Georgia State University

ScholarWorks @ Georgia State University

Philosophy Theses

Department of Philosophy

8-10-2021

What Is Wrong With Self-effacing Ethical Theories?

Austin Keefe

Follow this and additional works at: https://scholarworks.gsu.edu/philosophy_theses

Recommended Citation

Keefe, Austin, "What Is Wrong With Self-effacing Ethical Theories?." Thesis, Georgia State University, 2021.
doi: https://doi.org/10.57709/23185824

This Thesis is brought to you for free and open access by the Department of Philosophy at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Philosophy Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

WHAT IS WRONG WITH SELF-EFFACING ETHICAL THEORIES?

by

AUSTIN KEEFE

Under the Direction of Andrew I. Cohen, PhD

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2021

ABSTRACT

Self-effacing ethical theories are those that recommend their own erasure. Such theories are a controversial topic in contemporary moral philosophy. In this thesis, I shed light on what is and is not wrong with this type of theory. I examine two kinds of self-effacing ethical theories, a radical version of sophisticated consequentialism and developmental virtue ethics. I defend them against three common objections to self-effacing theories. I raise and develop two novel objections to self-effacing theories: a self-erasure objection and an incompleteness objection. I conclude by arguing that these two objections pose fewer challenges for developmental virtue ethics than radical sophisticated consequentialism.

INDEX WORDS: Self-effacing, Virtue ethics, Developmental virtue ethics, Consequentialism, Sophisticated consequentialism, Decision procedure

Copyright by
AUSTIN JEROME KEEFE
2021

WHAT IS WRONG WITH SELF-EFFACING ETHICAL THEORIES?

by

AUSTIN KEEFE

Committee Chair: Andrew I. Cohen

Committee: Tim O'Keefe

S.M. Love

Electronic Version Approved:

Office of Graduate Services

College of Arts and Sciences

Georgia State University

August 2021

DEDICATION

For my parents and sibling.

ACKNOWLEDGEMENTS

For whatever is lacking in this thesis, I have no one to thank but myself. Yet, for all that is good in it, I have three people, in particular, to thank. I am profoundly grateful, first and foremost, to Dr. Andrew I. Cohen, my committee chair, for both the guidance he gave me throughout this process and the insightful comments, questions, and suggestions he provided on each of the many drafts of this thesis. I am also deeply grateful to Drs. Tim O'Keefe and S.M. Love, my committee members, for the incisive comments they provided on the latest drafts of this thesis and for the probing and interesting questions they raised during my thesis defense.

Aside from my committee, no one provided me more support and advice during this process than Carly Lewis, for whose wisdom and love I am perpetually grateful. I am also grateful to Dan Song, and Adin Bohmiller, whose comments and questions on the penultimate draft of this thesis helped me to improve both the clarity and strength of my argument. My roommates and fellow philosophers, Richard Link and John Garrett, are deserving of thanks as well: their friendship has been among the most valuable parts of my time here at Georgia State University, and without their conversation and support, writing this thesis would have been significantly more difficult for me than it turned out to be.

I am indebted to those professors who mentored me during my time as an undergraduate at Holy Cross; in particular, Predrag Cicovacki, Theresa McBride, Denise Schaeffer, and Cynthia Hooper. Were it not for their mentorship and wisdom, I would likely not have been inspired to apply to, let alone have the honor of being accepted into, a masters program as wonderful as that at Georgia State University; and in consequence, I would likely not have been able to write this thesis.

Finally, I want to thank my parents, sibling, grandfather, and the rest of my family for all of their love and support over the course of my life and, especially, during the past two years. Without you, none of this would have been possible.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS V

PREFACE IX

1 INTRODUCTION: THE MOTIVATION PROBLEM AND SELF-EFFACING ETHICAL THEORIES 1

2 WHAT IS NOT WRONG WITH SELF-EFFACING ETHICAL THEORIES 3

***2.1* The Decision Procedure Framework (DPF) 3**

***2.2* Three Objections to Self-Effacing Ethical Theories 4**

2.2.1 Keller's Objection 4

2.2.2 Pettigrove's Objection 5

2.2.3 Martinez's Objection 6

***2.3* Two Kinds of Self-Effacing Ethical Theory 7**

2.3.1 Sophisticated Consequentialism 7

2.3.2 Developmental Virtue Ethics (DVE) 12

***2.4* Defense of R-SC and DVE 14**

2.5 Conclusion 20

3 WHAT IS WRONG WITH SELF-EFFACING ETHICAL THEORIES 21

3.1 The Self-Erasure Objection 21

3.2 The Incompleteness Objection 29

3.3 Reconsidering the Incompleteness Objection: A Case for Development 31

4 CONCLUSION..... 33

REFERENCES..... 35

PREFACE

This thesis is a study about ethical theories that recommend their own erasure. Such theories are commonly known as *self-effacing* ethical theories. In this thesis, I offer a qualified defense of some prominent kinds of self-effacing ethical theories against the charge that they are inadequate as ethical theories.

This defense is a timely one, for among contemporary ethicists, self-effacing ethical theories are the subject of significant controversy. On the one hand, various scholars think that there is something gravely wrong with self-effacing ethical theories. In particular, they think that self-effacing ethical theories fail to perform some of the basic functions of an ethical theory and, thus, are not good ethical theories. On the other hand, certain scholars view self-effacing theories in a more positive light. For consequentialists such as Peter Railton, self-effacing theories represent a way to overcome some powerful objections to consequentialism.[1] And for virtue ethicists such as Julia Annas, self-effacing versions of virtue ethics are effective at helping agents develop the ability to "act, think, and feel virtuously."[2] I consider how such theories fare against some new challenges to their self-effacement.

This thesis contributes to the current debate about what is and is not wrong with self-effacing ethical theories. In Section One, I examine and defuse three recent objections to self-effacing ethical theories. In Section Two, I raise novel objections to two kinds of self-effacing ethical theories: sophisticated consequentialism and the developmental version of virtue ethics propounded by Julia Annas[3] and Justin C. Clark.[4] I conclude this thesis by arguing that charges

---

[1] Peter Railton, "Alienation, Consequentialism, and the Demands of Morality," *Philosophy & Public Affairs* 13, no. 2 (Spring, 1984): 134-171.
[2] Julia Annas, "Virtue Ethics and the Charge of Egoism," in *Morality and Self-Interest*, ed. Paul Bloomfield (Oxford: Oxford University Press, 2008), 213.
[3] Julia Annas, *Intelligent Virtue* (Oxford: Oxford University Press, 2011).
[4] Justin C. Clark, "Eudaimonistic Virtue Ethics and Self-Effacement," *Journal of Value*

of self-effacement pose fewer challenges for developmental virtue ethics than for sophisticated consequentialism.

Inquiry 50, no. 3 (September 1, 2016): 507–24.

1 INTRODUCTION: THE MOTIVATION PROBLEM AND SELF-EFFACING ETHICAL THEORIES

It seems reasonable to maintain that part of the goal of moral philosophy is not just to help us identify the right thing to do but also to help us do it—or at the very least, to not directly undermine our capacity to do it. Some ethical theories fail at this task.[5] Such theories suffer from what I call *the motivation problem*. Stated generally, the motivation problem is this: it seems that if we are motivated to act by a desire to realize the good(s) that these ethical theories tell us to realize, then we will often *fail* to realize said good(s).[6] In particular, we risk failing to realize important moral values, especially when doing so is conditional upon our having the right sort of motivation when acting.[7] This is a particular concern in personal relationships such as friendships.

One way for an ethical theory to solve the motivation problem is for it to become a *self-effacing* ethical theory. Considered broadly, self-effacing ethical theories are those that, at some level, recommend their own erasure. More specifically, self-effacing ethical theories tend to give us the following advice: to realize the good(s) the theory tells us to realize, we need not always be motivated to act by a desire to realize said good(s). It might seem straightforward how

---

[5] Bernard Williams has identified various forms of consequentialism as theories of this sort. Bernard Williams, "A critique of utilitarianism," in *Utilitarianism: For and Against* by J.C.C. Smart and Bernard Williams (Cambridge: Cambridge University Press, 1973), 77-150. Further, Michael Stocker goes so far as to accuse all modern ethical theories of failing at this task. Michael Stocker, "The Schizophrenia of Modern Ethical Theories," *Journal of Philosophy* 73, no. 14 (August 1976): 453-466.

[6] Notice that the problem with ethical theories that suffer from the motivation problem is *not* that they fail to help us identify the right thing to do. Rather the issue is, to quote Michael Stocker, that "when we try to act on [these] theories, try to embody their reasons in our motives—as opposed to simply seeing whether our or other's lives would be approved of by the theories—then in a quite mad way, things start going wrong." Stocker, "The Schizophrenia of Modern Ethical Theories," 466.

[7] See Neera Badhwar, "Why It Is Wrong to be Always Guided by the Best: Consequentialism and Friendship," *Ethics* 101, no. 3 (April, 1991): 483-504. See also Scott Woodcock, "Moral Schizophrenia and the Paradox of Friendship." *Utilitas: A Journal*,*of Utilitarian Studies* 22, no. 1 (March 1, 2010): 1–25.

becoming self-effacing can help a theory solve the motivation problem. Self-effacing ethical theories solve the motivation problem by informing us that, to some degree, we need not be motivated to act by a desire to realize the good(s) the theory tells us to realize. They tell us that an action that best contributes to human happiness, or that will be most conducive to our either living a good life or becoming a virtuous person, need not always be an action motivated by a desire to realize these goods.

While the ability of self-effacing ethical theories to solve the motivation problem undoubtably counts as a mark in their favor, moral philosophers are wary of them and worry that there may be something wrong with them. In this thesis, I present a clearer picture of what is and is not wrong with two prominent self-effacing ethical theories: a radical version of sophisticated consequentialism (R-SC) and developmental virtue ethics (DVE). In Section 2, I dispel some common worries about self-effacing ethical theories. I focus particularly on whether such theories are inadequate because they fail to provide us with motives for action or they fail to provide determinate guidance on what we morally ought to do. In Section 3, I raise some novel problems with R-SC and DVE. I then conclude by arguing that we have good reason to believe that charges of self-effacement pose fewer problems for DVE than for R-SC.

2 WHAT IS NOT WRONG WITH SELF-EFFACING ETHICAL THEORIES

In this section, I consider three objections to self-effacing ethical theories. Put broadly, these objections hold that when a theory is self-effacing, it is thereby inadequate as an ethical theory. I argue such objections have little or no grip against two prominent self-effacing ethical theories: a radical version of Peter Railton's sophisticated consequentialism (R-SC) and Julia Annas's developmental eudaimonistic virtue ethics (DVE). Before considering these objections, I first consider the basic aims of ethical theorizing.

2.1 The Decision Procedure Framework (DPF)

Ethicists often suppose ethical theories have two basic aims. In *Moral Theory: An Introduction*, Mark Timmons writes,

> *Practical Aim*. The main practical aim of a moral theory is to provide a decision procedure whose use by suitably informed agents will reliably lead them to correct moral verdicts about matters of moral concern in contexts of moral deliberation and choice.[8]
>
> *Theoretical Aim*. The main theoretical aim of a moral theory is to discover those underlying features of actions, persons, and other items of moral evaluation that make them right or wrong, good or bad.[9]

Taken together, the practical and theoretical aims of ethical theorizing establish a basic framework for what an ethical theory should look like.[10] I call this "the decision procedure framework" (DPF). I do not offer a detailed defense here of the idea that the decision procedure framework is one to which a satisfactory ethical theory must conform. Such an argument would take another paper to make. Instead, I assume for the sake of argument that conforming to the

[8] Mark Timmons, *Moral Theory: An Introduction* (Lanham: Rowman & Littlefield Publishers, 2012), 3.
[9] Timmons, *Moral Theory: An Introduction*, 4.
[10] This is not to say that an ethical theory that meets the requirements of this framework necessarily counts as a good ethical theory; rather these requirements seem to be necessary conditions, though not sufficient conditions, for an ethical theory to be a good theory. At the very least, if an ethical theory does not achieve the practical aim and the theoretical aim, then there is something gravely wrong with that theory; it is a bad ethical theory.

decision procedure framework is a necessary condition for a satisfactory ethical theory and reserve another time for a defense of this idea.

2.2 Three Objections to Self-Effacing Ethical Theories

I consider three objections to self-effacing ethical theories. Each supposes that, if an ethical theory does not adequately fulfill the practical and theoretical aims of the DPF, then it is not a good ethical theory. Put simply, these objections argue that self-effacing ethical theories, in virtue of being self-effacing, do not adequately fulfill the practical aim of the DPF. Below, I will sketch and assess each of these objections in turn.

2.2.1 Keller's Objection

The first objection to self-effacing ethical theories that I consider comes from Simon Keller. Keller writes that self-effacing ethical theories "do not tell us what should motivate us, so they do not tell us how to live, so they do not perform a function that an ethical theory should perform."[11] Keller here supposes that a proper ethical theory will perform the function of telling us how we should live. There are different ways to interpret the phrase "telling us how we should live." I take it to mean that a proper ethical theory will fulfill the practical aim of the DPF. As Timmons writes, a proper ethical theory will "provide a decision procedure whose use by suitably informed agents will reliably lead them to correct moral verdicts about matters of moral concern in contexts of moral deliberation and choice."[12] The objection also assumes that an

[11] Simon Keller, "Virtue Ethics is Self-Effacing," *Australasian Journal of Philosophy* 85, no. 2 (June 2007): 222. Keller's objection is akin to that raised by Michael Stocker in "How Emotions Reveal Value and Help Cure the Schizophrenia of Modern Ethical Theories," in *How Should One Live: Essays on the Virtues*, ed. R. Crisp (Oxford: Oxford University Press, 1998), 173.
[12] Timmons, *Moral Theory*, 3.

ethical theory's ability to tell us how we should live is conditional upon its ability to tell us what should motivate us.[13]

If we take these two assumptions to be true, then it is easy to see why Keller maintains that self-effacing ethical theories are not good ethical theories. Recall that self-effacing ethical theories tell agents the following: To some degree, agents need *not* be motivated to act by a desire to realize whatever good(s) the theory tells them they ought to realize. Since self-effacing theories tell agents that there is a sense in which agents need not desire to realize the good(s) the theory tells them to realize, it seems reasonable of Keller to worry that such theories do not tell us what should motivate us. And if self-effacing ethical theories do not tell us what should motivate us, then (on Keller's second assumption above) they do not tell us how to live; and if they don't tell us how to live, then they do not fulfill the practical aim of the DPF. Self-effacing theories, on Keller's objection, are then poor ethical theories because they are *incomplete*.

2.2.2 Pettigrove's Objection

The second objection to self-effacing ethical theories that I consider comes from Glen Pettigrove. He writes,

> If a normative theory is meant to help an agent answer the question "What should I do?" then it is problematic if the agent must put the theory's chief tool for answering the question out of her mind when she acts. At the very least, it introduces a time-lag between when the agent may consult the theory for guidance and when she may act in accordance with that guidance.[14]

---

[13] I think that this notion is a reasonable one. To see why it is reasonable, consider an ethical theory like act consequentialism. Act consequentialism is commonly thought to provide us with a decision procedure that says: an action is right if and only if it is the action that, compared to the other actions available to us in the circumstance, will contribute most to the general happiness of humankind. While this decision procedure does not explicitly tell us what should motivate us, it does seem to imply that there is a sense in which the person using it should desire to maximize the happiness of humankind. If an informed agent did not feel a fairly passionate desire to maximize the general happiness of humankind, then why would he adopt a decision procedure that posits maximizing the general happiness of humankind as an overriding concern? For a similar view, see Williams, "A critique of utilitarianism," 128.

[14] Glen Pettigrove, "Is Virtue Ethics Self-Effacing?" *Journal of Ethics* 15 (2011): 192.

In contrast to Keller, who complained self-effacing theories are incomplete by not providing a decision procedure, Pettigrove's worry is that such theories provide a poor decision procedure, because it is insufficiently practicable. Pettigrove's objection rests on two main assumptions. First, if an ethical theory is self-effacing, then it will advise an agent to put its decision procedure "out of her mind when she acts."[15] Second, if an ethical theory is to achieve the practical aim of the DPF, it must provide a decision procedure that is suffiently practicable; i.e., one which agents can have in mind when they are acting.

2.2.3 Martinez's Objection

The third objection to self-effacing ethical theories that I will consider comes from Joel Martinez. Importantly, Martinez does not aim his objection at all self-effacing theories. Rather it applies only to what he calls "*immodestly self-effacing* ethical theories.[16] According to Martinez, a theory is immodestly self-effacing if it "recommends its own rejection, in the sense that it recommends that we must *always* act from considerations other than those that make the action right."[17] Martinez argues such self-effacing ethical theories do not guide action in the right way, "because as soon as they advise, they are discharged."[18] In contrast, an ethical theory that guides action in the right way will "provide ongoing advice about particular motives and viewpoints through reflective revision informed in part by the theory and in part by the normative considerations that are salient in particular circumstances."[19]

I take Martinez's objection to be one about the inability of immodestly self-effacing ethical theories to fulfill the practical aim of the DPF. Martinez thinks the decision procedures of

---

[15] Pettigrove, "Is Virtue-Ethics Self-Effacing?" 192.
[16] Joel A. Martinez, "Is Virtue Ethics Self-Effacing?" *Australasian Journal of Philosophy* 89, no. 2 (June 2011): 281.
[17] Martinez, "Is Virtue Ethics Self-Effacing?" 279-280.
[18] Martinez, 281.
[19] Martinez, 281.

immodestly self-effacing theories are poor decision procedures, because they do not provide agents with "ongoing advice."[20] For Martinez, this failure to provide ongoing advice is significant, since it "precludes regarding morality as an ongoing project that takes place over the course of a life."[21] If morality really is an ongoing project of this sort, then it seems like an ethical theory's decision procedure should reflect this fact rather than contradict it. Since the decision procedures of immodestly self-effacing theories' contradict this fact, Martinez's holds they are bad ethical theories.

2.3 Two Kinds of Self-Effacing Ethical Theory

I argue that the objections of Keller, Pettigrove, and Martinez to self-effacing ethical theories are unsound, at least when applied to two prominent kinds of self-effacing ethical theories: a radical version of Railton's sophisticated consequentialism (R-SC) and Annas's developmental virtue ethics (DVE). What are these two theories, and how do they avoid the above three objections to self-effacing theories?

2.3.1 Sophisticated Consequentialism

Peter Railton uses the label "sophisticated consequentialism" for consequentialist theories that do "not set special stock in any particular form of decision making."[22] Railton distinguishes sophisticated consequentialism from two other types of consequentialism: subjective consequentialism and objective consequentialism.

Unlike sophisticated consequentialist theories, subjective consequentialist theories advise "following a particular mode of deliberation in action."[23] For example, a subjective kind of act consequentialism might prescribe that, "whenever one faces a choice of actions, one should

[20] Martinez, 281.
[21] Martinez, 281.
[22] Railton, "Alienation, Consequentialism, and the Demands of Morality," 153.
[23] Railton, "Alienation," 152.

attempt to determine which act of those available would most promote the good, and should then try to act accordingly."[24] Similarly, a subjective kind of rule consequentialism might prescribe that "in deliberation we should always attempt to determine which act, of those available, conforms to that set of rules general acceptance of which would most promote the good; we then should attempt to perform this act."[25]

Objective consequentialism is "the view that the criterion of rightness of an act or course of action is whether it in fact would most promote the good of those acts available to the agent."[26] Objective consequentialism differs from subjective consequentialism and sophisticated consequentialism, because it is indifferent as to particular forms of decision making. To quote Railton, subjective and sophisticated consequentialism are distinguished by requiring (or not requiring) a "particular mode of deliberation in action,"[27] whereas "objective consequentialism […] concerns the outcomes actually brought about, and thus deals with the question of deliberation only in terms of the tendencies of certain forms of decision making to promote appropriate outcomes."[28] In this way, both subjective and sophisticated consequentialists are, in an important sense, committed to living objectively consequentialist lives, because both are motivated to live the kind of life that will "bring about the best outcomes."[29] Thus, the critical difference between subjective and sophisticated consequentialists is that they disagree about what mode of deliberation, or complex of modes of deliberation, will be most conducive to their living such a life. The subjective consequentialist holds that using a particular mode of

[24] Railton, 152.
[25] Railton, 156.
[26] Railton, 152.
[27] Railton, 152.
[28] Railton, 152.
[29] Railton, 152. Railton defines an objectively consequentialist act (or life) in the following terms: "Let us reserve the expression *objectively consequentialist act* (*or life*) for those acts (or that life) of those available to the agent that would bring about the best outcomes." Railton, 152.

deliberation will be most conducive to his leading an objectively consequentialist life. In contrast, the sophisticated consequentialist maintains that the objectively consequentialist life will be best achieved by one who takes a more adaptable approach to deliberation and does not "set special stock in any particular form of decision making."[30]

A sophisticated consequentialist theory may be more or less "sophisticated." This idea has the following upshot: it entails that to describe a consequentialist theory as a "sophisticated consequentialist theory" is to say that it falls somewhere on a *scale of sophistication*. At one end of this scale, we find forms of sophisticated consequentialism that, save for some minor differences, resemble subjective consequentialist theories. Considered generally, these kinds of sophisticated consequentialism tell us that there are a few circumstances in which we should not "undertake a distinctively consequentialist deliberation when faced with a choice."[31] Such circumstances tend to be those in which deliberation—that is, any form of reflective deliberation at all—would be either disastrous or inefficient. Railton points to cases such as emergencies that "require action so swift as to preclude consequentialist deliberation,"[32] and to situations in which the choice one faces is too "insignificant to warrant consequentialist deliberation […] or too predictable in outcome."[33] Thus, a less sophisticated form of consequentialism might prescribe

---

[30] Railton, 153.
[31] Railton, 153.
[32] Railton, 153.
[33] Railton, 153. I stipulate that any consequentialist theory that fails to allow this minor degree of sophistication is *prima facie* self-defeating and untenable. Indeed, as a Railton describes it, subjective consequentialism seems to be a theory that no reasonable consequentialist would endorse. For it appears to prescribe that one deliberate, using consequentialist reasoning, in any and every situation, including situations wherein successfully doing the right thing requires reflexive, non-deliberative action. While I cannot defend my position at length here, I want to note that it seems to parallel a position taken by J.C.C. Smart. In "An outline of a system of utilitarian ethics," Smart describes what I would call a sophisticated consequentialist, writing: "He knows that a man about to save a drowning man has no time to consider various possibilities, such as that the drowning person is a dangerous criminal who will cause death and destruction, or that he is suffering from a painful and incapacitating disease from which death would be a merciful release, or that various timid people, watching from the bank, will suffer a heart attack if they see anyone else in the water. No, he knows that it is almost always right to save a drowning man, and in he goes. Again, he knows that we would go mad if we went in detail into the probable consequences of keeping or not keeping every trivial promise: we will do the most good and reserve our mental energies for more important matters

that a consequentialist mode of deliberation should be used in most circumstances, albeit with some exceptions. I call this form of sophisticated consequentialism *moderately sophisticated consequentialism*.

As we move toward the other end of the scale of sophistication, we see something of the opposite sort. Specifically, we observe that more sophisticated forms of consequentialism advise against using a consequentialist mode of deliberation in most circumstances. At the furthest reaches of this end of the scale of sophistication, we encounter a form of consequentialism that prescribes that one should *never* employ a consequentialist mode of deliberation—not even as a form of meta-deliberation.[34] I call this form of sophisticated consequentialism *radically sophisticated consequentialism* or R-SC.

I think that there are two main ways to grasp how sophisticated consequentialist theories are self-effacing theories. First, one might say that moderately sophisticated consequentialism is self-effacing in the following sense: it advises agents that occasionally, they best aim for the good(s) that sophisticated consequentialism prescribes by abstaining from using a consequentialist mode of deliberation, or any mode of deliberation at all, when acting. Moderately sophisticated consequentialism might offer this kind of negative guidance with regard to situations like emergencies. Railton sketches a useful picture of what such guidance might look like, writing

> A sophisticated act-consequentialist may recognize that if he were to develop a standing disposition to render prompt assistance in emergencies without going through elaborate

---

if we simply habituate ourselves to keep promises in all normal situations." J.C.C. Smart, "An outline of a system of utilitarian ethics," in *Utilitarianism: For and Against* by J.C.C. Smart and Bernard Williams (Cambridge: Cambridge University Press, 1973), 43.

[34] Railton indicates that this radical variety of sophisticated consequentialism is one that an objective consequentialist would recognize, if not as plausible, then at least as possible. Discussing an objective rule-consequentialist, Railton writes, "an *objective rule*-consequentialist sets actual conformity to the rules with the highest acceptance value as his criterion of right action, recognizing the possibility that the best set of rules might in some cases—or even always recommend that one not perform rule-consequentialist deliberation." Railton, "Alienation," 156.

> act-consequentialist deliberation, there would almost certainly be cases in which he would perform acts worse than those he would have performed had he stopped to deliberate […] It may still be right for him to develop this disposition, for without it he would act rightly in emergencies less often—a quick response is appropriate much more often than not, and it is not practically possible to develop a disposition that would lead one to respond promptly in exactly those cases where this would have the best results.[35]

Second, one might say that radically sophisticated consequentialism is self-effacing in the following sense: it advises agents that they best aim for the good(s) that sophisticated consequentialism prescribes by ceasing to be consequentialists.

The main difference between these two interpretations is that moderately sophisticated consequentialism allows for sophisticated consequentialist agents to have a "commitment to leading an objectively consequentialist life,"[36] whereas radically sophisticated consequentialism does not. For it seems that a person has a commitment to leading an objectively consequentialist life *if and only if* he is a consequentialist. To be sure, this is in no way to imply the following: that one can only lead an objectively consequentialist life if one has a commitment to doing so. For radically sophisticated consequentialism proscribes having such a commitment precisely on

---

[35] Railton, 157.

[36] Railton, "Alienation," 153. In Railton's view, it seems that a person has a commitment to living an objectively consequentialist life if his "motivational structure" conforms to the following "*counterfactual condition*: he need not always act for the sake of happiness, since he may do various things for their own sake or for the sake of others, but he would not act as he does if it were not compatible with his leading an objectively hedonistic [or, in this case, consequentialist] life." Railton, 145. I take it that if one is to be accurately described as having a "commitment" to consequentialism or to living an objectively consequentialist life, then Railton's counterfactual condition must be normative rather than psychological. According to William Wilcox, "someone with a normative counterfactual condition in his motivational structure believes that were certain things to happen, he *ought* to change in specific ways […] The normative counterfactual "condition" consists of a normative judgement and a counterfactual (or at least condition) *intention* derived from that judgement." William H. Wilcox, "Egoists, Consequentialists, and Their Friends," *Philosophy & Public Affairs* 16, no. 1 (Winter 1987), 77. In contrast, a psychological counterfactual condition involves a reflexive aversion to a particular phenomenon, such as pain. For example, "an effective egoistic hedonist could be someone whose psychological counterfactual condition is unusually effective. Such a person can then be aware that if his caring about a friend for the friend's sake becomes more painful than pleasant […] then as a matter of psychological fact his concern will very likely cease." Wilcox, 77. Along with Wilcox, I hold that a person's motivational structure must have a normative consequentialist counterfactual condition in order for them to qualify as being a consequentialist and as having a commitment to leading an objectively consequentialist life. To quote Wilcox, "we must understand this counterfactual as being normative rather than merely psychological if the agent is to be a consequentialist rather than merely someone who acts as a consequentialist would require." Wilcox, 79.

the grounds that having it will preclude one from living an objectively consequentialist life. In this way, radically sophisiticated consequentialism is a more radically self-effacing kind of sophisticated consequentialism than its moderate counterpart. For moderately sophisticated consequentialsm merely tells agents that they should not be *preoccupied* in certain situations with their commitment to living an objectively consequentialist life, but radically sophisticated consequentialism tells agents that they should not even have such a commitment at all. In what follows of this chapter, I will be concerned strictly with radically sophisticated consequentialism. Now I will move to discuss a second self-effacing theory: Julia Annas's developmental virtue ethics.

2.3.2 Developmental Virtue Ethics (DVE)

Annas's DVE is a self-effacing ethical theory, because it denies that we should *always* be guided by the decision procedure it provides us. I next discuss, in turn, three aspects of how DVE is self-effacing: (1) the good(s) DVE tells people to achieve, (2) DVE's decision procedure, and (3) the circumstances in which DVE advises agents that they should not use DVE's decision procedure.

First, consider which good(s) DVE tells people to achieve. DVE tells people that they should live a virtuous life and be virtuous people. Further, since being a virtuous person involves behaving like a virtuous person, DVE tells people that they should do what the virtuous person characteristically would do. It also tells them that they should do what the virtuous person would do *as* the virtuous person would do it.

DVE provides a decision procedure that connects with what DVE tells people to value. This decision procedure consists of applying the following criterion of right action as a way to answer the question "what should I do": An action is right if and only if it is what a virtuous

person would do.[37] Importantly, the value of this decision procedure lies not just in its ability to help agents decide what to do in particular situations. It also lies in the fact that it helps people who use it to lead more virtuous lives and to become more virtuous people.

According to DVE, becoming a more virtuous person involves acquiring a richer understanding of what it means to be virtuous and do the right thing. As a person's knowledge of what it means to be virtuous becomes more complex, this person will cease to rely as much on DVE's decision procedure when deciding what to do. The idea here is that the more virtuous a person becomes, the less often they will feel themselves motivated to act by an *explicit* desire to do what the virtuous person would characteristically do. To quote Annas,

> a beginner in virtue will have to try explicitly to become a virtuous person, and to do so by doing virtuous actions; his deliberations will include such thoughts as that so and so is what a virtuous person would do, or what virtue requires. This is, indeed, how he guides his deliberations. The truly virtuous person, however, will not explicitly think about, for example, being brave or performing a brave action. Rather, he will, as a result of experience, reflection, and habituation, simply respond to the situation, thinking that these people in danger need help, without explicit thoughts of bravery entering his deliberations. Thoughts about bravery, or the virtuous person, are no longer needed.[38]

---

[37] Annas, *Intelligent Virtue*, 41. See also Rosalind Hursthouse, *On Virtue Ethics* (Oxford: Oxford University Press, 1999), 28. A common objection to this decision procedure is that it is too vague and too simple to be helpful to people when they are deciding what to do. In response to this worry, I think proponents of developmental virtue ethics have two main defenses. First, they can argue that this objection lacks sufficient motivation: it charges developmental virtue ethics with failing to provide a helpful decision procedure, yet it fails to specify a clear and uncontroversial standard for what a helpful decision procedure should look like. In *Intelligent Virtue*, Annas makes this same point, writing: "It is […] completely unclear how we are to assess objections like this, given that there is no agreement as to how specific, if at all, ethical directions should be. We can agree that ethical direction will be useless if too general […] and also that it is unrealistic and unmotivated for a theory to be required to tell us exactly what to do, at a level leaving no room for our own judgement. But there is no agreement as to how specific ethical direction should be, and so no independent traction to an objection that theories in which virtue is central give us ethical direction in too general and unspecific a way." Annas, *Intelligent Virtue*, 35. Second, proponents of developmental virtue ethics can reply that even if we accept that this decision procedure is vague when presented by itself and without exegisis, this procedure will become more precise once we specify the nature of the virtuous person. To specify the nature of the virtuous person, one might begin by presenting two principles. First, "a virtuous agent is one who has, and exercises, certain character traits, namely, the virtues." Second, "a virtue is a character trait that …" Hursthouse, *On Virtue Ethics*, 29. Different forms of virtue ethics will complete the second premise in different ways. The version of developmental virtue ethics that Annas discusses is eudaimonistic, so it holds that a virtue is a character trait that is necessary for eudaimonia, a happy life. Annas, *Intelligent Virtue*, 167. For the time being, I think that these two replies to the vagueness objection are sufficient to deflate its force.

[38] Annas, "Virtue Ethics and the Charge of Egoism," 212.

In Annas' view, "the self-effacingness here is as harmless as it is with a practical skill."[39] Hence, just as thoughts about how to swing a tennis racquet will become less explicit in the minds of a tennis players as their expertise grows, so too will thoughts about what the virtuous person would do be less explicit in the minds of people of deepening virtue.

DVE is then self-effacing by holding that agents will eventually come to rely very little upon its decision procedure. For in order to become virtuous people, agents will typically have to become less reliant on DVE's decision procedure by developing a richer understanding of what it means to be virtuous and to act as a virtuous person would.

2.4 Defense of R-SC and DVE

I argue that Keller's, Pettigrove's, and Martinez's objections to self-effacing ethical theories fail with regard to R-SC and DVE. Specifically, I maintain that each objection rests in various ways on the mistaken idea that all self-effacing theories fail to fulfill the practical aim of the DPF. Below, I address how these objections fail in turn.

To recap, Keller's objection to self-effacing ethical theories runs as follows: If an ethical theory does not tell us what should motivate us, then it does not tell us how to live; and if it does not tell us how to live, then it is not a good ethical theory. Now, since self-effacing ethical theories do not tell us what should motivate us, they do not tell us how to live. Hence, they are not good ethical theories. I think that this objection is too quick. While self-effacing theories such as R-SC and DVE tell us, in some sense, that we should not be motivated to act by certain desires, this does not entail that they tell us nothing about what should motivate us.

It is true that R-SC tells us that we should not be consequentialists and, by the same token, should not be committed to living an objectively consequentialist life. Thus, if we were to

[39] Annas, 212.

adopt an agent-neutral, eudaimonistic form of R-SC, then our ultimate and exclusive motive in life should *not* be to maximize the overall happiness of humankind. Despite this fact, I think that there are two clear senses in which R-SC does tell us what should motivate us. First, R-SC tells us that we should be motivated to act by whatever motive(s) will, among those available to us, be most conducive to our maximizing the good in the long run. The second sense in which R-SC tells us what should motivate us concerns the consequentialist commitment it tells us to eliminate. What should motivate a proponent of R-SC to eliminate his commitment to consequentialism? R-SC answers this question in the following way: A proponent of R-SC should be motivated to rid himself of his commitment to consequentialism *by* his commitment to consequentialism.

Keller's objection does not just fail against R-SC; it misses the mark with DVE as well. It is true that DVE gives us the following self-effacing guidance: we should develop a richer understanding of what it means to act and live virtuously, which will result in our being less frequently motivated to act by an *explicit* desire to do what the virtuous person characteristically would do. However, by providing agents with this guidance, DVE does not cease to tell agents what should motivate them. DVE merely tells agents that their motivation to live virtuously and become virtuous people will become less explicit in their minds as they become more virtuous people. If Keller's objection led us to worry that self-effacing theories, such as DVE and R-SC, are inadequate theories because incomplete theories, then we should worry no longer. For I have just shown how to complete them satisfactorily.

Unlike Keller, Pettigrove worried not that self-effacing theories would be incomplete but that they would provide a complete but faulty decision procedure. Pettigrove's concern is that this decision procedure will be inadequate. For Pettigrove, an adequate decision procedure is one

that an agent should not "need to put out of her mind when she acts."[40] If an ethical theory cannot provide a decision procedure that works in this way, then, on Pettigrove's view, there is something problematic about that theory.

Pettigrove's objection is flawed. To see why, consider why Pettigrove thinks it is problematic for an ethical theory to advise agents to put its decision procedure out of their minds when acting. He thinks that ethical theories are supposed "to help an agent answer the question 'What should I do?'" and he thinks that an ethical theory's decision procedure is "the theory's chief tool for answering the question." [41] Pettigrove worries that an agent who must put his ethical theory's decision procedure out of his mind when acting will be unable to identify the right thing to do, let alone be able to realize the right thing. With regard to R-SC and DVE, this worry is unwarranted.

While R-SC requires that agents always put a consequentialist decision procedure out of their minds when acting, it does not leave them without a way to identify the right thing to do. On the contrary, R-SC tells agents to develop a variety of alternative decision procedures for deciding the right thing to do. R-SC does this because it calculates that an agent who uses such alternative decision procedures will overall bring about better consequences than would an agent who uses a consequentialist decision procedure.

---

[40] Pettigrove, "Is Virtue Ethics Self-Effacing?" 192. Pettigrove does not go into detail about what it means to put an ethical theory's decision procedure out of one's mind when one is acting, or what it means for one to either have or not have a theory's decision procedure in mind when acting. Nevertheless, I think that such phrases call for some clarification. To this end, it will be helpful to refer back to Timmons' description of the practical aim of ethical theorizing; specifically, the part of his description that concerns the purpose of an ethical theory's decision procedure. According to Timmons, one purpose of a decision procedure is to help an agent decide what to do "in contexts of moral deliberation and choice." Timmons, *Moral Theory*, 3. Contexts of this character are contexts in which one asks oneself: what is the right thing (for me) to do here? In such contexts, reflecting on an ethical theory's decision procedure is supposed to help us identify the right thing to do. Thus, to say that a person has a decision procedure *D* in mind when acting is merely to say that said person, when faced with a choice of what to do, reflects upon *D* and chooses in accordance with *D*. In contrast, to say that a person does not have *D* in mind when acting is to say that said person, when faced with a choice about what to do, neither reflects upon *D* nor deliberately seeks to choose in accordance with *D*, though the choice said person makes might indeed accord with *D*.

[41] Pettigrove, "Is Virtue Ethics Self-Effacing?" 192.

Further, while DVE advises agents to develop a richer understanding of virtue, and eventually rely less on its decision procedure when acting, this advice does not diminish the ability of such agents to identify the right thing to do. DVE provides this advice because it is a theory that expects that agents with a more developed understanding of virtue will not have as much of a *need* for a decision procedure to help them identify the right thing to do. When someone who is just beginning to learn how to write must spell the word "receipt," they may need to remind themselves: "I before E except after C." Yet it would seem odd if a professional typist were to need phrases like this to help him identify the right way to spell "receipt." In a similar sense, when someone who is just beginning to learn how to be virtuous faces an unfamiliar decision, he may need to rely on DVE's decision procedure to ascertain the right thing to do. However, if a more virtuous person were faced with the same choice, then she would not find it as difficult to discern the right thing to do. Having worked long and hard to be virtuous, she has developed a capacity for "practical wisdom, or a capacity to perceive the feature of a situation most worthy of being acted on" without external guidance from a decision procedure.[42]

Like Pettigrove, Martinez argues that there is something problematic about an ethical theory that requires agents to put its decision procedure out of their minds when they are acting. Unlike Pettigrove, however, Martinez takes issue with self-effacing theories on the grounds that they do not provide decision procedures that give "*ongoing* concrete advice […] about particular motives and viewpoints through reflective revision informed in part by the theory and in part by

[42] Clark, "Eudaimonistic Virtue Ethics and Self-Effacement," 515. See also Annas, *Intelligent Virtue*, 28-30. There, Annas writes, "the virtuous person's response [to a crisis] is immediate, but it is an *intelligent response* […] For one important lesson we learn from the analogy with skill is that reasons for acting can efface themselves without evaporating entirely. Moreover, it is the fact that these reasons cease to take up psychological room at the time of action which enables the virtuous person to become someone who is disposed to act generously or sympathetically without hesitation or the need to work out the options."

the normative considerations that are salient in particular circumstances."[43] As I now discuss, this objection also fails against R-SC and DVE.

Since R-SC advises agents that they should cease to be consequentialists, there is a sense in which R-SC fails to provide agents with *some* ongoing moral advice. R-SC fails to provide ongoing *consequentialist* advice. R-SC nevertheless provides other ongoing advice to agents. R-SC helps agents to identify (and also prescribes they adopt) certain alternative non-consequentialist decision procedures —namely, those that, once adopted by the agent, would be more conducive to the R-SC's commitment to living an objectively consequentialist life than the alternative of following a consequentialist decision procedure.

Martinez might object that by equipping agents with a set of alternative decision procedures and sources of advice and by effacing itself from the scene, R-SC does not actually provide agents with ongoing advice; rather, the alternative non-consequentialist decision procedures provide guidance. Yet if Martinez is right about this, his objection to self-effacing theories seems unjustifiably biased against consequentialist theories in general and sophisticated consequentialist theories in particular. As Railton notes, objections such as Martinez's appear to arbitrarily "require that one class of actions […] *not* be assessed in terms of their consequences."[44] Specifically, Martinez's objection seems to require that acts like developing, maintaining and giving up moral commitments as well as selecting and employing decision procedures not be assessed in terms of their consequences. Martinez thus demands without justification, and against the ultimate consequentialist goals of R-SC, that the theory provide agents with an R-SC decision procedure.

[43] Martinez, "Is Virtue Ethics Self-Effacing?" 281.
[44] Railton, "Alienation,"155.

When applied to DVE, Martinez's objection misses the mark as well. While DVE does tell us that agents will rely less *explicitly* on DVE's decision procedure as they become more virtuous, this does not mean that DVE ceases to provide agents with ongoing advice. While agents who are more virtuous should not rely as often on the decision procedure that DVE provides, they may encounter situations that powerfully test their commitment to acting virtuously and to becoming virtuous people. Virtuous persons in such rare situations would be justified in relying on DVE's decision procedure for help with identifying the right thing to do.

Martinez might object that DVE fails to provide agents with ongoing advice after they become fully virtuous, for it would seem that the fully virtuous person would never be justified in relying on DVE's decision procedure to decide what to do.[45] In reply I note that this objection elides how DVE provides inexplicit advice to fully virtuous agents with the claim that DVE provides no advice to them at all. The fully virtuous agent does not receive *explicit* ongoing advice from DVE. When faced with what Timmons calls "contexts of moral deliberation and choice,"[46] the fully virtuous person will rarely if ever reflect upon DVE's decision procedure and then try to choose in accordance with it. Instead, the fully virtuous person will simply come to see the right thing to do by assessing the details of the situation at hand. Even so, for the fully virtuous person thoughts about DVE's decision procedure have not "have not evaporated" but "could be recovered if needed—if, for example, the person needs to explain what he did afterwards, say, to a child who is being taught" how to be virtuous.[47] In this way, the fully

[45] Here, I pass over the idea that the fully virtuous person is perhaps an *ideal* to which we should aspire rather than something we can actually become. Julia Annas considers the idea of the virtuous person in this way. On her view, "being *fully* virtuous does seem to be an ideal that we aspire towards but can never achieve. At best we can be virtuous in a less than full way, one marred not only by our own deficiencies but by the point that the structures of our societies preclude us from being fully virtuous." Annas, *Intelligent Virtue*, 64.
[46] Timmons, *Moral Theory*, 3.
[47] Annas, *Intelligent Virtue*, 159.

virtuous person is *implicitly* guided by DVE's decision procedure insofar as the fully virtuous person represents the fullest embodiment and personification of this procedure.

2.5 Conclusion

In this chapter, I resisted objections that self-effacing ethical theories are inadequate. I showed that many leading objections have no traction against R-SC and DVE. Against the objections of Keller, Pettigrove, and Martinez, I argued that R-SC and DVE may be able to tell us what should motivate us and how we should live.

3 WHAT IS WRONG WITH SELF-EFFACING ETHICAL THEORIES

In the previous Section, I defended two self-effacing ethical theories, radically sophisticated consequentialism (R-SC) and developmental eudaimonistic virtue ethics (DVE), against three different objections to self-effacing ethical theories. In this Section, I do something of the opposite sort: I raise two new objections linked to self-effacement. I show R-SC is vulnerable to what I call the *self-erasure objection*. Further, I show DVE is vulnerable to an *incompleteness objection*. Below, I outline and analyze both of these objections in turn. I close this chapter by arguing that the problems DVE faces are less worrisome and more surmountable than those R-SC faces.

3.1 The Self-Erasure Objection

In this Section, I sketch the self-erasure objection and examine the extent to which it poses problems for self-effacing ethical theories, such as R-SC and DVE. I begin by describing what I intend by calling an ethical theory *self-erasing*. Next, I examine some ways in which the charge of self-erasure constitutes an objection to R-SC. I conclude by arguing that the charge of self-erasure is less of a problem for DVE than it is for R-SC.

When I call an ethical theory a self-erasing ethical theory, I mean in part that it is a particular type of *self-changing ethical theory*. An ethical theory is self-changing if it instructs us to change some component of our conception of self.[48] An ethical theory might recommend self-change in a variety of ways. For example, it might ask us to integrate new self-components into our conception of self, or to give up or reinterpret self-components already present in our

[48] I use the phrase "conception of self" in reference to "the notion that each of us has […] an idea or set of ideas of who we are, and that this conceptualization is relatively constant over time." Daphna Oyserman, "Self-Concept and Identity," in *The Blackwell Handbook of Social Psychology*, ed. A. Tesser and N. Schwarz (Malden: Blackwell, 2001), 500.

conception of self. Additionally, ethical theories that prescribe self-change can do so in degrees. A self-changing ethical theory might instruct us to get rid of a relatively unimportant component of our conception of self (i.e., our love of gas-guzzling SUVs); whereas another self-changing theory might tell us to significantly modify or even replace our conception of self.

Self-erasing theories are a particular type of self-changing theory. They are distinguished from other types of self-changing theories by the variety and degree of self-change they instruct us to perform. In effect, self-erasing theories require us to give up some component(s) of our conception of self and to thereby undergo a radical degree of self-change. Put differently, they tell us to give up or erase a *fundamental component* of our conception of self. When I call a component of our conception of self "fundamental," I mean that it is a constitutive component of our conception of self, such that if it were erased we would seem like a different person.[49] As a result, self-erasing ethical theories effectively instruct us to erase who we are.

The radical form of Railton's sophisticated consequentialism (R-SC), which I discussed and defended in Section 1, is a self-erasing ethical theory. Specifically, it is a self-erasing theory to the degree that its defining prescription is directed toward people who are consequentialists, that is, people who have a commitment to living an objectively consequentialist life. To see why, recall that the defining prescription of R-SC is the following: In order to maximize the good, one should cease to be a consequentialist, that is, should cease to have a "commitment to leading an objectively consequentialist life."[50] I stipulate that if a person has a commitment to leading an

[49] Here and below, I use phrases like "seem like a different person" as colloquialisms, reflecting a common meaning: these phrases often pick out a change in the qualitative identity of a person. For example, suppose you encounter a old friend whom you have neither seen nor kept in contact with for a long time. In your memory, your friend may have been a devout believer of a particular religious faith. However, upon meeting him again, you observe that he is not a believer anymore (or perhaps he has converted to a very different faith). You might say to this friend, "wow, you seem like a totally different person." In this situation, you are pointing out a significant change in the qualitative identity of your childhood friend: he *seems* dissimilar to the childhood friend of your memories, but you recognize that, at bottom, he is still the same person.

[50] Railton, "Alienation," 153.

objectively consequentialist life, then this commitment is fundamental to his conception of self.[51] Thus, if we take R-SC's defining prescription to be directed toward people who are consequentialists, then R-SC is a self-erasing theory. It instructs such people to erase or relinquish a fundamental component of their conception of self, namely, their commitment to leading an objectively consequentialist life.

By counseling self-erasure, R-SC places consequentialist agents in a paradoxical position. For R-SC tells them to maintain their commitment to leading an objectively consequentialist life by erasing it and/or replacing it with a new, more optimific commitment. Hence, if a consequentialist were to successfully follow the guidance of R-SC, then he would thereby cease to be a consequentialist and, by the same token, cease to be committed to leading an objectively consequentialist life. What's more, if a consequentialist agent were to choose to act against the self-erasing guidance of R-SC to preserve his commitment to consequentialism, then he would, by making this kind of anti-consequentialist decision, effectively undermine his commitment to consequentialism. For by R-SC's standards, such a decision would be patently immoral.

There is a consequentialist response to the worry that R-SC places consequentialists in a paradoxical position. This response is that the position in which R-SC places consequentialists is really not so paradoxical as it is complex. R-SC does not require consequentialists to contradict their commitment to leading an objectively consequentialist life. For to be committed to leading

[51] Recent studies in psychology and experimental philosophy lend support to the idea that a person's moral commitments, such as a commitment to consequentialism, are an exceptionally important part of their conception of self—indeed, perhaps even more important than memories and preferences. While the authors of these studies disagree about whether changes in a person's moral commitments elicit changes in their numerical identity or their qualitative identity, they agree that changes in a person's moral commitments constitute a significant change in their conception of self. See Nina Strohminger and Shaun Nichols, "The essential moral self," *Cognition* 131 (2014): 159-171; Jesse J. Prinz and Shaun Nichols, "Diachronic Identity and the Moral Self," in *The Routledge Handbook of Philosophy of the Social Mind*, ed. Julian Kiverstein (Abingdon: Routledge, 2016), 449-464; and Paul Bloom and Christina Starmans, "Nothing Personal: What Psychologists Get Wrong about Identity," *Trends in Cognitive Sciences* 22, no. 7 (July 2018): 566-568.

an objectively consequentialist life is to be committed strictly to living a life "that would bring about the best outcomes."[52] This is different than living a life in which one has a commitment, which one retains over the course of their life, to living an objectively consequentialist life. In this way, the guidance that R-SC provides is complex and perhaps even counterintuitive, but it is not paradoxical; it does not involve self-contradiction or absurdity.

A more dangerous problem than paradoxicality for self-erasing theories like R-SC has to do with the notion that people ought to be treated as ends in themselves. Specifically, by advising consequentialist agents to erase a fundamental component of their conception of self—a component of their self that is constitutive of who they are—R-SC appears to contradict the idea the people ought to be valued as ends in themselves. Of course, when considered with regard to consequentialist theories in general, this accusation in not an uncommon one. For example, Bernard Williams has argued that to look at the world through a consequentialist lens requires that one "come to regard as one satisfaction among others, and a dispensible one, a project or attitude [that is, a commitment] round which he has built his life."[53] By requiring a person to view his commitments in such a light, consequentialism functions "to alienate him in a real sense from his actions and the source of his actions in his own convictions. It is to make him into a channel between the input of everyone elses projects, including his own, and an output of optimific decision."[54] However, when this accusation is applied to R-SC, I think it takes on a novel aspect. While consequentialists are commonly accused of failing to value other people as ends in themselves, less often are they criticized for failing to value themselves as such. Yet this is exactly the oddity the R-SC brings to the fore. R-SC takes consequentialist logic to its limits,

[52] Railton, "Alienation," 152.
[53] Williams, "A critique of utilitarianism," 116.
[54] Williams, 116.

requiring consequentialist agents to not only treat other individuals as expendable means to aggregate goods, but also to treat their greatest commitment, a fundamental component of their very self, as perfectly fungible as well. By offering this kind of self-erasing guidance, R-SC seems to render itself absurd because highly counterintuitive.

I think that there are two main consequentialist responses to this worry. The first response accepts that the self-erasing guidance that R-SC offers is counterintuitive enough to render R-SC untenable as a theory; yet it counters that it is implausible that a consequentialist theory would ever offer such guidance. I call this *the opportunity cost response*. The opportunity cost of erasing a fundamental component of one's conception of self outweighs the goods that this erasure would allow one to realize.[55] Much of the force of this response lies in the idea that it would be exceptionally difficult, time-consuming, and risky for a person to erase and revise fundamental components of their conception of self.

Perhaps some form of extended identity therapy or brainwashing might be required for the consequentialist agent. [56] While the phrase "brainwashing" may have negative connotations for many of us, for the sophisticated consequentialist, such connotations do not obtain. For the person who desires more than anything to live an objectively consequentialist life, brainwashing seems to be a welcome means to erasing whatever aspect of oneself is not maximally value-

[55] Railton might be read as suggesting a similar objection, when he writes: "while one can attempt to cultivate dispositions that are responsive to various factors […] refinements have their own costs and, given the limits of human resources, even the best cultivated dispositions will sometimes lead one astray." Railton, "Alienation," 157.
[56] Parfit raises a similar idea in *Reasons and Persons*. He writes, "suppose that C told us to believe some other theory. As I have said, it would be hard to change our beliefs, if our reason for doing so is not a reason which casts doubt on our old beliefs, but is merely that it would have good effects if we had different beliefs. But there are various ways in which we might bring about this change. Perhaps we could all be hypnotized, and the next generation brought up differently. We would have to be made to forget how and why we acquired our new beliefs, and the process would have to be hidden from future historians." Derik Parfit, *Reasons and Persons* (New York: Oxford University Press: 1984), 41.

promoting. Importantly, the opportunity cost response shows that consequentialists have no objection to the self-erasing guidance that R-SC might provide.

If the self-erasure R-SC might prescribe were easy to bring about, then the opportunity cost objection would lose much of its force, if not dissolve completely. Consider an example I call *Identity Optimizer (IO)*:

> *Identity Optimizer (IO) is* a device with two main functions. (1) when connected to a person, IO can quickly and effectively adjust or reprogram that person's conception of self. (2) IO can calculate and implement whatever motivations, dispositions, character traits, etc., would optimize the connected-person's ability to maximize the good.

IO would test the commitment of any consequentialist to consequentialism, for IO allows whichever consequentialist agent who connects to it to effectively achieve his final goal to the best of his ability. The reason that IO would test a consequentialist's commitment to consequentialism is that by connecting to IO one risks erasing whatever component of one's conception of self is not conducive to one's leading an objectively consequentialist life. In this way, connecting to IO comes with a catch for consequentialists. While connecting to IO allows the consequentialist to effectively achieve his final goal of living an objectively consequentialist life, the catch is that, when he disconnects from IO, he may no longer recognizes this goal as *his* goal; it may become rather the past commitment of a former self who seems completely different from him.

I think that IO reveals to us three important insights. First, IO shows us that the self-erasing guidance offered by R-SC is not unique to R-SC; rather it is guidance that many (and perhaps all) consequentialist theories would offer to an agent given the right circumstances. Second, IO shows us that the self-erasing guidance offered by R-SC is comprehensible yet counterintuitive to us. We understand that, for most consequentialist agents, it would make perfect sense to connect to IO in order to live a maximally value-promoting life, but we feel

uncomfortable at the thought of having to do so ourselves. Put differently, though we may grasp a consequentialist's motivation for connecting to IO, this motivation does not appear to be one that we share. Third, IO shows us that the opportunity cost response to the self-erasure objection is, in some sense, a red herring. For the opportunity cost response asks us to consider how very unlikely it is that a consequentialist theory would ever prescribe the kind of self-erasure that R-SC prescribes. But our initial worry about self-erasing guidance was not a worry about likelihood or frequency, it was a worry about *propriety*. There seems to us to be something gravely wrong about a theory that would advise us to erase a fundamental component of our conception of self, and the fact that it is unlikely that a consequentialist theory would offer such advice does not change the fact that, given the right conditions, it *would* offer it.

Why do we feel averse to the sort of self-erasure that R-SC recommends? While I do not have the space here to provide an extended response to this question, I nevertheless would like to present one hypothesis for why we might have an aversion to erasing a fundamental component of our conception of self, i.e., a moral commitment. This possibility has to do with the notion that our conception of self plays an important role in how we navigate social life and engage with members of our community. On this view, our conception of self is "a social force: it influences what is perceived, felt, and reacted to and the behavior, perceptions and reactions of others."[57] It may be that when we erase or alter a fundamental component of our conception of self, our behavior becomes less predictable to members of our community; in some cases, this might result in mutual misunderstanding, social ostracization, or worse. If this is true, then a willingness to alter fundamental components of one's conception of self might be a trait that

[57] Oyserman, "Self-Concept and Identity," 502.

would be evolutionarily selected against, whereas an aversion to altering fundamental components of oneself could be a more desireable trait.

The second consequentialist response to the self-erasure objection is to challenge whether the objection counts against consequentialism. The critic—who has said "there seems to us to be something gravely wrong about a theory that would advise us to erase a fundamental component of our conception of self"—seems to deny a consequentialist moral theory. From a consequentialist standpoint, not only are one's actions only as good as the states of affairs they bring about, so too are one's conception of self, one's life projects and commitments, one's character and motivations.[58] Thus, it may be true that for a hardcore sophisticated consequentialist, the self-erasure objection is hardly an objection at all.

One might wonder whether DVE is a self-erasing theory as well. After all, like R-SC there is a sense in which agents who follow DVE's guidance and develop a richer understanding of virtue appear to become different people as a result. Indeed, one might say that both SC and DVE require agents to change who they are to a substantial degree. A sophisticated consequentialist who connects to IO might be a very different person when she disconnects. In a similar way, one might say that someone who has developed a rich understanding of virtue over time may seem significantly different from the person she was when she first started trying to be virtuous.

[58] A similar point is made by Bernard Williams, who writes: "As a utilitarian […] he must be concerned with the utility in the world—and if utility can leak into or out of the world by channels which do not run totally through acts, then he would be mad to take no account of them." Williams, "A critique of utilitarianism," 120. In *Reasons and Persons*, Parfit makes a comparable observation not just about utilitarian theories but about consequentialist theories in general. He writes, "consequentialism covers, not just acts and outcomes, but also desires, dispositions, beliefs, emotions, the color of our eyes, the climate and everything else. More exactly, C covers anything that could make outcomes better or worse." Parfit, *Reasons and Persons*, 25.

Despite these similarities between R-SC and DVE, DVE is not a self-erasing theory. R-SC tells fully developed consequentialist agents to relinquish what is perhaps their most fundamental commitment; namely, their commitment to living an objectively consequentialist life. In contrast, DVE tells agents who are just beginning to learn how to be virtuous to develop a commitment with which they already identify themselves; namely, their commitment to living a virtuous life. Thus, although DVE does urge that one change one's commitments, it urges this in the same way that a gardener urges the seeds she sows to change into trees. In this way, the content of one's commitment to virtue becomes broader and deeper as one becomes more virtuous, but its form remains the same over the course of one's ethical development.[59]

3.2 The Incompleteness Objection

While DVE has the merit of not being a self-erasing theory, it is vulnerable to a different objection. This is the objection that DVE does not equip agents with a complete account of either virtue or eudaimonia.Imporatntly, my incompleteness objection is different than the objection that Keller levels against self-effacing ethical theories. Keller's objection contends (incorrectly) that self-effacing ethical theories are incomplete, because they do not provide us with a decision procedure and, as a result, do not fulfill the practical aim of the DPF. On the other hand, my incompleteness objection proposes that there may be a sense in which DVE does not fulfill the theoretical aim of the DPF.

To ground my objection, it will be useful to highlight two points about virtue and happiness that Annas emphasizes throughout *Intelligent Virtue*. The first point has to do specifically with virtue. According to Annas, "we cannot understand what a virtue is without

[59] Clark makes a similar point about DVE. He describes DVE as "*self- embracing,*" for as "some of the theory's reasons, values, and justifications are effaced from the agent's motive, the theory offers additional reasons, values, and justifications for the agent to embrace." Clark, "Eudaimonistic Virtue Ethics and Self-Effacement," 511.

coming to understand how we acquire it."[60] This claim admits of multiple interpretations. One way to read it is as implying the following claim: One cannot understand what a virtue is before one develops it. This interpretation of Annas's claim is, I think, unacceptable; for Annas maintains that people typically have some understanding of virtue even before they start trying to be virtuous.[61] However, there is another way to read Annas's claim about the relationship between virtue and development. Namely, Annas can be read as saying that we cannot achieve a *complete* understanding of what a virtue is or what the virtues are prior to developing into virtuous people. Indeed, it seems that rather than provide us with a complete account of virtue, DVE merely promises us such an account. It tells us that we will eventually fully understand the nature of virtue and the virtuous life but only after we have become virtuous ourselves. A similar problem arises in Annas's discussion of eudaimonia.

In *Intelligent Virtue*, the sort of happiness (or eudaimonia) with which Annas is concerned is not "of the pleasure- or desire- or life-satisfaction kinds."[62] On Annas's view, happiness serves as an overall end that cannot be defined independently of virtue. She writes, "[h]appiness is the unspecific overall aim that we find that we have in mind in some form in what we are doing. What we take it determinately to consist in is not given in advance of our becoming virtuous."[63] In sum, this is to say that as we deepen our understanding and practice of virtue, "our conception of happiness *becomes* more determinate."[64] In other words, the process of becoming virtuous is itself a process of becoming more aware about "what matters in [one's life]

[60] Annas, *Intelligent Virtue*, 23. Annas reiterates claims like this throughout *Intelligent Virtue*. For example, she writes that "[v]irtue, on an account of this kind, cannot be properly understood without understanding how it comes about and develops." Annas, 83.
[61] Annas writes, "[b]y the time we think about the virtues, we all have some […] For we have all been brought up, and that is where we have learned to be virtuous (or not, or not very)." Annas, 21.
[62] Annas, 155.
[63] Annas, 156.
[64] Annas, 156.

and what [one's] view is of [one's] happiness."[65] What this discussion of happiness tells us is that Annas connects happiness to development in a manner similar to the way in which she connects virtue to development. In both cases, Annas seems to tell us that we should not be concerned if DVE does not provide a full account of virtue and happiness. For such accounts are available only to the truly virtuous themselves.

It does seem that DVE does not provide all agents with a complete account of virtue and happiness. Whether this is a problem turns on how we interpret the phrase "does not provide." Suppose we interpret "does not provide" as meaning that DVE does not *immediately* equip agents with a fully fleshed out account of what it means to be virtuous and to live a happy life. On this interpretation, DVE does not provide all agents with a complete account of virtue and happiness. On an alternative reading, however, we might interpret the phrase "does not provide" as meaning that DVE *never* provides nor promises to provide all agents with a complete account of virtue and happiness. As we have seen, though, DVE does, in fact, provide an account of virtue and the happy life. The catch, however, is that one cannot gain access to this account without a significant amount of effort; that is, one must become virtuous oneself to acquire it.

3.3 Reconsidering the Incompleteness Objection: A Case for Development

It may not be so great a problem that DVE is vulnerable to the incompleteness objection. I take it that completeness is a feature that one desires in an ethical theory, because this feature helps one achieve clear answers to questions like what should I do and why should I do it. However, DVE is not a complete theory in this way. Instead, it provides agents with a roadmap for attaining answers to ethical questions, but it does not present them with a cheat sheet for such questions. To quote Annas, this roadmap "guides us […] toward emulating people who are

[65] Annas, 160.

braver, more generous and generally better than we are, and does so in a way which recognizes the contstaints put on this by the level of our development as well as that of the people we emulate."[66] Similarly, the conceptions of virtue and happiness that DVE provides to agents who are novices at virtue are not determinate. One reason for this is that, for a long period of our youth and even our adulthood, we may feel that we desire happiness but are not quite sure what we mean by happiness. In typical cases, however, we start with a vague idea of what a happy life would look like and, over time, fill out its content until we achieve a more developed understanding of what it means to live happily. Thus, we see that the incompleteness objection is not so great a problem for DVE after all.

[66] Julia Annas, "Being Virtuous and Doing the Right Thing," *Proceedings and Addresses of the American Philosophical Association* 78, no. 2 (2004): 73-74.

4 CONCLUSION

My aim in this thesis has been to contribute to the current debate about what is and is not wrong with self-effacing ethical theories by engaging with some of the most important concerns in this debate. In chapter one, I resisted three objections to self-effacing ethical theories. I showed that these objections have no traction against two prominent kinds of self-effacing ethical theories: a radical version of sophisticated consequentialism (R-SC) and developmental virtue ethics (DVE). I also showed, contra the worries of Keller, Martinez, and Pettigrove, that self-effacing theories, such as R-SC and DVE, are able to tell us what should motivate us and how we should live.

In chapter two, I raised two novel objections to R-SC and DVE. I called the first objection the "self-erasure objection." I argued that the self-erasure objection shows R-SC to be a comprehensible yet counterintuitive theory: while we understand why R-SC would recommend self-erasure, we nevertheless find this recommendation improper and unnerving. I also argued that DVE avoids the self-erasure objection. Far from being a self-erasing theory, DVE is a self-developing theory. In other words, while DVE urges one to develop one's initial commitment to virtue, it urges this in the same way that a gardener urges the seeds she has sown to develop into trees. In this way, the content of one's commitment to virtue becomes broader and deeper as one becomes more virtuous, but its form remains the same over the course of one's ethical development.

I called the second objection the "incompleteness objection." This objection reflected a concern that DVE does not equip agents with a complete account of either virtue or eudaimonia. I argued that, while DVE seems ill-equipped to avoid this objection outright, DVE nevertheless can deflate its force: while DVE does not *immediately* provide agents with a complete

understanding of virtue or eudaimonia, it does provide them with a roadmap for developing such an understanding *in the long run*.

Over the course of this thesis, I have shown that self-effacing ethical theories are more viable ethical theories than they appear at first glance. However, I have also shown that R-SC and DVE are imperfect theories. R-SC is vulnerable to the self-erasure objection, and DVE can deflate, but not completely avoid, the incompleteness objection. Although it is unable to completely avoid the incompleteness objection, I think that DVE is a more intuitive theory than R-SC. While the guidance DVE offers us may seem, when compared to that of R-SC, to be less precise because less complete, DVE's guidance nevertheless comes across to us as more palatable. For unlike R-SC, DVE does not counsel self-erasure, does not tell us to relinquish commitments fundamental to our conception of self. Instead, what DVE counsels is self-development: it tells us to develop a commitment to virtue with which we already identify ourselves and to become better, more virtuous people thereby.

REFERENCES

Annas, Julia. "Being Virtuous and Doing the Right Thing." *Proceedings and Addresses of the American Philosophical Association* 78, no. 2 (2004): 61-75.

Annas, Julia. *Intelligent Virtue*. Oxford: Oxford University Press, 2011.

Annas, Julia. "Virtue ethics and the charge of egoism." In *Morality and Self Interest*, ed. Paul Bloomfield, 205-223. Oxford: Oxford University Press, 2008.

Badhwar, Neera. "Why It Is Wrong to be Always Guided by the Best: Consequentialism and Friendship." *Ethics* 101, no. 3 (April, 1991): 483-504.

Barry, Paul. "Schizophrenia and the Virtues of Self-Effacement." *Les Ateliers de l'éthique* 11, no. 1 (December 1, 2016): 29–48.

Blasi, Augusto. "Moral Cognition and Moral Action: A Theoretical Perspective." *Developmental Review* 3 (1983): 178-210.

Bloom, Paul & Christina Starmans,"Nothing Personal: What Psychologists Get Wrong about Identity." *Trends in Cognitive Sciences* 22, no. 7 (July 2018): 566-568.

Clark, Justin C. "Eudaimonistic Virtue Ethics and Self-Effacement." *Journal of Value Inquiry* 50, no. 3 (September 1, 2016): 507–24.

Copp, David & Sobel, David. "Morality and virtue: An assessment of some recent work in virtue ethics." *Ethics* 114, no. 3 (2004): 514-554.

Das, Ramon. "Virtue Ethics and Right Action." *Australasian Journal of Philosophy* 81, no. 3 (September 1, 2003): 324–39.

Keller, Simon. "Virtue Ethics is Self-Effacing." *Australasian Journal of Philosophy* 85, no. 2 (2007): 221-231.

Martinez, Joel A. "Is Virtue Ethics Self-Effacing?" *Australasian Journal of Philosophy* 89, no. 2

(June 2011): 277–88.

Oyserman, Daphna. “Self-concept and identity.” In *The Blackwell Handbook of Social Psychology*, edited by A. Tesser and N. Schwarz, 499-517. Malden: Blackwell, 2001.

Parfit, Derik. *Reasons and Persons*. New York: Oxford University Press: 1984.

Pettigrove, Glen. “Is Virtue Ethics Self-Effacing?” *Journal of Ethics* 15 (2011): 191-207.

Pincoffs, Edmund. “Quandary Ethics.” *Mind* 80, no. 320 (October 1971): 552-571.

Prinz , Jesse J. & Shaun Nichols. “Diachronic Identity and the Moral Self.” In *The Routledge Handbook of Philosophy of the Social Mind*, edited by Julian Kiverstein, 449-464. Abingdon: Routledge, 2016.

Railton, Peter. “Alienation, Consequentialism, and the Demands of Morality.” *Philosophy & Public Affairs* 13, no. 2 (Spring, 1984): 134-171.

Smart, J. J. C., and Bernard Williams. *Utilitarianism : For and Against*. Cambridge: Cambridge University Press, 1973.

Stocker, Michael. “How Emotions Reveal Value and Help Cure the Schizophrenia of Modern Ethical Theories.” In *How Should One Live: Essays on the Virtues*, edited by R. Crisp, 173-190. Oxford: Oxford University Press, 1998.

Stocker, Michael. “The Schizophrenia of Modern Ethical Theories.” *The Journal of Philosophy* 73, no, 14 (1976): 453-466.

Strohminger, Nina. & Shaun Nichols. “The essential moral self.” *Cognition* 131 (2014): 159-171.

Timmons, Mark. *Moral Theory : An Introduction*. Lanham: Rowman & Littlefield Publishers, 2012.

Wilcox , William H. “Egoists, Consequentialists, and Their Friends.” *Philosophy & Public*

Affairs 16, no. 1 (Winter 1987): 73-84.

Woodcock, Scott. “Moral Schizophrenia and the Paradox of Friendship.” *Utilitas: A Journal of Utilitarian Studies* 22, no. 1 (March 1, 2010): 1–25.