist metaethical views and the view I develop, perspectival naturalism, is a modification of such views. It is naturalist in the sense that it is committed to our best scientific / naturalist understanding of the world *and* it does not appeal to non-natural moral truths. It is perspectival in that it is interested in far more of our moral minds than our moral beliefs and in far more of our moral views than the propositional claims they make. Perspectival naturalism modifies one popular type of what we might call functionalist anti-realist or constructivist view (according to which moral-theoretical justification derives from how well the moral theory in question facilities the fulfilment of some narrow function of morality) (Kitcher 2011; see also Mackie 1991's critical discussion of morality as a device for fixing limited sympathies) by treating morality (descriptively *and* normatively) as a *design project* rather than a technology. Thus, according to perspectival naturalism, the best moral theory for creatures like us is the moral theory that is best designed for creatures like us: that best manages the many (sometimes competing) interests of our goals, values, interests, wants, needs, circumstances, psychologies, and tendencies. Thus, moral theories are *constructed* by us, for us. Importantly, perspectival naturalists treat this as a thoroughly unradical claim: this project is *exactly* what humans at all times and in all places have been doing when we engage in ethical discussion and deliberation. And, we are then in the position to make the case for dislodging our most strongly-held moral intuitions without totally discarding them. We should distrust extremely strongly-held moral-theoretical intuitions because (a) they are *not* justified via

knowledge that they are objectively true, (b) they are *not* justified via knowledge that they are self-evident, (c) on the most plausible metaethical view, they descriptively *are* the results of contingent social construction processes, (d) on the most plausible ethical view, they are always subject to change and improvement if a *better-designed* moral theory is available, and (e) on the most plausible metaethical view (or, more accurately, its independently plausible psychological commitments about perspective and our moral psychologies), there can be no *perfect* ethical theory (that is, an ethical theory that perfectly captures all of our moral intuitions in a perspectival framework), and so we must always remain open to the possibility of changes to our held moral theories that we would accept. And so, the burden is very slightly lifted – at least enough to justify the project of arguing for the Consequences and Character Conditions despite the fact that doing so necessitates clashing with or ignoring a great many discussions, debates, and disagreements about rightness and wrongness. But further, this is not done at the cost of forcing us to wholly reject all of our moral and moral-theoretical intuitions: since perspectival naturalism is essentially constructivist, moral justification *must* come ultimately from our moral and moral-theoretical intuitions (of course, via a process of refinement: for perspectival naturalism, a design process). And further still, we now have a (in some sense) novel account of moral theoretical justification that we can put to use in making the case for the Consequences and Character Conditions: we should adopt them if moral theories that do are, *ceteris paribus* better designed for us than those that do not.

Thus, the arguments for the Consequences and Character Conditions unfold similarly: we look at the logical alternatives to each of them and show why the Consequences and Character Conditions should be preferred. We compare the Consequences Condition to the Always-Consequences Condition (which states that the consequences should *always* conclusively

determine an actions rightness or wrongness) and the Never-Consequences Condition (which states that the consequences should *never* conclusively determine the rightness or wrongness of an action). We then compare the Character Condition to the Only-Character Condition (which states that character is the *only* consideration that should influence our determinations of rightness or wrongness) and the Not-Character Condition (which states that character should not at all influence our determinations of rightness or wrongness) Part of these arguments depend on each other: if the Consequences Condition is independently plausible, then we should reject the Only-Character Condition (because they are inconsistent), and if the Character Condition is independently plausible, then we should reject the Always-Consequences Condition (again, because they are inconsistent). And so, much of the work will be showing that the Consequences Condition is preferable to the Never-Consequences Condition (that is, showing that at least sometimes the consequences *should* conclusively determine the rightness of an act) and that the Character Condition is preferable to the No-Character Condition (that is, that showing that at least sometimes character *should* influence our determination of the rightness of an act). In both cases, the arguments will show that the Consequences and Character Conditions better manage our natural evaluative tendencies and our values than these alternatives.

1.4 The Plan

It will be helpful here to briefly outline the structure of the dissertation. Most generally, in Chapter 2, I develop and describe in detail Consequences and Character Conditions as well as the metaethical view that underlies the project of this dissertation, perspectival naturalism. In Chapter 3, I make the case for perspectival naturalism and develop the ground-clearing or burden-shifting argument that our strongest moral-theoretical intuitions should be dislodged but not abandoned. In Chapter 4, I make the case for the Consequences and Character Conditions.

Chapter 2

Up to this point, I have not spelled out in much detail either the Consequences and Character Conditions or the metaethical view I'll be using, perspectival naturalism. I take up the task of doing so in Chapter 2. I discuss in more detail how I am interpreting the key concepts of *consequences* and *character* as well as the types of cases that I take each condition to cover. Then, I describe and develop key features of the metaethical view that will underpin the argument in the rest of the dissertation.

Chapter 3

The goal of Chapter 3 is to provide a ground-clearing or burden-shifting argument that will justify the relatively revisionary and high-level nature of the argument for the Consequences and Character Condition. That argument will unfold in several steps. It begins with the development of what I call the "radical debunking argument," a modification of more familiar evolutionary debunking arguments against moral realism that aims to show that we should not accept (strong versions of) moral realism or straightforward versions of moral intuitionism (according to which we have moral knowledge of certain self-evident truths). I then make the case for perspectival naturalism, the metaethical view that I believe is most plausible, given the radical debunking argument. Finally, I lean on perspectival naturalism to provide several reasons for dislodging our most strongly-held moral-theoretical intuitions without forcing us to abandon them altogether, which will provide us with the ground-clearing and burden-shifting argument we needed.

Chapter 4

In Chapter 4, I make the case for the Consequences and Character Conditions, based on the metaethical view and its account of moral-theoretical justification developed in the previous chapter. Beginning with the Consequences Condition, I contrast it with its logical rivals, which claim that the consequences alone should *always* conclusively determine the rightness of an action or that they *never* should. I provide several reasons to prefer the Consequences Condition over these alternatives. Then, I turn to the case for the Character Condition, which is the bulk of the focus of the chapter. I lay out several arguments in favor of the Character Condition. I then finish the chapter by considering and responding to several objections to the case for the Character Condition.

Chapter 5

In Chapter 5, we will take a step back and look at what I have argued so far. I will provide a summary of the overall argument of the dissertation. Then, I will expand the brief discussion we had earlier in this chapter about the implications of views I've developed, and we will look in much more detail at how perspectival naturalism and the Consequences and the Character Conditions should inform our opinions and evaluations of several extent views in metaethics and normative ethics. Finally, I will suggest several potential directions for future research based on the groundwork that I have laid in this dissertation.

References

Anscombe, G.E.M. 1958. "Modern Moral Philosophy." *Philosophy* 33 (124): 1–16.
Arpaly, Nomy. 2003. *Unprincipled Virtue*. Oxford: Oxford University Press.

Brown, Campbell. 2016. "The Rightest Theory of Degreees of Rightness." *Ethical Theory and Moral Practice* 19 (1): 21–29.

Cokelet, Bradford. 2015. "Dispositions, Character, and the Value of Acts." In *Character: New Directions from Philosophy, Psychology, and Theology*, edited by Christian B Miller, R. Michael Furr, Angela Knobel, and William Fleeson, 233–50. Oxford: Oxford University Press.

Hurka, Thomas. 2006. "Virtuous Act, Virtuous Disposition." *Analysis* 66: 69–76.

Hursthouse, Rosalind. 1999. *On Virtue Ethics*. Oxford: Oxford University Press.

Kitcher, Phillip. 2011. *The Ethical Project*. Cambridge, MA: Harvard University Press.

Mackie, John L. 1991. *Ethics: Inventing Right and Wrong*. Penguin Books.

Mill, John Stuart. 1979. *Utilitarianism*. Edited by George Sher. Indianapolis: Hackett.

Noddings, Nel. 2003. *Caring: A Feminine Approach to Ethics and Moral Education*. 2nd ed. University of California Press.

Peterson, Martin. 2013. *The Dimensions of Consequentialism*. Cambridge: Cambridge University Press.

Pettit, Philip. 2015. *The Robust Demands of the Good: Ethics with Attachment, Virtue, and Respect*. New York: Oxford University Press.

Ross, David. 1930. *The Right and The Good*. Oxford: Clarendon Press.

Slote, Michael. 2001. *Morals from Motives*. Oxford: Oxford University Press.

Sverklik, Steven. 2011. *Motive and Rightness*. Oxford: Oxford University Press.