UCLA
UCLA Electronic Theses and Dissertations

Title
The Third Motive On Non-Hedonic, Self-Regarding Reasons for Acting

Permalink
https://escholarship.org/uc/item/4m53g346

Author
Teti, Trent Ross

Publication Date
2021

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Los Angeles

The Third Motive
On Non-Hedonic, Self-Regarding Reasons for Acting

A dissertation submitted

in partial satisfaction of the requirements for

the Degree of Doctor of Philosophy in Philosophy

by

Trent Ross Teti

2021

© Copyright by

Trent Ross Teti

2021

ABSTRACT OF THE DISSERTATION

The Third Motive
On Non-Hedonic, Self-Regarding Reasons for Acting

by

Trent Ross Teti

Doctor of Philosophy in Philosophy

University of California, Los Angeles, 2021

Professor Barbara Herman, Chair

Moral philosophy frequently treats self and other-regarding reasons as opposed poles in our practical deliberations, where the former are presumed to be motivated by hedonic ends and the latter by morality. I believe this mischaracterizes our deliberative process by neglecting a third motive and portrays us as baser creatures that we actually are. In what follows, I offer another account that I think better explains our life-structuring decisions and sheds light on an otherwise neglected aspect of our character.

I begin by considering a body of hedonic literature that contends that people's happiness is tied to internal set points and that the range and extent of deviations from these points are determined by a variety of internal mechanisms that attenuate our joys

and sorrows. As a result, our relative levels of happiness fail to correlate with the goodness or badness of events in our lives, as instead we find ourselves on what has been termed "the hedonic treadmill."

This raises a question about what rationally motivates many of our self-regarding decisions, if not happiness. I suggest that our awareness of ourselves as objects we care about generates a class of non-hedonic, self-regarding reasons, what I call "self-making reasons," and that these guide us in life-structuring decisions. I argue that a person's interest in defining herself in light of her existential circumstance gives her a motive to act in a particular way and develop certain capacities, even if neither will increase her happiness. I try to show how a person's interest in herself as a thing in the world, and the self-making reasons that follow, form the basis of her character and are appropriately subject to judgment, both from herself and others.

In closing, I argue that self-making reasons are normatively interesting despite exceeding the bounds of traditional, other-regarding morality and they give us reason to consider a more expansive conception of ethics beyond what is required and proscribed.

The dissertation of Trent Teti is approved.

Seana Shiffrin

Gavin Lawrence

Calvin G. Normore

Barbara Herman, Committee Chair

University of California, Los Angeles

2021

For my father.
You are still with me.
I love you always.

Many thanks to Barbara
Without you, this would not be.

Table of Contents

Introduction 1

1. On Happiness, Old and New 17

2. Utilitarianism: Happiness, the Experience of Goodness 47

3. Kant's Ethics: Moral Law, Disclosed by Rationality 96

4. What We Want for Ourselves 117

5. A Non-Hedonic Self-Regarding Motive 146

6. Capacities: Fodder For Character 173

7. Supererogation and Character 209

Postscript 241

References 246

Trent Teti

EDUCATION

U.C. Berkeley	BA, Philosophy 1996
University of Virginia	Philosophy Department, Ph.D program, 1997-1999
U.C.L.A	Philosophy Department Ph.D program 1999-2004

ACADEMIC WORK EXPERIENCE

Graduate Student Instructor

U.C.L.A. Philosophy Department — Fall 2000 – 2003

Lectured and evaluated assignments for various philosophy courses Including Ethics, Applied Ethics, Philosophy of Mind, Skepticism and Rationality.

Instructor

University of Virginia Philosophy Department — Summer 1998

Designed curriculum and taught a full length course in Applied Ethics

Teaching Assistant:

University of Virginia Philosophy Department — Fall 1997 – Spring 1999

Lectured and evaluated assignments for various philosophy courses including Applied Ethics, Philosophy of Law, Political Philosophy, and Formal Logic

HONORS AND AWARDS

UCLA University Award, 1999 & 2000.

Dupont Fellowship of the University of Virginia, 1996 - 1998

Presidential Fellow of the University of Virginia, 1996 - 1998

Benjamin C. Moomaw Fellowship, U. Va 1998

UV.a Philosophy Department Most Outstanding Graduate Instructor Award, 1999

Seven Society's Superb Instructor Fellowship Finalist, 1999

Thomas Jefferson Literary and Debate Society "Greatest Orator at UV.a", 1998

Western United States Collegiate Debating Champion (W.S.P.D.A.), 1993

Introduction

Morality and happiness appear to be different animals. Morality, after all, concerns the proper treatment of others and is essentially prescriptive. Happiness refers to a pleasant subjective state and is thus descriptive. Morality is generally thought to pertain to interpersonal relations, whereas happiness is intrapersonal. Morality is experienced as a duty whereas happiness is a state we desire. Nevertheless, we believe their orbits intersect.

Among moral theories, the relation of happiness and morality admits of some variation. If we take happiness to be a fundamental human entitlement, morality might forbid us from diminishing the happiness of others. If we believe happiness is an end to be maximized, we might recognize a duty to promote the happiness of others. A common theme is that the happiness of other persons is morally relevant. But an equally common theme is that morality should allow for one's *own* happiness.

Moral theories tend to accept that however costly morality is, it should not vanquish the possibility of one's own happiness. Accordingly, both expansive and restrictive concepts of morality almost inevitably try to accommodate a subject's happiness. This can be difficult because a basic tension lies between our ordinary notion of happiness and our traditional conceptions of morality. Happiness is seen as our fundamental self-regarding desire, the thing we want for most for ourselves. Morality, on the other hand, is essentially other-regarding, the system of obligations we owe to others. They form the opposed ends of a spectrum running from desire to obligation, from freedom to

responsibility, from selfishness to selflessness. A person's moral fitness and the rigor of a moral theory are often measured by where they each fall on this continuum. The hedonist eschews moral obligation for pleasure, where the saint does the opposite. It's often believed that some versions of utilitarianism demand everything from us, where Kant's ethics leave us freer. So ingrained is this tension that we judge both individuals and moral systems by their response to it.

Various philosophers, Mill and Aristotle among them, have sought to overcome this tension by showing that a subject's happiness and moral fitness can be consistent and even mutually dependent. Generally, though, moral philosophers accept the tension and try to show how a person should balance the demands of morality and happiness in a life. Contemporary philosophers such as Bernard Williams and Samuel Scheffler, for example, have tried to establish a protected space where our happiness and personal projects are safe from what might otherwise be boundless moral obligation.

What makes this tension significant in the landscape of moral philosophy is that happiness and moral obligation are not just seen as two of the many interests that motivate our action, but rather are frequently treated as the respective ends for *all* of our instrumental reason. According to this view, our actions are motivated by two broad categories of reasons; self and other-regarding. Our ultimate self-regarding reason is *to promote our own happiness*; we believe the things we do for ourselves will, in one way or another, make us happier. Our ultimate other-regarding reason is articulated by the demands of morality, and the things we do for its sake are motivated by *our concern for the good of other persons*. Major decisions in our lives (e.g. which profession to enter,

who to marry) and minor decisions (e.g. how to spend a free evening, what to purchase with a bit of extra income) are commonly seen as being structured by these two competing concerns. We chastise ourselves for working toward hedonic ends, knowing that moral concerns should be weighed as well. We'd be tempted to use our disposable resources for ourselves, but we realize obligations to others should sometimes supersede such wants. We resist the impulse to spend all of our extra income on a shiny new car because we know some of it should be directed toward relieving famine in the developing world or helping a neighbor who fell on hard times. We are, according to this view, creatures striving for happiness, but restrained by a guiding notion of what we owe to those around us.

In what follows, I try to resist this picture and offer another. I think conceiving of our deliberations as running along a continuum from self-regarding happiness to other-regarding obligation misconstrues our basic motives and yields a distorted picture of our agency. Because this picture is so deeply rooted in our thought, I try to create space for an alternative by marshaling support from an uncharacteristic source: contemporary psychological research. A bevy of recent studies indicate that our happiness doesn't function in the way we've traditionally thought. Research indicates that satisfying, or alternatively aggravating, our desires, does not lead to a sustained increase or decrease in happiness. We rather experience brief positive or negative deviations, but return quickly to an internal set point that varies from person to person. One person might have a relatively melancholic nature to which she returns even after great successes, while someone else might have a more chipper disposition to which she floats back even after relatively more failure. The range of deviation from one's set

point also seems to vary among people; one person might be largely unflappable where another's mood might fluctuate widely. This body of literature suggests that each person's set point and hedonic range remain largely fixed during their entire lives. A person's happiness simply moves between these upper and lower limits, almost always drifting back to a set point. A variety of internal psychological processes appear to buffer our reactions to both good and bad events in our lives to promote this equilibrium. As a result, neither our misery nor our bliss last very long and they don't correlate with the objective goodness or badness of events in our lives. Our hedonic pattern is thus largely the same whether our lives go well or badly.

If true, this would create a difficulty for our generally accepted model of rational deliberation because (1) if our self-regarding actions are motivated by a desire to increase our happiness and (2) if we can't make ourselves meaningfully happier, then some of our most pivotal decisions are irrational. Contemporary hedonic research suggests that the second claim is true. Though perhaps counterintuitive, the notion is that one would *not* be noticeably happier living in the city as opposed to the country, working as an artist instead of a banker, or marrying Albert over Jonathan *and* neither would one be much happier making the opposite choice. These life-structuring decisions appear to have relatively little effect on one's long-term happiness. So, if one were choosing whichever would make maximize her happiness, she'd be choosing irrationally. Much of the initial reaction to these findings has emphasized changing our behavior so that we're able to maximize whatever happiness is available to us, thus, reducing the extent to which we're acting irrationally.

I think, however, this is the wrong response. The notion that this well-ingrained picture of rational deliberation might entail that much of our behavior is irrational invites us to consider whether the picture itself is inaccurate. Put another way, if we believed that a great deal of our behavior was fueled by a certain motive that we learned was unattainable, we might wonder whether that was motivating our action in the first instance. If it wasn't, we might might not have to choose between the conclusions that we're fated to be irrational actors or that we should change the way we act to preserve whatever modest increases in happiness are possible for us. In such a case, it might be that we were acting well all the while and that the mistake was thinking that our deliberative process primarily involved balancing our desire to increase our own happiness with a need to respect our obligations to others.

I believe this is the case. I think we need *not* have been irrational in our self-regarding choices because I *don't* think we were moved to increase our happiness in many of these actions. I rather think we have an entirely separate, self-regarding motive to mold ourselves as objects in the world and this is what drives our life-structuring decisions. I think this non-hedonic, self-regarding motive constitutes a third motivational pole in our deliberative framework, and that it renders intelligible behavior that otherwise would appear irrational. But more profoundly, I think it shows that we're self-creating agents, defined by the ends we've chosen. Moreover, even if some of these hedonic findings overstate the extent to which our happiness is disconnected from the good in our lives, I believe this third fundamental motive for acting nevertheless has purchase and better explains our life-structuring decisions.

The fully articulated version of that view is admittedly a long way off and this account proceeds in stages. In the first chapter, I unpack the everyday notion of happiness presupposed by our mainstream moral theories and contrast it with the starkly different picture drawn by modern hedonic research. Here, I try to show that whereas our ordinary notion holds that happiness is tied to states of affairs in the world and increases without obvious limit as circumstances improve for us, contemporary research suggests that our happiness is tied to an internal set point in us and rarely deviates from a fairly narrow range. Whereas the ordinary notion suggests that happiness increases linearly and, given equivalent circumstances, that it's more or less equally available to all, this research claims that its pattern is cyclical and both the range and set point vary among persons. Accordingly, while our ordinary notion supports the idea that happiness is a good worth promoting, it's far less clear that the happiness illustrated by this research has an unambiguous moral relevance.

This prompts one to wonder what role happiness should play, in either moral theory or in our self-regarding deliberations. To provide a foundation for this question, I explore why moral theories are expected to account for happiness in the first place. I try to explain why, for both constructivist and realist conceptions of ethics, underwriting the intuition that morality should allow for our own happiness is a criterion of acceptance for moral theories generally. Setting the stage for the second and third chapters, I discuss, in broad terms, why utilitarianism is often seen as more successful than Kant's ethics on this score.

The second chapter explores the question of whether classical utilitarianism does indeed find a better place for happiness. In treating happiness as the highest human good and positing that our happiness should ultimately coincide with that of others, utilitarianism dissolves the opposition between self-regarding desires and our other-regarding obligations. In this way, it presents an attractive deliberative picture. But when we delve into the place happiness holds in each of utilitarianism's two foundational principles, I think we find that it strains each. It isn't entirely obvious, as the first principle holds, that satisfying our informed preferences will increase happiness or, as the second principle holds, that happiness is our highest good. Because these are familiar critiques, I explore what I take to be the strongest utilitarian defenses of each principle to see what sort of response can be marshaled for the doctrine. I argue that these defenses independently fail, and that both of utilitarianism's two principles are properly vulnerable to criticism.

For my purposes, the important lesson here doesn't pertain to utilitarianism *per se*. It's rather that the two principles share a middle term, happiness, and my goal in this discussion is to illustrate why this shared concept exposes both principles to criticism. The reasons why satisfying our informed preferences won't reliably lead to increased *happiness* and why *happiness* isn't our highest good are related. It's that the happiness we experience is fickle, internally regulated, and highly variable among persons. It is, thus, unlikely to consistently follow from a pattern of behavior across persons and it's not the sort of thing that would be expected to connect with an independent moral concept such as goodness. On this latter point, I use Mill's status as a non-utilitarian to show the complexities that arise when trying to connect happiness to richly normative

notions of human goodness. The general theme of chapter two is that happiness maximization doesn't sufficiently explain the rationality of our self-regarding, life-structuring decisions.

I intend this to prepare the ground for chapter three, where I inquire into the alleged austerity of Kant's Ethics. As I explore some features of Kant's ethics generally and his treatment of happiness in some detail, I try to explain why they can appear stern. Kant is uncompromising in his view that we should embrace our highest nature, which he locates in our rationality, and resist our inclination for happiness, which he sees as a vestige of our animal nature. He admits because our impulse toward happiness is unrelenting throughout our lives, we can expect a tension will persist in our deliberations. The greatest harmony we can experience, for Kant, is not a unity between our desire for happiness and moral law, but rather a strength of character that allows us to overcome our persistent inclinations.

At first glance, this does appear to be a demanding view. But my hope is that the first two chapters give us the context to see Kant's ethics as something else. If one accepts the notion that our happiness is as disconnected from the world as this psychological data suggests and thus locating central place for happiness in a moral theory causes the persistent difficulties that even strongest defenders of utilitarianism seem unable to resolve, then Kant's relegation of happiness begins to look less like a deficiency and more like a virtue. My general hope is that by chapter three we become more sympathetic to Kant's attempt to push against the notion that our desire for happiness should guide our deliberation and action.

This, however, leaves an explanatory gap as to what motivates our life-structuring decisions, from the self-regarding standpoint. To inform this question, I explore Kant's notion of imperfect duties; the duty for self-perfection and for the happiness of others. These duties are a notable departure from the uncharitable caricature of Kant's ethics as simply consisting of prohibitions on action, as they offer an incipient glimpse into a realm of positive obligations and a broader conception of moral guidance. I suggest that while these are closer to explaining our life-structuring behavior than happiness maximization, unclarity about the scope and rigidity of imperfect duties motivate us to look beyond for an explanation.

If the goal of the first three chapters is open a space where we can see beyond happiness, all that follows is devoted to explaining what could fill this space. In chapter four, I consider three general concepts of well-being, subjective, objective, and hybrid theories, to see which, if any, could motivate our life-structuring decisions. I begin by briefly describing subjective well-being views, (the description of these is shorter than the others because my extended discussion of utilitarianism in chapter two already established the basic parameters of this type of theory). I argue that subjective well-being views lack the resources to adequately sanitize the preferences they promote and are vulnerable to happiness machine worries.

I then describe objective-list theories of well-being and try to show that while these views are responsive to the concerns attendant to subjective views, they are open to motivational challenges that compromise their explanatory value and that these challenges are difficult to overcome even when modifications to the theories are made.

I go on to consider hybrid well-being views acknowledging that, *prima facie*, they appear to resolve the difficulties of the prior two theory types. Despite that, I try to show that such views place excessive demands on the emotional constitution of agents, especially those raised in non-ideal circumstances. For this reason, while such theories provide appealing explanations of behavior, I argue that these explanations better fit creatures more psychologically unified than ourselves. I, thus, conclude that none of the motives posited by these theories of well-being are suited to explaining our life-structuring decisions.

In closing this chapter, I outline an alternative that I think better explains our life-structuring (and other) decisions. I suggest that our awareness of ourselves as self-creating objects in the world generates a class of self-regarding, non-hedonic reasons that can inform and even guide us in these deliberations. Very generally, the notion is that a person is not merely a collection affective states, but is also a special sort of object in the world. This object is constituted not only by its physical extension but, more profoundly, by a complex web of action-motivating reasons and commitments that define the sort of thing it is. While some of these defining reasons are hedonic in nature and others are moral, I suggest that this object is also defined by a class of self-regarding and non-hedonic considerations that I call *self-making reasons*.

In chapter five, I develop the notion of non-hedonic, self-regarding reasons in some detail. To do this, I consider why a person might pursue a voluntary end with enthusiasm and persistence when she has no expectation that attaining this end will increase her happiness. A reason, I offer, is that she have been moved by a non-

hedonic motive in the first place. Many life-structuring decisions don't generate more happiness in part because they bring more stress than alternatives and, more generally, because our psychology is unlikely to allow for a sustained change in happiness. What such decisions do profoundly affect is the object we are in the process of becoming. I argue that we pursue these ends because we have self-regarding motives that are *not* hedonic in nature and a species of these are self-making reasons.

I subsequently explain how one's interest in oneself as an object generates a reason to act in a *particular* manner. The part of oneself subject to judgment I call one's *character*. As I use the notion, a person's character is determined by how one chooses to respond to the circumstances in which she finds herself given the resources open to her. While a variety of decisions affect one's character, life-structuring decisions almost inevitably do because they are subject to extensive deliberation and they express a hierarchy of value. One may, along these lines, decide to become a parent or enter a challenging profession, not because of any expected happiness or perceived obligation to others, but because she feels that this use of her capacities is appropriately responsive to her circumstance. One might, I argue, be driven to become a certain thing, not merely to feel a certain way.

I close chapter five by laying out the general structure of self-making reasons. I explain that these reasons are not reducible to more basic hedonistic motives because becoming the object one aspires to be need not, and typically will not, result in substantially increased happiness. I argue that self-making reasons present a more attractive picture of our agency than the traditional hedonistic model because they allow

us to avoid the conclusion that much of our life-structuring behavior is irrational without having to resort to implausible explanations of our action. Lastly, I suggest that what should deeply count in their favor is the extent to which self-making reasons show that we're more interesting creatures, leading more interesting lives than the traditional model would allow. We are not, according this this view, simply chasing different varieties of happiness. We are rather more centrally concerned with the kind of thing we are and can be. Our lives are not spent simply in the pursuit of pleasure, but in pursuit of an evolving concept we have of ourselves.

Having laid out self-making reasons' general structure, I move on in chapter six to discuss how acting on reasons defines a person's character and, more specifically, which reasons should govern our life-structuring choices. To address this admittedly large question, I consider a species of decision that does a great deal to determine the trajectory of our lives: the choice of which capacities to develop in ourselves. I argue that when we develop our capacities we are, in a sense, making ourselves because the development of our capacities (1) recasts the question to which character is an answer and (2) generates further associated values. As such, our reasons for developing the capacities we choose *are* the reasons we become what we ultimately will be.

I start by considering two commonly adopted motives for developing a certain capacity or talent, namely that one's talent is either rare or uncommonly strong or, alternatively, that the development of this talent is likely to increase one's happiness. After exploring these motives, I conclude that both underdetermine our choice in different ways. A person might reasonably decide to forgo the development of a rare

talent because it would serve an ugly or trivial end, or, alternatively, develop a lesser talent toward a more noble end. In either case, the relative strength or rarity of the talent isn't pivotal to her choice. I also argue a person shouldn't develop a talent on the assumption it will make her happier than an alternative because a myriad of factors will influence her happiness more than this choice. I think the conclusion that we don't develop particular capacities because they're rare or likely to make us happy, gestures toward our actual motive; that at least in an inchoate way we understand that the capacities we choose to develop, and our reasons for choosing them, make us the thing we are and that we're ultimately guided by a concern for this object.

I proceed to discuss how our choices and the reasons for them determine our character, as I use the term. I suggest that for choices to have character significance at all, they must be made in the context of some limited constraint. I'll argue that both excessive abundance and scarcity eliminate the mixture of choice and constraint required for our decisions to express relative value. The principle that should govern our decisions about which capacities to develop instead emerges when we consider circumstances of *relative* abundance and scarcity. I try to show that such circumstances indicate that we should be guided by a sense of what we believe is (1) valuable in the world and (2) responsive to our circumstance. The general notion is that these capacities and the actions they generate become part of the world, and thus the reasons governing their development are, in part, the reasons the world is what it is. I argue that our choice to become a certain kind of thing should be delimited by a sense of what we want to write upon the structure of the world as we find it. I'll suggest that in the end, we are little more than this decision, extended in time.

In the seventh and final chapter, I consider two connected issues that remain after the preceding discussion. The first issue is the relation of self-making reasons, and character itself, to moral theory. While I've argued that self-making reasons are self-regarding, I've also claimed that we *should* be moved by them and that they are normatively significant. Given such language, one might reasonably wonder if self-making reasons are moral in nature. In this chapter, I argue the answer has more to do with what we take the proper scope of moral theory to be than with the importance of objective reasons and character.

The second issue I discuss concerns the *range* of self-making reasons a person might have. In prior chapters I focused primarily on self-making reasons that were life-structuring in nature because they help demonstrate the existence this class of reason generally. But these are just one sort of such reasons when, in fact, many self-making reasons are *not* life-structuring. Rather than simply enumerating other kinds, I explore various character properties from which different sorts of self-making reasons would flow.

These two issues are related because the range of different character properties informs the relation of character and morality. I engage these issues by examining the concept of supererogation presented in J.O. Urmson's "Saints and Heroes." Urmson considers supererogatory actions to be extra-moral and gestures toward the notion that they are just one instance of normatively interesting phenomena that don't fit into our existing moral taxonomy. I believe Urmson is right in this thought and, following his model, I identify four adjacent classes of normatively interesting behavior that don't fit

into our current moral framework: the *extraerogatory*, *suberogatory*, *undererogatory*, and the *nonerogatory*.

I try to show that in these classes of action we can a find plethora of normatively interesting behavior, some inspiring and some disappointing, but none of which is either morally required or proscribed. As such, I believe this behavior issues not from that which a person *owes* another, but from a concern for her character. I discuss this normatively interesting, extra-moral behavior for two reasons. First, it enables us to see the aspects of character that give rise to self-making reasons that are less world-affecting. Second, considering this range of behavior gives us greater insight into how character and morality might be related. I conclude that whether character is considered moral or extra-moral, these classes of behavior indicate fertile areas for normative inquiry beyond the permissible and impermissible.

What follows is not an incontrovertible, deductive argument, but something closer to an offer. It recognizes the grip a certain well-established picture has on us and invites one to consider a different view. It calls into question certain widely accepted facts about our experience of happiness and traces the implications through our notions of what a good person and life can be. With such a large target, it cannot aspire to be the final statement on this subject. I rather hope it can provoke us to think about the motives that ultimately drive us and the kind creatures we take ourselves to be.

Before even starting, I must admit that at least superficially, the traditional hedonic picture is more comforting. That notwithstanding, I'll try to show that for all its ease, it's

not suited for the kind of beings we are. I'll argue that it's predicated on sanitized concepts of rational deliberation and that it incorrectly presupposes that our level of satisfaction can serve as a reliable metric of our lives' success. But more profoundly, I'll suggest that the traditional model paints us as somewhat vulgar creatures. Our lives, according that view, are spent in pursuit of happiness, even if we are constrained by a notion of the good. It holds that morality is something we owe to others, but that, for ourselves, we simply seek different modalities of pleasure. It's a picture that focuses on how our lives *feel*, but it neglects a rich account of what we *are* independent of this feeling. It's missing a notion of a person as an evolving object, reaching for something greater than happiness because it doesn't have a place for the enduring dissonance we feel when our best decisions aren't met with corresponding pleasures. Most importantly, it doesn't express what might be the most redeeming part of our humanity: that we're ultimately driven not by a desire to feel, but by a desire to be.

Chapter One
On Happiness, Old and New

> *If we were to ask the question, "What is human life's chief concern?", one of the answers we would receive would be, "It is happiness." How to gain, how to keep, how to recover happiness, for most men, is the secret of all they do.*
>
> William James' *Variety of Religious Experience*, pg. 83

It would be pleasant to think that the moral life must be happy and that the happy life must be moral, but we know this isn't so. We can imagine both kind, melancholic persons and those who are happy but malevolent. As an empirical matter, morality and happiness don't appear to be inexorably intertwined.

One might instead insist that at least the moral life *should* be happy. Even if morality and happiness don't always proceed together, one might hold that happiness is a central goal of any rational person's life and that a reasonable chance at happiness is a basic right. If so, it would seem that an acceptable moral theory should make space for happiness. We tend to think, in the normal case, that morality should not so obligate a person that her own happiness is made impossible. But to understand how this might be so, we need an account of how morality and happiness fit together. We hope they're symbiotic and fear they're antagonistic, but in between lie a myriad of potential relations– that they're partially coextensive, contingently correlated, or perhaps entirely unrelated. To see how they could fit together, we must understand the kind of things they are.

People experience morality as special kind of obligation, one that should take precedence over all others. Though we have many commitments (social, familial, and

professional), in a conflict our moral obligations are expected to be overriding.[1,2] Moral obligation also presents itself as an *external* authority, something imposing on us from the outside, to which we should yield irrespective of our inclination. Initially, we experience morality as a foreign guide to be obeyed and later an internalized barometer to which we should listen, but in both cases we experience it as something prior to ourselves, as part of the fabric of the world we find ourselves in.[3]

Happiness, on the other hand, isn't experienced as either external or prior to ourselves. Happiness is experienced as our own, arising from and tailored uniquely to ourselves. We needn't surrender our other inclinations to happiness or learn to adopt it as an internal motivation. Happiness is rather taken to be the source from which our inclinations derive, the foundation of all our self-regarding motives. Our disposition to happiness, thus, is an *impulse toward,* not an *obligation to*.

---

[1] Some social obligations have moral entailment (e.g. it would be both a social *faux pas* and a moral transgression to strike a friend at a dinner party) and some instances of professional obligations appear to be transparently moral obligations (e.g. an accountant's professional responsibility not to defraud his clients might be little more than a particular application of the moral proscription against stealing). Nevertheless, there are clearly social, familial, and professional obligations that are *not* obviously moral obligations (i.e. the obligation to send birthday cards to friends, visit family on holidays, and a stage actor's professional obligation to do character research), and in a conflict a moral claim is generally accepted to take precedence over such obligations. The limiting case would likely arise in a circumstance where an agent had a very strong professional obligation and a discretionary moral obligation. If, for example, Kant's imperfect duties are properly interpreted as being defeasible, a professional obligation might take precedence (i.e. a doctor's obligation to not harm a patient might override his defeasible obligation for charity).

[2] Although some moralists such as Phillippa Foot hold that, like other sorts of obligations, moral obligations are simply hypothetical imperatives, one might argue that moral obligations must still take precedence. For even if all obligations are hypothetical, it doesn't follow that all are equally contingent, as certain antecedent conditions might be lexically prior to others. See Philippa Foot, "Morality as a System of Hypothetical Imperatives,", *Philosophical Review*, Vol. 81, No. 3 (Jul., 1972), pp. 305-316.

[3]. I intend the claim that morality is *part of the fabric of the world we find ourselves in* to allow for both the possibility that it is a socially constructed part of the community we're born into and the stronger claim that morality exists prior to social institutions and rather follows from our fundamental nature. The reason for this vagueness is that I want to distinguish our experience of morality, irrespective of its ontology, from our experience of happiness. For both the socially constructed and realist conception of morality, it is experienced as an external obligation that can grow into an internalized one.

It has been traditionally thought that happiness is sought for its own sake by all rational subjects. Pascal famously wrote: “All men seek happiness. This is without exception. Whatever means they use, everyone tends toward this end. It is the motive of every human being, even those who hang themselves.”[4] Whereas we might at times feel plagued by morality or conveniently forget our obligations, happiness never plagues us, nor can we easily forget about its demands. Where morality is the most important of mind-independent guides of our behavior, happiness is the mental state we most keenly seek.

Happiness, in common parlance, doesn’t refer to a single positive affective state. Joy, ecstasy, and glee, for example, are each happy affective states. It might be tempting to think that happiness refers to favorable judgment of one’s overall circumstances, but one can imagine a person who believed her circumstances were good overall, but was nevertheless unhappy.[5] Happiness is sometimes associated with a feeling of satisfaction, but we can imagine a happy and yet unsatisfied person. A celebrated line of Aristotelian thought treats happiness as a kind of flourishing, such that that a person who has led a virtuous and fulfilling life would, *ipso facto*, be happy.[6] But again, this diverges from our normal use of the term happiness, since we can imagine a person who lived a fulfilling life, even by her own measure, who was nevertheless unhappy day to day.

---

[4] Pascal, Blaise *Pensees* No. 425

[5] See Sarah Buss’ treatment of happiness in “The Irrationality of Happiness and the Paradox of Despair,” Journal of Philosophy, April 2004, pg. 170

[6] Such a use echoes with certain objective theories of well-being which I talk about at length in chapter five.

Our ordinary notion of the term ‘happy’ most closely corresponds to a subjectively experienced, positive affective state or, as it’s commonly known, *subjective well-being*. An initial requirement of subjective well-being is a subject’s internal endorsement; one has to *feel* happy to be happy.[7] Subjective well-being includes a family of pleasurable affective states, when these states are not overwhelmed by countervailing negative states.[8] This happiness doesn’t require an absence of pain, for example, but it does require that whatever measure of pain a person feels is outweighed by some feeling of pleasure (e.g. the exhausted marathoner). We cannot think of someone as happy if her pain or misery is so great that it eclipses all of her good feeling. When a person going through great turmoil is nevertheless happy, we think this person must have some positive feeling or commitment in which her happiness can take root.

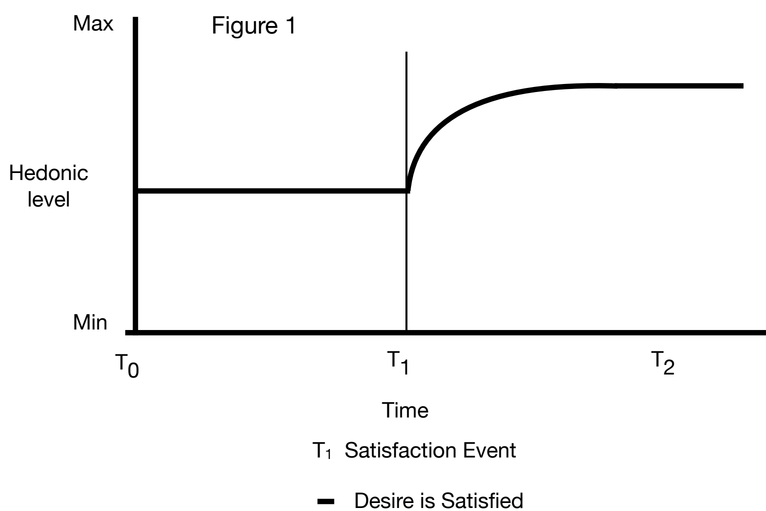
As subjects, we take ourselves to be authorities on what our own happiness requires. In the prototypical case, we believe (1) that objects or states of affairs will make us happy and (2) we pursue them for that reason. We expect their attainment will increase our happiness and their lack of attainment will depreciate our happiness (See Figure 1 and 2)[9]. In unfortunate cases, we can fail to know in what our happiness lies,

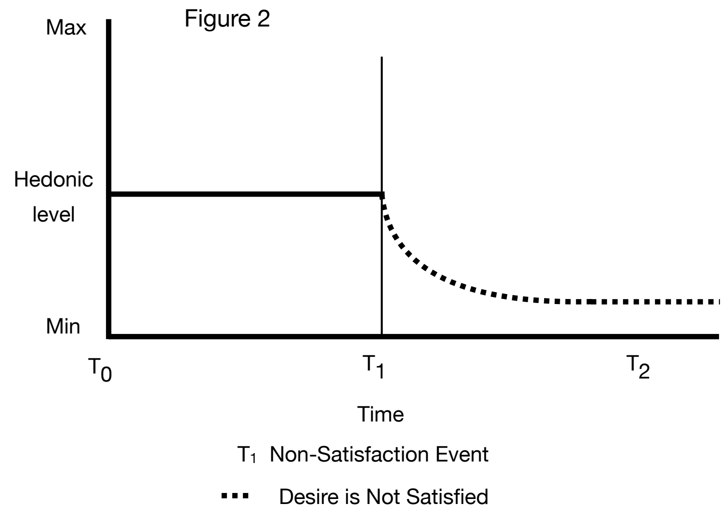
---

[7] Such a requirement doesn’t immediately explain the intelligibility of the notion that *I was happy then, I just didn’t know it*, but upon reflection, these become more compatible. The retrospective assessment of one’s own happiness involves identifying a feeling one had in the past that one failed to recognize as happiness, but which now, perhaps on the basis of new information or greater experience, one sees as happiness. But here there is internal endorsement, simply one that arrives well after the fact. The limiting case against an internal endorsement requirement would be one in which a person was said to be ‘happy’ even though she never believed herself to be so. Such a case would seem incompatible with our everyday use of the term.

[8] Hereafter all mentions of “happiness” will refer to “subjective well-being” unless otherwise noted or attributed (e.g. ‘objective happiness’ or ‘happiness for Aristotle’).

[9] All figures presented here are for illustrative purposes only. Although they are intended to express the hedonic pattern described, they themselves are not draw to scale or derived from clinical results. That said, the hedonic patterns they illustrate detailed in the cited literature from Kahneman, Gilbert, Easterlin, DIener, and others and these findings are supported by numerous clinical studies.





or we can believe something will bring us happiness when it does not. But we see these as failures of self-knowledge or mistakes about circumstances in the world and, in either case, we believe these are unusual. We think we know ourselves well enough to see that the third piece of cake will make us more unhappy than happy and that we won't enjoy the expensive scarf enough to justify its cost. We feel chagrined when we are wrong, but we do not find the sources of our happiness sufficiently mysterious that we believe another party better understands our happiness than we do ourselves.

In our ordinary view, happiness increases linearly and without obvious limit. We believe that the more our informed preferences are satisfied, the happier we will be. Though of course we don't expect happiness to increase infinitely, we don't ordinarily think there is any *particular* limit, such that after a certain point greater happiness is unavailable (or even that our pleasures must diminish in magnitude as we get happier). We rather think of pursuing happiness as stoking a fire, which if properly attended can continue to burn hotter and hotter.

We also think that, if the world didn't intervene, happiness would be available to each of us equally. We realize, of course, that unequal distributions of goods and liberties give some greater prospects of happiness than others, but we tend to believe that antecedent to these distributions we all have a roughly equivalent potential for happiness. We tend to look at the most fortunate among us and think if we had their share, we'd be happier, and, conversely, when we imagine having as little as the poorest people in the world, we think we'd be miserable. With regard to happiness, we're still partial to the enlightenment notion that we're largely blank canvases and if any two of us were painted by the similar circumstances, we'd each feel roughly the same.

I take this to be the ordinary conception of happiness presupposed by our common parlance.[10] When philosophical theories try to accommodate our happiness, they are, broadly speaking, trying to accommodate this notion; one that takes happiness to consist of positive affective states, ascending linearly, without a necessary limit, equally

---

[10] In reading the notion of happiness above, one might be left with the sense that it's deficient, easily assailed, or in some other sense represents a straw-man. Against this view, I'd note that this characterization is charitable relative to the dominate notion of happiness in psychology prior to the introduction of adaptation and set-point theory. As recently as 1967, the widely-accepted treatment of happiness in psychological literature was Warner Wilson's "Correlates of Avowed Happiness," which is far more anachronistic than the characterization put forth above. Here Wilson concludes a happy person is "…young, healthy, well-educated, well-paid, extroverted, optimistic, worry-free, religious, married person with high self-esteem, job morale, modest aspirations, of either sex and of a wide range of intelligence." (pg. 294) Such a notion strikes us immediately as antiquated, since it's now believed that youth, extroversion, education, and pecuniary gain (post-subsistence) are not strongly correlated with *SWB*. The characterization above does not rely on any such associations.

Most importantly, in the next chapter I'll argue that a desideratum of moral theories generally is that they support our intuition that morality should allow a place for one's own happiness. The core of this intuition is that happiness *as we ordinarily think of it* shouldn't be rendered impossible by morality. Moral theories seeking to accommodate this intuition must therefore try to accommodate our ordinary notion of happiness. I take our intuitive, everyday notion of happiness to be most clearly identified by our normal-use which is what I'm trying to capture above. A theory that simply redefined happiness as a wholly objective state of virtue without any subjective entailment, for example, wouldn't support our intuition in the relevant sense (I address this at greater length in chapter four's discussion of objective theories of well-being).

available to all, attained through satisfaction of informed preferences. It is a perfectly reasonable and widely accepted picture. The difficulty, according to contemporary hedonic research, is that very little of this is true of happiness as we experience it.

According to this body of research, the clearest divergence from our ordinary notion is the way in which desire-satisfaction fails to produce happiness in the straightforward fashion we suppose. We do not want something, acquire it, experience increased happiness, want something else, acquire it and experience still greater happiness, as if we're climbing the steps of a hedonic pyramid. The metaphor of a pyramid is apt, but, as it turns out, incomplete. It's more accurate to see the steps as moving and that all of our climbing merely sends us back and forth within a very narrow range. As soon as we acquire the object of our desire, we experience an increase in happiness, but it's relatively short-lived because we cease wanting that thing and instead pin our happiness on some other object or state to be acquired. Hence, the phenomenon happiness researchers have termed "the hedonic treadmill."[11]

Let us imagine, for instance, an agent desires a necklace, believing it will give her a needed accessory for special occasions. Upon purchasing it, she is not disappointed and, indeed, the necklace increases her happiness. To this point, the traditional account of happiness suffices; the acquisition of a desired object increases an agent's happiness. Following this, however, our ordinary account begins to unravel. Shortly

---

[11] Philip Brickman and Donald Campbell coined this phrase in their well-known "Hedonic Relativism and Planning the Good Society" (1971), which appeared in M.H. Apley, ed., *Adaptation Level Theory: A Symposium,* New York: Academic Press, 1971, pp 287-305.
For an overview of the effect of adaptation on subjective well-being see "Subjective Well-Being: Three Decades of Progress," Diener, Ed and Suh, Eunkook and Lucas, Richard and Smith, Heidi, (January 10, 2013). *Psychological Bulletin*, Vol. 125, No. 2, 1999, especially 276-277, 285-295.

after acquiring the necklace, the happiness it originally brought dissipates. Perhaps, its owner reasons, this is because she doesn't wear it as often as she had hoped or because it tends to dominate her outfit when she does. These claims may not be true, but even if they are, they are not the reason for her decreasing satisfaction according to this research. Rather, her own need for rationalization requires that she attribute her waning happiness to something or other, when this depreciation is actually the result of an independent process.[12] As her increased happiness dissipates, she'll determine that she really needs a certain pair of earrings to match this necklace and will begin nurturing a desire for those. In our traditional account, there's no reason her happiness should be so short-lived or that its cessation should give rise to a desire for something else.

Also incompatible with our ordinary notion is an alternate version of events in which our agent returns to buy the necklace only to learn that it has already been purchased by someone else. While initially she might despair, this will quickly give way to a reassessment that she didn't actually fancy the necklace as much as she'd thought. She'll reason that it wasn't actually flattering on her or that it projected an air she'd wanted to avoid. Much more to her liking is a more modest, but still desirable, scarf she recently saw in a shop window. But, like her declining satisfaction after the purchase in the earlier case, this reassessment of the necklace need not be rooted in fact.

Our ordinary notion predicts she would experience *more* than a momentary dip in happiness upon not acquiring the necklace. This is important for desire-satisfaction

---

[12] When her decreasing satisfaction *is* the result of some defect in the object, it is simply seen as a case of learning and is thus unproblematic for the traditional account of happiness.

theories as they are predicated on the notion satisfying desires increases happiness and leaving them unsatisfied decreases happiness.[13] If leaving an agent's desire unsatisfied didn't decrease her happiness, there would be little justification for allocating the resources to satisfy it. In this case, however, while the gain and loss of the necklace are respectively accompanied by increases and decreases in happiness, both are relatively small and short-lived. More significantly, after a fairly brief period the agent's overall level of happiness returns to the same level, whether she acquires the necklace or not. For the traditional model, satisfying and not satisfying the desire would respectively engender increasing and decreasing hedonic curves (Figure 1 and 2). However, for the new model, these actions would be followed by momentary blips, each followed by a return to the original state (See Figure 3 and 4).

Figure 3

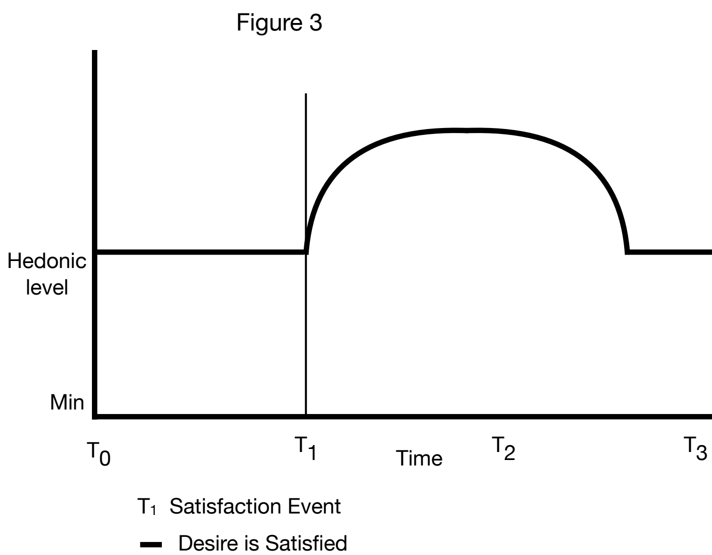
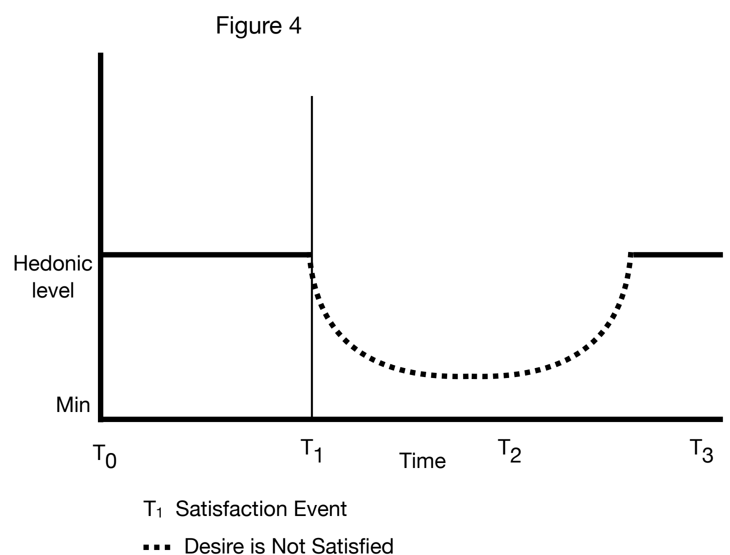


Figure 4



[13] For an explanation of our tendency to overestimate their affective reactions to negative events, see Daniel Gilbert's "Immune Neglect: A Source of Durability Bias in Affective Forecasting" in the *Journal of Personality and Social Psychology*, Vol 75, No. 3, Sep 1998, 617-638.

Moreover, the issue is not that the necklace is simply too small an acquisition to create a sustained increase in happiness. The attractive partner, prestigious job, and large home all are accompanied by the same pattern: desire, acquisition, and a brief increase in happiness followed by a return to the prior hedonic level.[14] What's interesting is that neither the agent's lack of sustained happiness after acquiring the necklace nor her quick recovery when she's deprived of it are explained by a deficiency in the necklace. Rather, this pattern will obtain even when she is correct in thinking that the necklace is of high quality and that she desires it. Her error is in assuming that satisfying her informed desire would increase her level of happiness in any sustained way. This mistake, in turn, follows from her failure to see that she is tethered to a particular hedonic level, where she is likely to spend her entire life apart from brief excursions into happiness and sadness.

The modern explanation of this phenomenon is that preconscious mechanisms adjust our level of happiness both when our desires are satisfied and when they are not. One of these mechanisms, termed *adaptive preference* by Jon Elster in 1983, buffers us from the unhappiness that we'd expect to follow from the non-satisfaction of our desires.[15] Adaptive preferences adjust our aspirations (wants, desires, or preferences) to favor options that are possible for us, when we would favor other options if they were attainable. On the traditional model, our preference changes when we learn some new fact about a desired object, and in such a case its availability or unavailability is largely

---

[14] See Subjective Well-Being: Three Decades of Progress, pgs. 286-294. On long term financial improvement not improving hedonic levels, see also Richard Easterlin's "Explaining Happiness" *Proceedings of the National Academy of Sciences*, Vol. 100, No. 19, pp. 11176-11183, September 2003.

[15] See *Sour Grapes*, Jon Elster, Cambridge, UK: Cambridge University Press, 1983. pp.110.

incidental (e.g. when the agent loses her desire for the necklace because she learns it's a fake, its availability or lack thereof becomes irrelevant). Adaptive preferences differ from preference changes due to learning because they don't require the addition of any further information about the object, except that its availability has changed.[16] Moreover, in the case of adaptation, a change in availability is itself sufficient to reverse a preference, as opposed to being one of a variety of factors that goes into an overall evaluation.[17] This is why the envisaged agent stops wanting the necklace soon after learning it has been purchased and is no longer available. We stop wanting even that which we desire most when we come to believe it is unattainable.

One might reasonably wonder why our preferences non-consciously adapt instead of remaining aggravated by our circumstances. According to Elster and others, the purpose of adaptive preferences is to reduce an agent's cognitive dissonance, or the extent to which non-fitting relations among cognitions cause psychological discomfort for the agent.[18] Cognitive dissonance will typically ensue when an agent has a desire she cannot satisfy.[19] Someone who itches, but is bound and cannot move, would experience dissonance from the unsatisfiable desire to scratch. This person might plausibly try to convince herself the she did not itch as a way of reducing the tension

---

[16] The one bit of information that is necessary for a preference to adapt is the belief that it is unavailable. But in this case the additional information is not so much about the object as about the impossibility of attaining it.

[17] Elster, pg.113

[18] Elster pp. 36, 110, 126. See also Daniel Kahneman's "Objective Happiness" in Well-being: The Foundations of Hedonic Psychology , (pp. 3-25), especially 14-15. Kahneman, Daniel (Ed); Diener, Ed (Ed); Schwarz, Norbert (Ed). (1999). New York, NY, US: Russell Sage Foundation, xii, 593 pp.

[19] Festinger, Leon, "A Theory of Cognitive Dissonance," Evanston, Ill: Row Peterson, 1957. Also see Festinger, L. & Carlsmith, J.M. (1959). "Cognitive Consequences of Forced Compliance" *Journal of Abnormal and Social Psychology, 58*, 203-210.

between her preference and her circumstance. The physical reality of the itch makes this more difficult, but our minds are better able to adjust our wants. Whereas our traditional notion holds that when an agent's desire is at odds with her environment she should naturally change her environment to fit this desire, adaptive preferences change her desire to suit her environment. This phenomena is consistent with research that shows test subjects are more satisfied with a result when they believe they cannot change it than they are with that *same* result when they believe that they can.[20] It is not the substantive character of the result that causes satisfaction, but rather the impossibility of changing it.

More recent research has referred to adaptive preferences as part of a *psychological immune system* (PIS) that is intended to insulate agents from psychological harm.[21] Just as our bodies induce a state of 'shock' to prevent us from feeling physical trauma, this hypothesis suggests our minds induce an analogous state to buffer us from mental traumas. The more severe the trauma, the more powerfully this system works, creating the counterintuitive phenomenon that traumatic injuries often abate faster than milder ones and therefore that one might actually experience *more* suffering from slight harms than severe ones (See Figure 5). Whereas this psychological immune system might enable an agent to recover from the death of a parent with disquieting ease, it isn't triggered by the creaking screen door that annoys her to no end.[22] Our relative level of happiness might thus systematically fail to track the severity of events in our lives. This

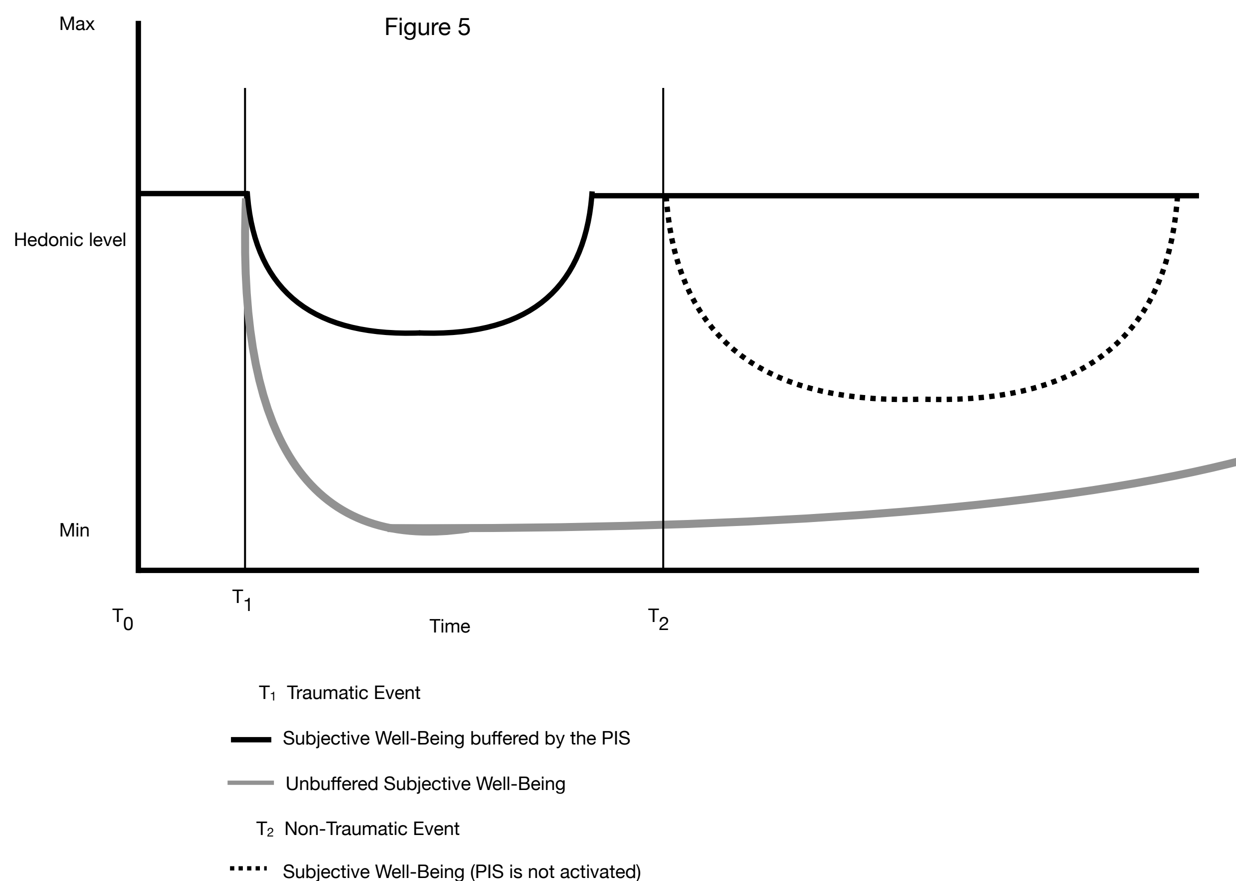
---

[20] See Daniel Gilbert and Jane Ebert's "Decisions and Revisions: Affective Forecasting of Changeable Outcomes," *Journal Personality and Social Psychology*, 2002 Vol. 82, No. 4, 503-514.

[21] Gilbert, Daniel, *Stumbling on Happiness*, Knopf, 2006, (pg. 162).

[22] See "The Peculiar Longevity of Things Not So Bad": Daniel T. Gilbert, Matthew D. Lieberman, Carey K. Morewedge, Timothy D. Wilson, *Psychological Science*, Vol. 15, No. 1. (2004), pp. 14-19.

is, of course, strongly inconsistent with the traditional notion that treats happiness as a proxy for our objective well-being.



Adaptive preferences have a dialectical twin that actually *decreases* our happiness when we attain our ends. Recall how after acquiring a desired necklace the imagined agent experienced a brief increase in happiness, only to return to her prior hedonic level and cease fancying the necklace. According to Elster and a bevy of hedonic research, this loss of favor was the result of a systematic process called *counter-adaptation*.[23] Just as it seems that we want what we do not have, it seems that we typically cease wanting that which we've already obtained.

[23] Elster, Jon pp. 111

Counter-adaptation is in some ways a stranger phenomenon than adaptation, since it artificially *diminishes* an agent's level of happiness. We can understand why our psychology would be tailored to prevent us from falling into despair, but it isn't immediately clear why excessive happiness should be a concern. However, the manner in which counter-adaptive preferences operate offers an answer. When a desire is satisfied and the object of that desire is devalued, a new desire is typically deployed in place of the first. When the agent ceased being content with her necklace, she began to desire a pair of earrings. When those too inevitably lose their luster, she'll develop another desire on which she'll believe her future happiness depends. An effect of this process is that she is kept in a perpetual state of goal pursuit. To the extent that these goals would tend to make life go better (or longer), counter-adapting preferences keep a person forever productive in the pursuit of new ends and in this way confer a survival advantage.

One might still think that adaptive and counter-adaptive preferences are different animals. Adaptive preferences appear psychologically benign because they increase our happiness, whereas counter adaptive preferences seem less so because they decrease our happiness even if this is to improve our productivity. But adaptive preferences do more than simply increase our happiness, for they could accomplish that by making us revel in our failure or dispose us to only take on modest goals in the first instance. Instead, adaptive preferences allow us to adopt lofty goals and only force our preferences to adapt when we believe the end is truly out of reach. Were we not to have such a mechanism, a person might continue striving toward an impossible end, risking despair or injury, all the while wasting her effort. When adaptive preferences

remove a desire for an impossible end, they redirect it toward a more manageable goal. In doing so, they themselves maximize productivity by functioning as a sort of safety value that prevents a person from endlessly toiling in the service of an impossible end.[24]

In this light, adaptive and counter-adaptive preferences appear to be complementary parts of a psychological mechanism that maximizes human productivity. Counter-adaptive preferences keep us working because they stave off satisfaction. Adaptive preferences enhance our efficiency by directing our efforts toward manageable goals by redirecting us when we become stuck. There might, in this way, be an evolutionary explanation for our happiness and the complicated psychological architecture that sustains it. Knowing that the predator runs faster when chasing prey just out of reach, nature might have found a way to keep our quarry far enough away that we run, but not so far that we give up the chase.

In addition to erecting boundaries so that our happiness doesn't rise or fall beyond certain levels, cognitive mechanisms ensure our level of happiness will tend to float back to a neutral point within that range.[25] Research in the last few decades supports 'set point' theory, which holds that a person's happiness is *set* within a relatively narrow range by factors early in life and doesn't vary a great deal.[26] A person will find her neutral state in the middle of this range. While positive and negative outcomes will temporarily push her level of happiness up or down, she'll almost inevitably drift back to

---

[24] See Shane Frederick and George Lowenstein's "Hedonic Adaptation" (pg. 303) *in Well-being: The foundations of hedonic psychology*, Kahneman, Diener, Schwarz, Norbert (Eds). (1999). (pp. 302-329).

[25] Headey, B., & Wearing, A. (1989). Personality, life events, and subjective well-being: Toward a dynamic equilibrium model. *Journal of Personality and Social Psychology*, 57, 731–739.

[26] Easterlin, Richard "Explaining Happiness" *Proceedings of the National Academy of Science*s, Vol. 100, No. 19, pp. 11176-11183, September 2003.

this default spot.[27] Few if any events an ever push an agent out of her hedonic range, and it's similarly rare for one's set point to be altered. Such permanent changes in one's set point are, sadly, almost always decreases in happiness, following traumatic events such as the loss of a child or spouse or a prolonged period of unemployment.[28] But most other profound changes in one's life, positive or negative, do not alter one's set point or even result in a prolonged deviation from one's neutral state.[29] One of the most often cited hedonic studies shows that both lottery winners and recent paraplegics return to their prior levels of happiness within just a few months of these extraordinary events.[30] With very few exceptions, both our default hedonic state and our possible ranges of deviation are set very early in our lives.

However, that fact that *each* of us has a default hedonic point and range within which our happiness moves should not be taken to imply that *all* of our default points are the same *or* that our ranges are equally large. One person can plausibly have a set point substantially lower than another, and a person can likely likely have a broader spectrum than another. On a hedonic scale of one to ten, for example, Alan might have a set point of four and a range that travels from three to five (See Figure 6). On the same scale Beth might have a set point of six and a half and a range of four to ten (See

---

[27] See "Longitudinal Analyses of Psychological Well-being in a National Sample: Stability of Mean Levels," Costa, Paul, Zonderman, Alan, McCrae, Robert, Cornoni-Huntley, Joan, Locke, Ben, and Barbano, Helen, in *Journal of Gerontology* (1987) 42 (1): 50-55

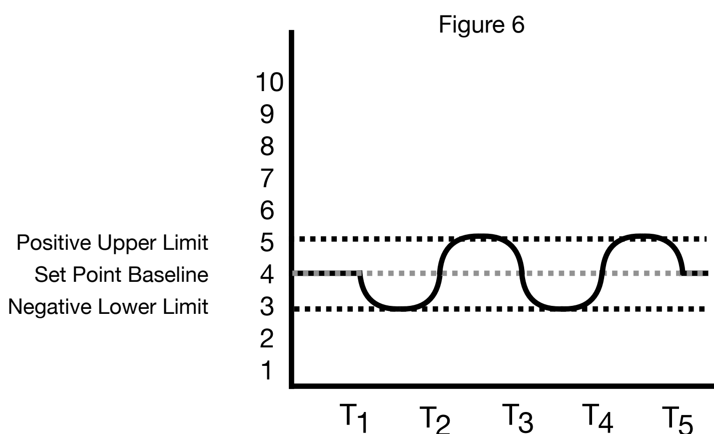
[28] For difficulty in hedonic adaptation following widowhood see "Reexamining Adaptation and the Set Point Model of Happiness: Reactions to changes in Marital Status", *Journal of Personality and Social Psychology*, 2003. Vol. 84, No. 3, 527-539. For difficulty in hedonic adaptation following a prolonged period of unemployment see "Unemployment Alters the Set Point for Life Satisfaction," Lucas RE, Clark AE, Georgellis Y, Diener E. *Psychological Science.* 2004 Jan;15(1):8-13.

[29] Lykken, D., & Tellegen, A. (1996). "Happiness is a stochastic phenomenon" *Psychological Science*, 7, 186–189.

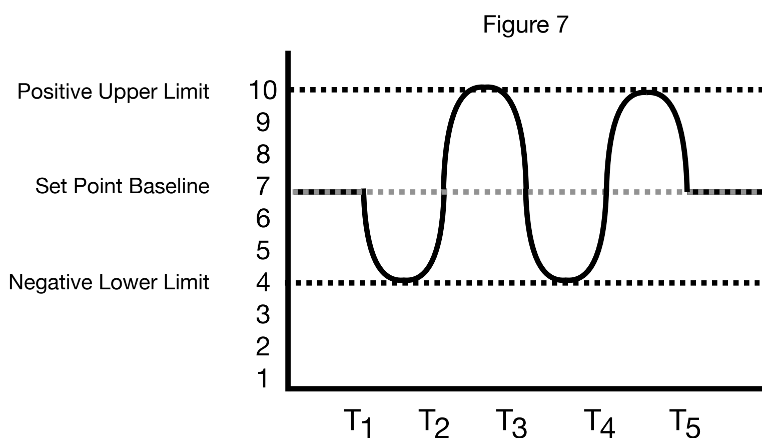
[30] "Lottery Winners and Accident Victims; Is Happiness Relative?" Philip Brickman, et. al, *Journal of Personality and Social Psychology*, 1978, Vol. 36, No. 8, 917-927.

Figure 7). Beth will then be chronically happier than Alan and her mood will be more variable. Where Alan will be a somewhat dour person who will never be much happier or sadder than his default state, Beth will be a cheerful person capable of much larger swings between happiness and sadness. Their adaptive and counter-adaptive preferences will keep them within their respective ranges, always gravitating back to their neutral states. Thus, contrary to desire-satisfaction theories, whether one is cheerful or morose appears to be largely determined by factors independent of her effort or share of distributable goods.[31]

Alan's Hedonic Spectrum



Beth's Hedonic Spectrum



T_1, T_3 Happiness Depreciating Event
T_2, T_4 Happiness Increasing Event

---

[31] This assumes that a person is not experiencing poverty, chronic pain or existential threat, all of which can sustainably depreciate a person's hedonic level and cause sustained misery. It is only after a person has achieved freedom from these that increases in wealth stop materially increasing a person's happiness. Below this point, the traditional model has purchase and a greater share of distributable goods can indeed help those who suffer intensely.
On the issue of increased happiness not following higher income in developed countries see Easterlin, Richard A. (1974) "Does Economic Growth Improve the Human Lot? Some Empirical Evidence" in Paul A. David and Melvin W. Reder (eds.) *Nations and Households in Economic Growth: Essays in Honor of Moses Abramovitz* (New York: Academic Press).

One must concede that neither adaptive preference nor set point theories are without their critics. Debate exists about the limits and duration of adaptive preferences where, for instance, some have noted that paraplegics don't completely adapt to their new circumstances.[32] Critics also dispute the range and possible change of set points, where certain events seem to actually move set points and other events allow longer than predicted deviations from a neutral state.[33] But this level of debate is itself telling. What one does *not* find in contemporary literature is a wholesale rejection of the set point theory and hedonic adaptation *or* defense of our normal language use. Some events exceed our ability to adapt and some of us have a greater adaptive propensity than others. Some events alter set points and some people have broader ranges than others. But these are relatively minor variations on a widely accepted theme. What's not disputed is that our happiness is strongly influenced by a variety of internal limits and cognitive compensations.

Taken together, set point theory and the presence of a psychological immune system complicate the relationship between happiness and conditions in the world. The veracity of set point theory would show that we can't infer a person's happiness *relative to others* from her circumstances. The fact that Alan is routinely less happy than Beth need not imply that things are worse for Alan. In fact, if conditions were equivalent for the two of them, *ex hypothesi*, Beth would be happier. The presence of the so-called psychological immune system would show that a particular agent's level of happiness

[32] "Reexamining Adaptation and the Set Point Model of Happiness: Reactions to Changes in Marital Status", Lucus R, Clark A., Georgellis Y., Diener E., *Journal of Personality and Social Psychology*, 2003, Vol. 84, No. 3, 527–539

[33] Ibid., See also Easterlin, Richard, *Explaining Happiness*, *Proceedings of the National Academy of Sciences* (PNAS), 2003, pp. 11176-11177

will not necessarily correlate with how good or bad conditions are for that agent. Beth's happiness might be more affected by minor physical ailment than by the death of her husband because adaptation and a host of other buffering influences prevent her from ever feeling the full force of the latter event.[34] Thus, set point theory prevents us from using material circumstances to compare happiness across persons, and the psychological immune system prevents us from using it to assess a person's own happiness.

This picture of happiness differs markedly from our ordinary notion, but also explains why our ordinary notion has purchase on us. We believe that satisfying our desires will make us happier, and it does, even if only temporarily. When we're successful, we experience a dose of happiness sufficient to motivate us to work harder. When we fail, we experience disappointment, but only until we find another hill to climb. We're given just enough material success and just enough cognitive adjustment to convince ourselves that we're making progress.

Though our ordinary first-person perspective treats happiness as the ultimate *end in itself*, from the standpoint of the objective observer, happiness appears to more clearly be a *means* to productivity. In its pursuit, we strive after an endless series of goals, redirecting our efforts when necessary. What's remarkable is that despite all of our triumphs and failures, our level of happiness rarely changes dramatically. We spend our lives working toward ends that always appear to be just within reach, but never arrive in a permanent way. We believe that we're making real progress, that we're much happier

---

[34] See Gilbert, Daniel, Lieberman, Matthew, Morewedge, Carey, Wilson, Timothy "The Peculiar Longevity of Things Not So Bad" in, *Psychological Science*, Vol. 15, No. 1. (2004), pp. 14-19.

and better off than we were before. But we seem not to notice that after a few pleasurable moments, we drift back to a default state so the cycle can repeat. Nor do we notice that we're staked to poles at various points on the continuum between happiness and misery, some of us with longer tethers and some of us closer to happiness. Instead, we think that there is just one happiness, that it's equally open to all of us, and, that if we could somehow achieve it, we might remain happy indefinitely. Perhaps the most impressive part of the trick is how well it works on us. We not only fail to see we're puppets, even being told that we are and seeing our limbs rise and fall as strings are pulled, goes no distance toward freeing us.

The fact that alternatively satisfying or not satisfying desires doesn't improve or reduce one's happiness generates two reactions. An immediate one is that we needn't worry about desire-satisfaction. If happiness remains more or less static, it might be more efficient to leave desires unsatisfied and redirect resources elsewhere. At the very least, it does seem that our practical calculus should change from assuming that satisfying desires linearly increases happiness to assuming something closer to a sine curve, where we instead measure the interval of happiness or unhappiness after satisfying or failing to satisfy a desire. The model would shift from seeing maximized happiness as consisting of an ever-increasing sum to one that sought to minimize negative deviations from a set point. But even this change to the standard model seems insufficient for these findings.

A second reaction, and one which comes more slowly, is that perhaps we shouldn't worry about happiness. It seems to be largely impervious to our attempts to increase it

and, more significantly, it doesn't track the good in our lives. One wonders, along these lines, whether happiness should be the sort of thing with which morality concerns itself in the first place. There is some reason to believe that in a society such as ours, the fact that one person's life is less happy than another's might be irrelevant from the standpoint of morality or distributive justice. Seen in this light, greater happiness needn't have an obvious moral significance, as happiness doesn't track virtue and isn't equally available to all even *before* the morally arbitrary assignment of goods. Differentials in happiness might *not* have political relevance, if, indeed, altering material distribution is unlikely to improve matters for the less happy person.[35]

But perhaps the most profound question framed by this new picture is whether greater happiness is even preferable from the *self-regarding* perspective. It isn't entirely obvious that one should prefer a happier life to a less happy life, especially if the latter contains other important goods. The new picture of happiness helps make this question intelligible. On the traditional model we'd be tempted to believe the worthwhile goods in life should contribute to our happiness, making the life with more goods necessarily happier. But seeing that happiness and the goods in our lives can diverge, we are now in a position to ask what should be guiding our action and structuring our lives. Moreover, if we discover another motive that better explains how we structure our lives, it might continue to have explanatory purchase for us even if we ultimately determine

[35] I want to emphasize that the claim that increases in material well-being do not contribute to happiness pertains *only* to relatively wealthy, post-industrial, democratic societies, and thus the new facts about happiness cannot be properly interpreted as a reason to ignore third-world suffering. Relieving poverty, illness, or existential threat reliably and sustainably increases happiness, as well as making available the non-hedonic goods I discuss later. Persecuted, starving, or suffering persons will not fully adapt to their circumstances, so from a distributive standpoint relieving these conditions *will* meaningfully increase net well-being. The new model of happiness has purchase only after these harms have been removed, as they arguably have in many wealthy democratic societies.

that our happiness is *more* closely connected to what's good for us than this hedonic research suggests. However, in order to put us in a position to identify this motive, we should first address the the relation of happiness to moral theory.

To varying degrees our prevalent moral traditions assume the old model of happiness described above. This assumption is part of what gives rise to the notions that happiness is a natural entitlement or the singular end of human activity, neither of which seem obvious given these hedonic findings. Nevertheless, assessing this picture's impact on moral theory generally isn't a trivial matter because a concern for happiness is more integral to some moral views than to others.

A Desideratum of Moral Theory

Before proceeding, I should say something about why a moral theory should take account of happiness in the first place. There are, of course, a variety of theory-specific reasons. If a theory held happiness was a natural entitlement, for example, the theory should allow for one's own happiness and that of others. If happiness were considered an end worthy of maximization, a theory would tend to promote it. But without assuming the substance of any one view, I think providing *some* space for a person's happiness is a desideratum of all moral theories, and the reasons for this are telling of what we expect from a moral theory generally.

It's generally accepted that, among other things, the purpose of moral theory is to explain how we should live and treat one another, using general principles that can be extended to our particular circumstance. There is admittedly debate about the ontology

of moral facts; whether these principles are inherent in the fabric of things or rationally constructed to give order to an inert world. For this reason, what makes a *good* moral theory will turn, in part, on one's conception of what a moral theory should be. But both realist and constructivist conceptions of morality must answer to certain core expectations, or desiderata, we have for any moral theory. Apart from the common criteria for theory acceptance (simplicity, scope, commensurability with other accepted theories), moral theories are uniquely expected to cohere with certain intuitions.[36]

For one who takes morality to be a social construction, the need for consistency with our intuitions is clear. In such a case, we're not so much choosing a moral system that accurately describes the world as we found it, as we're choosing one that describes a moral world we want to make. As such, a constructivist might require that morality should not only describe a system of principles that governs our conduct, but that the system itself must be *agreeable*. Such a requirement is not to be overstated, for the truth about how to live shouldn't bow entirely to one's fancy. But since it is a truth about *persons*, it should be sensitive to human capacities and dispositions. A constructivist can, thus, reasonably hold that a plan for living that rational persons can't endorse, *ipso facto*, cannot be the correct plan.

While that might suffice for those who take morality to be a social construction, one might understandably think a moral realist cannot demand that moral theory must be agreeable in this way. After all, to the extent a theory, moral or otherwise, is supposed to be entirely descriptive, its intuitive appeal, or demandingness, shouldn't be an

---

[36] This is perhaps the most audacious of moral philosophy's burdens and one that is not shared by the study of metaphysics, epistemology, or science. In these latter pursuits, a theory's agreement with our intuitions might be defeasibly confirming data, but not a criterion of adequacy.

overriding concern. Facts about the structure of the world, for example, need not conform to our intuition (i.e. the fact that we might find it unsettling that the material world isn't as it appears to us doesn't bear on whether that's the case). One might wonder, by analogy, how, for a moral realist, a theory's intuitive appeal could be relevant to its truth.

But if a moral realist believed we could treat our intuitions as provisional knowledge, she could consistently hold that a theory's demandingness is relevant to its acceptance. The notion is that while she might not know the structure of morality in its entirety, there are particular touchstones of which she's certain. Her conviction that suffering is wrong and that personal dignity should be respected might plausibly prevent her from endorsing a moral system that denied these things. Where the constructivist can reject a moral theory because it doesn't prescribe a world she could endorse, the realist can reject it because it doesn't describe the world she takes herself to inhabit.[37] But both

[37] The distinction I've drawn between constructivists and moral realists above is somewhat superficial, and further inquiry into their respective use of moral intuitions shows the limits of this distinction. The moral intuitions the constructivist uses to inform her view about the world she would like to make are largely similar to those the moral realist draws on to inform her view about the world she takes herself to inhabit. In both cases, these views are predicated on assumptions about the kinds of creatures we are, on our capacities and dispositions, and what can be reasonably expected of us.

If facts about us help determine the best way for us to act, as surely they must, then our prescription of the world we want is predicated on our description of the world as it is. Another way of putting this is that the world we want is *part of* the world we have, at least to the extent that it's implied by facts about us. The ontology of moral facts for a realist may be predicated on our features and thus fundamentally hypothetical (e.g. given that we are x, the best way for us to behave is y…) and the constructivist might hold the same view of moral facts. These two views would diverge if the realist held that moral facts were somehow prior to or independent of our constitution as persons, or if a constructivist held that moral facts were nothing more than intersubjective agreements. But since the former is scarcely intelligible and the latter is little more than subjectivism, one might think that the most plausible versions of realism and constructivism are co-extensive in this regard. I don't explore these issues more rigorously above because my purpose here is just to show that moral intuitions can be treated as defeasible evidence on a variety of moral conceptions.

are able to treat a moral theory's compatibility with our intuitions as relevant to its acceptance because they both see our moral intuitions as a kind of knowledge.

It it because moral theories treat compatibility with certain intuitions we hold as a criterion of adequacy that they make space for one's happiness. For among these deeply held intuitions is the notion that there must be a way to live morally that doesn't eliminate the possibility of one's own happiness. We would be skeptical of a picture that paints the moral life as so demanding that a modicum of happiness is an immoral indulgence. We can certainly imagine environments so brutal that happiness is impossible and yet morality still has a role to play, but these circumstances are brutal in large part *because* they prevent the coexistence of happiness and morality. While we can imagine them occurring, this is certainly not the world in which we find ourselves.[38]

As a practical matter, moral theories have a motivational incentive to allow for an agent's happiness. If they did not, a person's natural, hedonic impulse would generate a persistent motive to act immorally. This could render the theory unsustainable in the

---

[38] The expectation that morality and happiness must be able to coexist might follow from our relatively privileged circumstance. Ours is a *very* moderate scarcity by any historical measure, since we reasonably expect to be healthy, fed, and safe. This, of course, is strikingly not true of the entire world. In fact, many of the world's people routinely suffer from sickness, hunger, or threat and, predictably, such persons tend to be less happy and have lower expectations for their future happiness. Our high expectations do not fix any entitlement or desert, my point is merely that, rightly or wrongly, we treat the intuition that morality should *not* prohibit our own happiness as a criterion of evaluation for moral theories.

To the extent that morality should be sensitive to our circumstances, differing existential conditions allow different moral predicates to gain purchase. The facts about happiness and what count as reasonable expectations for our lives differ so strikingly between the post-industrial and the non-developing world that different norms of conduct should likely apply unless we're willing to accept the notion that we all belong to the same moral world and we're willing to take on the obligation to improve the conditions of those radically less fortunate than ourselves.

If we postulate that there are two entirely separate worlds, one whose occupants' heath, nourishment, and safety are reasonably assured and one whose occupants' is not, certain moral predicates would likely not find purchase in the more austere world (and correspondingly other different moral predicates might find an application). In such a case, different moral obligations and prerogatives would likely apply, not merely because the two would have different moral rules, but more profoundly because they might constitute distinct and incommensurable moral worlds.

same way that an excessively austere diet might be impossible to follow.[39] But the intuition seems to be *more* than simply a function of motivational necessity. We seem to believe that we needn't, and even shouldn't, have to choose between the possibility of own happiness and acting morally. We acknowledge that at times we must sacrifice morality for happiness, but we have the intuition that their coexistence must be possible. We thus impose this expectation on moral theories and treat their ability to underwrite it as provisionally confirming or disconfirming evidence.

Moreover, the intuition is *not* just that we should have freedom from moral obligation to pursue happiness, but that a moral theory should provide an account of *how* our happiness relates to our obligations. We'd tend to prefer a theory that showed how our moral life can be happy to one that simply required very little and left us to find our happiness elsewhere. Even if the theory holds that other obligations will sometimes override it, we want a moral theory to give an account of our happiness and its importance. Thus, the intuition is twofold: 1) we think our happiness has moral significance and, (2) we believe whatever obligations we owe to others should not excessively compromise our happiness. Underwriting *this* intuition is a desideratum of any moral theory.

On this score, utilitarianism has generally been seen as an unusually appealing theory. Predicated on the notion that happiness is the highest good, classical utilitarianism holds the greatest good follows from maximizing the net happiness of the greatest number. Expected happiness can be quantified and compared across persons,

---

[39] In chapter 4 I argue that objective well being views can fail this motivational requirement for a similar reasons. Even if such views prescribe ends that are objectively valuable, if the agent living that life cannot bring herself to endorse these ends, she will difficult to pursue them in a sustained way.

thus allowing the theory to fully direct both individual and social behavior. Moreover, the theory's calculus is simple: when deciding between two courses of action, that which can reasonably be expected to promote the greatest happiness is preferable. This resembles the rational procedure we use in our practical deliberations; we intuitively tend to compare outcomes and favor those that will make us happiest. This connection can run both ways; we're happiest when we get outcomes that we favor, and we favor the outcomes that will make us happiest. In both cases, utilitarianism supports our sense that happiness should run alongside morality.[40]

By contrast, the Kantian moral world is generally believed not to find as clear a place for happiness, and is sometimes seen as austere. It doesn't promise that our inclinations will lay down with our moral obligations, that a moral life will be a happy one, or that happiness is a goal worthy of our pursuit. Kant's ethics instruct us to eschew our powerful, immediately available emotions in favor of a rational nature that must be cultivated and maintained. While on his view moral education and conditioning can help us embrace our rational natures, he admits our inclinations may continue pulling at us

[40] The connection between happiness and moral obligation is so close in utilitarianism that the theory is sometimes criticized for reducing to a formal principle that one should simply do what's best. Modern variants of utilitarianism, notably consequentialism, border on being unfalsifiable because in simply holding that the action leading to the 'best outcome' is to be preferred. If one establishes that a given consequentialist prescription would not maximize a utility function, the consequentialist would simply change the prescription. The consequentialist is committed to *whatever* result would have the best outcome, but isn't committed to any particular outcome. For this reason, showing that the consequentialist would seem to be committed to advocating a negative outcome, doesn't go *any* distance in showing a deficiency in consequentialism. The consequentialist can simply respond that in such a case, she hadn't properly weighed the outcomes initially, but on the basis of these new considerations, she's prescribe an alternate course of action. It seems sensible, even if a little hollow, to hold that the best outcome is preferable.

and that we can expect to struggle against them throughout our lives. It is, in this way, a demanding view.[41]

Utilitarianism also promises a less schismatic deliberative process. While Kant's ethics treat happiness as a vestige of our appetitive nature and something to be overcome, utilitarianism treats it as a proxy for goodness and thus doesn't require the subordination of our emotional nature. If fighting against one's emotional nature creates unhappiness, *ipso facto*, it's not to be preferred. To the extent that cultivating a psychology that lies down with one's rational convictions promotes happiness, it is to be preferred. So too, if aligning one's interest with that of others promotes greater happiness, utilitarianism prescribes that as well.[42]

This is not to say Kant's ethics lack intuitive appeal. There is something undoubtedly attractive about Kant's view that our exercise of rationality allows us to self-legislate principles of action that define us. The notion that we're autonomous creatures capable of choosing the best way of living corresponds to an intuition we have that our lives are our own and that we have a responsibility to live well. The very notion of self-legislation underscores our sense that we are defined by the rules we use to guide our actions. Nevertheless, Kant's notion that leading a moral life needn't, and often won't, promote one's happiness is less attractive and *this* worry finds no utilitarian target.

---

[41] This is part of why we must recognize that the intuition has two parts. If we simply wanted freedom to pursue happiness, Kantian ethics are generally regarded as providing more space where we're free of moral obligation than utilitarian ethics leaves us. Nevertheless, Kant's ethics are generally seen as being less accommodating of happiness in part because it doesn't treat happiness as a central end to be promoted in the way utilitarianism does.

[42] To this effect, John Stuart Mill argues that incorporating the greatest happiness principle (that the greatest happiness of the greatest number is to be preferred) as one's own motivational maxim constitutes highest evolution of utilitarianism. *Utilitarianism* 4.5-4.8, introduction (pp. 25)

To the extent that a moral theory should accommodate our intuitions about happiness, there is some onus on Kant's ethics to explain why we should accept a picture of our moral agency as a struggle between our rationality and our inclinations. One answer might be that, however bleak, Kant's is a more accurate description of our existential circumstance, and the proper role of a moral theory is to describe the world in which we find ourselves. But our moral intuitions are in some way *about* our basic circumstance, about whether or not we take ourselves to be in such a dire state of affairs and if we don't think our rational capacity and our inclinations must be opposed, we might favor a picture that promotes their harmony to one that doesn't.

Over the course of the following two chapters, I want to explore this commonly held view that utilitarianism finds a more intuitively appealing place for happiness than Kantian ethics. I'll argue that while it initially appears to have purchase, when we push on the details of happiness, the intuitive appeal of utilitarianism begins to unravel and Kant's view starts to look less austere. Moreover, I think this happens even *before* we apply the new understanding of happiness discussed in this chapter, though, predictably, these findings magnify this effect.

My hope is that this inquiry produces a clarified sense of what we want to ask of moral theories generally, not merely a defense of one theory and a criticism of another. Most generally, my purpose is to show that assigning happiness a central role in a moral theory strains that theory in ways that are difficult to avoid. I think the impact it has on utilitarianism nullifies much of the initial intuitive appeal and that after seeing this, we're less inclined to treat Kant's relative indifference to happiness as a deficiency. But more

profoundly, I think this discussion prompts us to ask how much space a moral theory should provide for happiness. If we realize we weren't, and shouldn't have been, chasing happiness in the first place, are we then in a position to understand what we were after all along, and to ask whether this is the sort of thing our picture of self-regarding interests should accommodate? My view is that we are and that it should.

In order to provide a foundation for these questions, I'll begin with a discussion of the two principles of utilitarianism and confront some problems that arise when we apply even our everyday understanding of happiness to them. Because these are familiar issues, I'll consider the strongest utilitarian defenses and try to show why these don't sufficiently resolve some fundamental concerns with the two core principles. The view I'm pursuing is these issues don't merely concern utilitarianism, but rather any view that treats our somewhat fickle happiness as a reliable indicator of goodness or as a worthy end in itself.

Chapter Two
Utilitarianism: Happiness as Goodness

Utilitarianism is reputed to be a profoundly intuitive moral theory, one that's *easy* in precisely the way Kant's ethics are *hard*, and much of this is owed to its treatment of happiness. Below, I try to determine whether utilitarianism deserves this reputation, though my deeper purpose is to examine the appropriate relationship between happiness and moral theory. There are certainly a multitude of versions of utilitarianism, many of which differ from one another dramatically. As such, I will not be providing anything like a comprehensive rebuttal to the doctrine, but instead will focus on particular issues that follow from utilitarianism's two core principles.

The strain I want to consider is a clear progeny of Sidgwick and Bentham's classical utilitarianism and it rests on two related theses. The first thesis is that *satisfying an agent's (authentic or true) preferences will increase her level of happiness and not satisfying them will decrease her level of happiness*. Often referred to as *hedonic internalism*, this claim is predicated on our intuition that an agent is in an especially good position to know which conditions would increase her own happiness and, thus, satisfying her *true* preferences will tend to increase her happiness. Most utilitarians of this ilk rely on an intuitive notion of happiness, one that treats subjective well-being, or moment-to-moment pleasure experienced by the agent, as the barometer of happiness. The substance of the first thesis, then, is that our happiness is dependent on the satisfaction of our true preferences.

The second thesis is that *a person's good consists in maximizing her own happiness*. This makes the moral relevance of happiness clear; happiness is a proxy for human goodness and thus its maximization is obligatory. Accordingly, all things being equal, the more authentic preferences fulfilled, the happier the agents are in the aggregate, and the greater the good attained.

Both these principles have been subject to familiar critiques. The first principle has been criticized on grounds that not all desire satisfaction leads to increased happiness, and, in fact, that satisfying certain desires will likely decrease a person's happiness. I examine this criticism and a frequently resuscitated defense against it that originates in Henry Sidgwick's *Methods of Ethics* that has been modernized by Richard Brandt and then again by Peter Railton. The second principle has been criticized because the connection between personal happiness and human goodness seems to fray. Many things that seem to make us happy aren't good for us, and much of what's good for us doesn't make us happy. I consider this worry in some detail and a telling defense against it, found ironically in the work of John Stuart Mill.

This is not just ironic because many contend Mill isn't properly a utilitarian at all. It's rather that as we watch Mill's defense of utilitarianism, struggling to connect the empirical sources of human happiness with the objective good for persons, we see why he invokes ideas incompatible with the doctrine. The irony I think we ultimately see, and my purpose in considering Mill here, is that I think his defense succeeds at the cost of his position as a utilitarianism. In this way, I think Mill's struggles vividly illustrate the complexities of treating happiness as the highest human good.

I will argue that both the first and second principles encounter meaningful difficulties and that the intuitive appeal of utilitarianism is, in some sense, a chimera. But I think the lesson to be learned is not about utilitarianism *per se*, but about the way in which our happiness and human good have different overarching trajectories. Whereas utilitarianism assumes that satisfying desires increases happiness, which in turn constitutes goodness, I think these connections break at both levels, especially in light of the conception of happiness advanced in the last chapter. I'll argue that satisfying even our informed preferences won't make us happy (in the utilitarian sense) and that happiness isn't anything like our highest good. The upshot I'd like to move us toward is the hard recognition that happiness shouldn't be our highest, or even a particularly high, end because it needn't constitute human goodness and it isn't sufficient for a good life. Ultimately, I think, we want something else, something selfish, and we're overwhelmingly willing to trade happiness for it if given the chance. My treatment of utilitarianism, like my treatment of Kant's ethics, is intended to move us closer to understanding what that thing is.

The First Principle

We should immediately note that the concept of happiness discussed in the last chapter is straightforwardly problematic for the first principle of utilitarianism. The notion that satisfying preferences, however well-informed, will reliably increase one's happiness seems generally inconsistent with the presence of cognitive mechanisms that (1) prevent one's happiness from rising above certain levels, (2) that cause this

happiness to return to a baseline, and (3) that attenuate particular sources of happiness to keep us motivated.

According to the first principle of utilitarianism, satisfying a person's preferences should increase her happiness, all thing equal. But if a person has a ceiling on her hedonic range, her happiness cannot increase beyond that point, granting more extravagant preferences will not tend to have proportionally larger gains in happiness than less extravagant preferences. Additionally, if a person has a set point that pulled her back to her hedonic baseline, granting this preference would not result in a *sustained* increase in happiness. Similarly, if another pre-conscious mechanism (the purpose of which is to encourage more activity) causes the the joy that follows from satisfying this desire to dissipate quickly, the result would be the same. She would not experience a *sustained* increase in happiness following the satisfaction of any desire.

One could modify a utilitarian distribution to try to accommodate these factors. To maximize hedonic benefit of resources, it would seem most efficient to satisfy those preferences that have the greatest benefit relative to their cost (i.e. smaller satisfactions are less likely to be restricted by the upper limits of one's hedonic range).[43] If agents inevitably return to a hedonic baseline, or if motivational factors non-consciously attenuate the happiness that follows from preference satisfaction, we might revise the utilitarian calculus to track the duration and extent of a positive deviation from the

---

[43] Here again, the limiting case of the non-developing world offers an illustration. If, as surely is the case, a modest redistribution of resources from the wealthy, industrialized world to those in great need would generate substantially more net happiness, such redistribution would clearly be justified on utilitarian grounds. It's also worth noting that such redistribution could be justified by a variety of non-utilitarian considerations (it could plausibly enable the dignity and flourishing of other persons or, alternatively, provide them with goods to which they have a right or entitlement.)

baseline, instead of treating gains in happiness as static or permanent. Conversely, if not satisfying a preference doesn't permanently lower happiness and negative deviations are limited by a similar tendency to drift back to a set point, modifying the calculus to minimize negative deviations could deal with this as well. While such considerations make the calculation more complicated, a utilitarian account might be able to accommodate these. But fundamentally, I think this new picture of happiness suggests that preference satisfaction will *not* lead to the increases in happiness that the first principle of utilitarianism implies because it suggests that the aggregate happiness in a life isn't going to change substantially through our choices. The overall picture is that we're self-regulating machines and despite all of our toiling, our happiness will generally find its own level.

However, in what follows I want to consider a difficulty the first principle of utilitarianism encounters even on our ordinary conception of happiness. The worry begins at a reasonable place. As intuitive as it is to associate desire-satisfaction with happiness, the connection begins to fray. There are desires whose satisfaction will not lead to increased happiness, such as the desire to have another piece of cake when one is already full or the fleeting desire to strangle one's spouse. In fact, when we reflect upon our actual desires, we realize the fulfillment of many would decrease our happiness.

An immediate and sensible modification to the first principle of utilitarianism has thus been that only the satisfaction of *informed preferences* contribute to subjective well-being. An agent might fail to have an informed preference if she is (1) reasoning from

false information, (2) lacks information relevant to the desire in question, or (3) commits serious errors in reasoning with the information she has.[44] Fleeting desires might either follow from faulty information (violating #1) or present themselves as more permanent than they actually are (thus violating #2).

While they may seem innocuous, these requirements might be interpreted so broadly as to severely constrain one's set of informed desires. The restriction on incomplete or imperfect information prompts one to wonder *how* complete and accurate the information regarding the desire must be. One wonders if this requirement pertains only to knowledge of the object or extends to the agent's self-knowledge. For my desire to eat a piece of cake to be informed, for example, I must know certain things about the cake (i.e. that it is edible, that it tastes more or less as I expected). But certain self-knowledge is also relevant to the desire (i.e. that I'm not presently full and that I enjoy eating cake generally).

One wonders if other sorts of self-knowledge are required for the preference to be informed. It's relevant if my desire will pass quickly, so some further level of self-knowledge is surely required for a preference to be informed, but it isn't clear *how much*. One might argue, for instance, that for a preference to be informed, I must know whether my desire for the cake is really just a craving for sugar caused by my dearth of certain neurotransmitters. This bit of self-knowledge might show that I don't actually have a preference for cake itself (thus violating #2), but rather a desire for more

[44] My desire to take the highway on my commute home in uninformed if I believe that the highway runs in the direction of my house, when, in fact, it does not (#1). Similarly, my desire to take the highway to save time is uninformed if, unbeknownst to me, there is highway construction occurring that will slow me down (#2). Lastly, my desire to take the highway in order to get home more quickly is uninformed if I am aware either that there is construction on the highway or that the highway doesn't lead toward my home (#3).

serotonin which I might satisfy by physically exercising or developing more fulfilling interpersonal relationships. My failing to realize this might render my preference uninformed, hence the response, *you don't really want the cake, you're just depressed*.

There are also concerns about whether a preference must be consistent with the rest of a person's desires to be informed. One wonders, for instance, if the fact that I want to be healthy counts as "relevant information" when I form a desire to eat cake. Relatedly, it isn't entirely clear whether failing to consider my preference for health when forming my desire for the cake should count as a serious error in reasoning (thus violating #3). One worry is that requiring this degree of psychological insight and consistency would likely negate two of utilitarianism's principal virtues, universality and quick application. With such robust knowledge demands, it would seem the utilitarian calculus could only be accurately and quickly performed by the wise.

But rather than dwelling on utilitarianism's practical applicability, I want to ask what we would want if our desires *were* sufficiently informed. The assumption underlying an informed preference requirement is that we're aiming at happiness and when we miss, it's due to some defect in our faculty of desire, such as reasoning poorly, or with false information, or relying on an unsteady psychology. But the assumption is that if we refined our set of preferences by removing all these irrationalities, we'd then desire things that *would* reliably and incrementally increase our happiness. The possibility that satisfying even these sanitized preferences might not increase our happiness would be blow to utilitarianism's practical value. For in such a case, even if

happiness were the highest good, utilitarianism wouldn't give us a reliable mechanism to access this happiness.

I want to try to determine, then, whether satisfying even our *truly* informed preferences would reliably increase our happiness, first using our ordinary conception of happiness and only then layering the complexities imposed by modern hedonic research. In order to give this the best chance of success, I'll use one of the most rigorous standards to which we could hold these preferences. The thought is such a standard would only admit our most reasonable and healthy preferences, which are the most likely promote our happiness. The most influential standard for informed preferences originates in a provocative, but often misappropriated passage from Henry Sidgwick's *Methods of Ethics*.[45] Sidgwick writes,

> ...a man's future good on the whole is what he would now desire and seek on the whole if all consequences of all the different lines of conduct open to him were accurately foreseen and adequately realized in imagination at the present point in time.[46]

This passage has been read to suggest that *fully informed* preferences are those a person would prefer even after recognizing and comparing all actions open to her and

---

[45] See Sobel (1994) pg. 792, Rosati (1995) pg. 297, Qizilbash (2006) pg. 85, each of whom cites this passage of Sidgwick's *Methods of Ethics* when discussing full information accounts of well-being. However, for a discussion of how Sidgwick is inappropriately invoked in the discussion of full-information accounts, also see Robert Shaver's "Sidgwick's False Friends" (*Ethics*, Jan 1997, 314-320). Also see footnote #47.

[46] Sidgwick, Henry *Methods of Ethics* (pp. 111-112)

their consequences.[47] Sidgwick also mentions that he is supposing "...the desirer [would] possess a perfect forecast, emotional as well as intellectual, of the state of attainment or fruition," thus assuring the desirer would have an ideal intellect and disposition in addition to all the information available.

This staunch requirement is reflected in twentieth century utilitarian Richard Brandt's proposal that the only preferences that should count toward an agent's well-being or good are those that would survive a process of aggressive cognitive psychotherapy. Such a process would, in Brandt's view, allow an agent to experience 'value-free reflection' because it:

> …relies simply upon reflection of the available information, without influence by prestige of someone, use of evaluative language, extrinsic reward, or punishment, or use of an artificially induced feeling-state like relaxation.[48]

Brandt believes that such a process would avoid preferences formed by false information, generalizations, culturally transmitted norms, and by trauma, deprivation, or excess in early childhood.

[47] As a matter of scholarly interpretation, such an attribution is almost certainly unfair to Sidgwick. Though this passage can be read as advancing an audacious standard, Sidgwick was not proposing constraints on informed preferences as such. While in the pages preceding this passage, Sidgwick does claim that our actual preferences will frequently be defective, Sidgwick does not claim that the preferences that emerge from his thought experiment would necessarily contribute to our *happiness*. Sidgwick's concern in these pages is to identify which preferences would identify "*the good on the whole*" (112) for that person, not necessarily what would make that person *happy*. Sidgwick does not commit himself here to the claim that the good for a person lies entirely in his own happiness, and as such, is not here arguing for utilitarianism. While Sidgwick is commonly considered one of utilitarianism's principal architects, it's more charitable to read *Methods of Ethics not* as a defense of the doctrine, but rather as an examination of the way major ethical theories are intertwined in our ordinary thought. On this last point see J.B. Schneedwind's *Sidgwick's Ethics and Victorian Moral Philosophy*, pg. 192.

[48] Richard Brandt, "A Theory of the Good and the Right", pp. 234-45

One worry is that the idealizations that Sidgwick and Brandt suggest might result in a self with such rarified tastes they would have little relevance to an agent's present self.[49] Our intellectual and psychological imperfections might make certain desires appropriate for us, where they would be inappropriate for more perfect creatures. Sidgwick and Brandt's restrictions, then, run the risk of what might be called a *mismatched preferences* problem, namely these restrictions might impose on us the preferences of beings very different than ourselves. The mismatched preferences problem could give rise to some of the paternalism that utilitarianism characteristically avoids, since some of these idealized preferences might be so alien to an agent that she could not endorse them. One wants the *idealized-self* to use her powers of reflection and clear temperament to prescribe a course of action *not* for herself, but for our present and imperfect self.

Peter Railton has, more recently, proposed a version of the idealized-self that seeks to resolve this problem by disambiguating the powers of the judge and the limitations of the actual agent. He writes:

> ...an individual's good consists in what he would want himself to want, or pursue, were he to contemplate his present situation from a standpoint fully and vividly informed about himself and his circumstances, and entirely free from cognitive error or lapses of instrumental rationality.[50]

The notion is that an agent, Sarah, should imagine a being like herself but with unqualified powers of intellect and imagination and full information regarding her physical, environmental and psychological circumstances. This idealized judge or

---

[49] David Sobel treats this problem as a factor that motivates Peter Railton's idealized judge account. See Sobel (1994) pg. 792-793

[50] Peter Railton, "Facts and Values," *Philosophical Topics* 14 (1986) 5-31

advisor, let's call her Smarter Sarah, would know all of the possible consequences of any of Sarah's actions, and could see all of the possible lives which Sarah might lead. Whatever Smarter Sarah would prescribe for Sarah would, *ipso facto*, be Sarah's fully informed preference. Railton's move from considering *what an agent would want* to considering *what an agent would want herself to want* enables him to distinguish between what Smarter Sarah would want for herself and what she would want for Sarah. This way, the judge's perfect powers of reason and stable temperament don't distort the prescription, but are rather used only to identify what would be best for the actual agent.

This version of a full information account has certain advantages. It purports to preserve some aspect of hedonic internalism, since the agent determining the good is still herself, she's just using a refined version of herself.[51] This, in turn, affirms our intuition that we ourselves would be in the best position to know what is good for us, were it not for certain intellectual and psychological shortcomings.[52] When we ponder our good in ordinary circumstances, we regularly try to discount for our own psychological distortions and, in this way, Railton's view formalizes a common method of identifying the good for ourselves. It does seem that if we could somehow see our

[51] On the other hand, one might argue that it isn't clear how much of our identity would be preserved after one removed all of our cognitive errors and lapses of reason. Hence Francis Bacon's dictum, "There is no excellent beauty that hath not some strangeness in the proportion." It's likely that perfect rationality would result in some alteration to our set of preferences at least to the extent that we would eliminate those that are inconsistent or ill-thought in the first instance.

[52]. It isn't obvious, however, that Railton's view does preserve the intuition that we're in a particularly good position to determine what's best for us. While the preferences aren't mismatched in his version, one might argue that a third party is nevertheless prescribing an agent's good. While the actual-self can take solace in the fact that her prescribed good is determined by a wiser and more balanced source than herself, she isn't determining it and thus we might lose the intuition that she knows what's best for herself. It isn't clear that Railton's view, that the sources of your happiness are best seen by a perfectly rational and psychologically neutral version of yourself, preserves the intuitive appeal of hedonic internalism.

future lives and compare them from a place of equanimity, we would then be in the best position to determine what would make us happiest, what is good for us, and thus what we should want.

This being said, there is a question about which preferences a full-information account such as Railton's would prescribe for us.[53] In order to avoid issues about the possibility of such foreknowledge, let us assume that these life-structuring decisions have just one determinative outcome and a judge is able to foresee these without distortion.[54,55] Let's imagine the judge was given a comprehensive 'report' displaying all available information about the two possible outcomes of a life structuring decision.[56] The question is whether with this wealth of information we can expect the idealized judge to unreservedly recommend the path of greater happiness.

---

[53] Whether as a matter of historical fact Sidgwick, Brandt, or Railton actually hold such views is not at all central to my purpose. Instead, I want to consider whether even the most robust kind of full information view could identify the conditions of an agent's happiness. If it cannot and, as we have seen, neither the use of (1) less robust informed preferences nor (2) actual preferences can identify the sources of a person's happiness, I think that we must consider that hedonic internalism is implausible generally. If fully informed preferences *can* identify the conditions of an agent's happiness then we would be in a position to advance to utilitarianism's second thesis and inquire as to whether that agent's good consisted in maximizing that happiness

[54] Full information accounts like these seem to suppose that there is a "normal" psychology that is distinct from distortion. It might be that there is no such neutral psychology and that persons simply experience varying levels of anxiety and compulsion. Nevertheless, for the sake of argument let's assume there is some *normal* psychological makeup and that our idealized judge possesses this.

[55] As David Sobel and Connie Rosati have noted, there seem to be two principle ways such future possibilities could be imagined. (See David Sobel (1994) pgs. 796-807. Connie Rosati (1995)) The information about a possible path for one's life could (1) be reported, depicted, or otherwise relayed to the idealized agent or (2) the idealized agent could actually experience these possible paths as lives she lived. While Sobel and Rosati both advocate the experiential model, it suffers from a dual-consciousness problem. While the experiential model creates the expectation that one compare the subjective experience of two different futures, it is impossible to inhabit two consciousnesses simultaneously. As such, the experiential model is either incoherent (if imagined simultaneously) or collapses into the report model (if imagined serially). As such, I adopt the report model as the most plausible version a full information account.

[56] Also imagine the judge could interview the agent, if necessary, to understand how it felt to live the life.

To see, let us suppose Sarah is trying to determine whether she should pursue a career as an artist or as a banker. Sarah is drawn to the creative process but is concerned that such a life will not afford her sufficient means to sustain herself. She doesn't have an interest in business and frowns upon people who have materialistic leanings. Because she's unable to choose, Sarah asks an idealized version of herself, Smarter Sarah, to evaluate both her future as Artist Sarah and as Banker Sarah and determine which is best for her.

Imagine Smarter Sarah sees that twenty years hence Artist Sarah is ambivalent about her choice. On one hand, she occasionally derives satisfaction from the creative process. On the other, Artist Sarah's lived on a modest income for many years and is frequently concerned about her finances. While she's never suffered poverty, she envies the comforts afforded by wealth. With the passing of time, she's become increasingly inured to the artistic value of her work and sensitive to her economic condition. Artist Sarah has, at times, lapsed into using her salary, not her art, as a metric of self-worth. Despite this, as Smarter Sarah presently looks onto this future, she is struck by the beauty of Artist Sarah's creative work and sees Artist Sarah's life as elegantly sparse as opposed to meager. Smarter Sarah is impressed by her future self's intellectual and artistic depth, but troubled to see that financial concerns weigh upon her and that she isn't particularly happy.

When Smarter Sarah considers Banker Sarah's future, she's dismayed at how content she is in this career. Because of the professional demands on her time, Banker Sarah stopped making art or even thinking of art on a regular basis. Instead, she grew

to appreciate market efficiency and the fruits it afforded. Banker Sarah developed expensive, if somewhat gaudy, tastes, and came to believe her prior interest in art was sophomoric. Looking at this future, Smarter Sarah is disappointed in Banker Sarah's expensive tastes and indifference to art, but she sees that Banker Sarah is undeniably happy in this life.

For utilitarian purposes, the choice should be clear. Banker Sarah is clearly happier than Artist Sarah, so if we truly are seeking happiness above all else, Smarter Sarah should unhesitatingly recommend the life as a banker. But it doesn't seem obvious that she should. One might believe recommending the banker's life would be excessively paternalistic of Smarter Sarah, as this is a life that Sarah presently finds objectionable and, more profoundly, Smarter Sarah would be recommending that Sarah *become* a person that Sarah presently finds unattractive. It wouldn't be unreasonable for Smarter Sarah to find the banker's life too alien to endorse and thus recommend the artist's life to Sarah. However, one sees that in doing so Smarter Sarah would be advocating a life that Sarah will ultimately enjoy less and about which she will be increasingly ambivalent.

Railton's invocation of an ideal observer doesn't entirely settle the issue because Artist and Banker Sarah are not only distinct from one another but also Sarah as she presently is. Sarah's decision not only dictates which of two different lives she will experience, but which of two different objects she will become. The ideal observer, then, is trying to determine not only which of two lives Sarah would prefer to live but which of two people Sarah would prefer to be.

Beneath the surface, the choice between these lives turns on two means of assessing value, a subjective measure of how a life feels to live and an objective assessment of which life, or person, is more worthwhile.[57] The former would recommend Banker Sarah because she's happier, which, of course, is a characterization of how the life feels from the *inside*. The latter would recommend Artist Sarah's life because it appears more noble when viewed from the *outside*. One couldn't reasonably prefer to *feel* the way Artist Sarah does, but one could prefer to *be* what she is. Our desires for ourselves, then, appear to concern not merely future states of feeling, but also future states of being. The fact that we might prefer Banker Sarah's life indicates that we have an interest in promoting a pleasant subjective experience. But the fact that we might prefer Artist Sarah's life indicates our interest in happiness can be outweighed by a concern for what we are or will become.

Our concern for our subjectivity is often overstated, whereas our interest in ourselves as objects goes often unnoticed. We sometimes confuse this latter concern with our interest in autonomy or self-determination. However, autonomy concerns the freedom to choose, whereas our interest in ourselves as objects in the world concerns that which is chosen. One could autonomously choose happiness as easily as one could autonomously choose a less happy but more meaningful path. The temporal aspect of Sarah's case illustrates this difference. There's a sense in which choosing to become the banker actually supports Sarah's future autonomy. After all, when she

---

[57] By *objective* assessment of a life, I don't intend that this measure is more correct or impartial. Rather when I refer to the objective assessment of a life or a person, I'm referring to the assessment of that thing as an object, when viewed from the outside. The distinction is not between appearance and reality, but between the way a thing feels (from the inside) and how it appears (from the outside).

becomes Banker Sarah, she'll be pleased with the path she chose, whereas she'd be consigning Artist Sarah to a life she won't be able to fully endorse. The reason to choose Artist Sarah's life is that, from Smarter Sarah's present standpoint, Artist Sarah strikes present-Sarah as a finer object and more worthy of respect, than Banker Sarah. Since Artist Sarah is less happy and since she would prefer Banker Sarah's life, any argument in favor of choosing Artist Sarah's life indicates that the interest we have in the objects we will become might outweigh *both* our interest in our future happiness and our interest in autonomy.[58] It's revealing that there's *any* argument at all self-regarding to be made in favor of a person consigning herself to a life that she'd enjoy less than an alternative (and a life that later she would prefer to opt out of.)

I think what's telling about this thought experiment is that the answer isn't obvious. It wouldn't be entirely unreasonable of Smarter Sarah to recommend the banker's life either because it is a happier one or because her future self would prefer it. But neither does it seem unreasonable that she recommend the artist's life on the grounds that as Sarah is presently constituted she believes Artist Sarah is a finer thing to be. Nevertheless, if one did advocate the artist's life, the act of making oneself into a person with which her future self will be unsatisfied should, I think, give one pause. What's illuminating about this complexity is that these appear to be distinct interests that guide

[58] The issue of autonomy is more complicated than described above. One might argue that choosing Artist Sarah's life shows a respect for Sarah's *present* autonomy because she's choosing a life that from our present standpoint seems best according to some measure. While this is true, she is still consigning her future-self to a life she'll be at best ambivalent about. The choice does then prompt questions about our present self's prerogatives and obligations in choosing a life that a future self will have to lead. Nevertheless, Sarah presumably maintains the option to opt out of either life, so some degree of autonomy is always preserved for her.

our preference formation; we have an interest in happiness, an interest in autonomy, and a separate interest in the thing we become.

However, I think the consequence for the first principle of utilitarianism is clear. This principle holds that if we could see the future without distortion, we would choose the outcome that would make us happiest. If granting a person's most informed preferences might *not* increase her happiness, because she might reasonably prefer something else, then happiness is not her *singular* end. The first principle, that satisfying a person's informed preferences will increase her happiness, is a seductive thought, not only because we expect to want happiness above all else, but also because one might think that getting what we want most would *itself* make us happy. But neither seems obviously true. We might prefer an outcome that makes us less happy, get precisely what we expected, and still believe that we chose well.

While this result does not depend on the new data about happiness, incorporating this data into our life-structuring decisions would tend to make us discount the significance of future happiness more when weighing outcomes. One might then argue that Sarah would have far *less* reason to prefer the banker's life because its primary appeal is the greater happiness it promises. But if the new picture of happiness is indeed accurate, it suggests that the hypothetical is impossible in the first place because, barring unforeseeable tragedies and contingencies, the level of happiness in the lives would be more or less equal. The causes of unhappiness might differ; the artist might worry about her financial condition where the banker might worry about professional rivalries, but the level of unhappiness would be broadly similar.

One life might contain worries we, looking in from the outside, consider serious where another life's worries strikes us as trivial, but their relative severity need not be proportionate to the level of distress they cause.[59] The banker's unhappiness with her failed orchid garden might equal the artist's sadness that she cannot afford to purchase a home. We'd tend to be less sympathetic to the former because of the relatively smaller impact on a life it will have when viewed from the outside, not because it causes less unhappiness. Our intuition is not that the failed orchid garden *doesn't* cause as much misery, but that it *shouldn't*.

We believe this because we tend to improperly, though understandably, assume that the significance of a harm on a life when viewed from the outside should determine the amount of unhappiness it causes. We expect that the person facing the loss of a loved one or marital instability will experience more unhappiness as a result than the person who fails in her gardening hobby. But easier, more privileged lives can plausibly contain less happiness than more difficult lives. We might want to argue the more privileged party is to be faulted, that her discontent is unjustified and that it follows from a failure to understand how lucky she is relative to others. We might think this is a species of uninformed preference, either reasoning from false information (failing to realize that others have more difficult lives) or reasoning badly (knowing that other lives are harder but failing to account for it). But the new hedonic data suggests this needn't be the case. The privileged person could realize her life is easier and that she *should* be happier than she is and yet this might go no distance to increasing her happiness

---

[59] "The Peculiar Longevity of Things Not So Bad": Daniel T. Gilbert, Matthew D. Lieberman, Carey K. Morewedge, Timothy D. Wilson, *Psychological Science*, Vol. 15, No. 1. (2004), pp. 14-19

because she's tied to a lower set point. It doesn't seem to be an issue of uninformed preferences, but rather that our happiness doesn't track the relative severity of conditions in the world.

There are some implications for our discussion of utilitarianism's two principles to be found here. First, it's clear satisfying informed preferences will *not* lead to sustained increases in happiness if people are bound to set points, hedonic ranges, and have the propensity to counter-adapt to the attainment of their ends. But second, and perhaps more tellingly, our lack of sympathy for the privileged person suggests that we are not merely concerned with a person's happiness, but its origin or legitimacy (which foreshadows our discussion of utilitarianism's second principle, the connection between a person's happiness and what is good for her).

Lastly, if, despite the most informed preferences imaginable, we can reasonably prefer outcomes that make us no happier, and even less happy, than alternatives, then utilitarianism's first principle becomes problematic even before applying the modern concept of happiness. The reason might be so obvious we tend to miss it: *satisfying our most informed preferences might not make us happier, because we might not prefer greater happiness to an alternative*. Something entirely apart from happiness and other-regarding moral obligation might be at play in these deliberations. The new hedonic data simply emphasizes how disconnected our happiness is from the world and the choices we make.

The Second Principle

Whereas the first principle is an empirical claim about the nature of human psychology, the theory of good utilized by utilitarianism lies in the second principle: that *happiness is the highest good* and that *the right action is that which maximizes net happiness*. When combined with the first principle, that satisfying informed preferences increases happiness, the claim that happiness is the highest human good allows the inference that *satisfying informed preferences enhances the highest human good*. Because the right action is that which maximizes the highest good, the right action is that which increases net happiness.[60] While there are certainly versions of consequentialism that don't presuppose that happiness is the highest human good, these theories aren't relevant to the present discussion about happiness's appropriate place in moral theory and, thus, I don't deal with them here. That said, the notion that happiness is indeed the highest good is common in various versions of utilitarianism and in what follows I consider whether the doctrine has the resources to defend it.

Two issues immediately surface when one treats happiness as the highest good and the spring from which all other goods flow.. First, while we think that happiness is good generally, other things, such as equality, honesty, and kindness strike us as being good as well. Utilitarians can consider these *derivatively* good, or good because they

[60] This is actually a simplification. Utilitarianism holds that the right action is the one that produces the *most utility or* good and the classical utilitarians, Mill and Bentham among them, interpret good as happiness. So for them, the right action is that which produces the most happiness. Mill himself writes, "The creed which accepts as the foundation of morals, Utility, or the Greatest Happiness Principle, holds that actions are right in proportion as they tend to promote happiness, wrong as they tend to produce the reverse of happiness," though his view is more complicated as I discuss in the following pages (*Utilitarianism* 2.2.1)

contribute to happiness.[61] But if these were good in themselves *or* if their goodness derived from a source other than happiness, it would be problematic for the strain of utilitarianism I discuss here.

This connects with a second concern; namely, the sources of our happiness don't always seem particularly good. Sometimes it seems (1) that we're happiest performing actions that don't appear to contribute to our good and (2) that the actions that do contribute to our good often don't appear to make us happy. We might be happiest indolently passing our days, manifesting nothing that even resembles the virtue or flourishing we generally associate with human goodness. The exercise of intellect or virtue that we would tend to more clearly associate with human goodness often generates more frustration than happiness in people. In finer versions of ourselves there might be greater overlap between our experience of happiness and the good, but one wonders if the overlap in us is sufficient to treat *our* happiness as unqualifiedly good at all.

Of these concerns, the second is the more fundamental to the extent that the first pertains to whether happiness is the *only* or *highest* good and the second concerns whether happiness is good *at all*. It's possible that happiness might be a good but fail to be the highest good, but if happiness were not properly a good at all, it trivially couldn't be the highest. Because the second is more foundational, I'd like to focus inquiry on whether happiness is indeed an unqualified good. The underlying issue, stated generally, is that the second principle of utilitarianism identifies a psychological

---

[61] "It results from the preceding considerations, that there is in reality nothing desired except happiness." *Utilitarianism*, 4.8.1. See also 2.2.10 - 2.2.14.

phenomenon (happiness) with a normative state, (goodness) and these don't appear to be entirely coextensive. The task for the utilitarian, or anyone who would want to associate happiness and goodness, is to explain how happiness as it occurs in us constitutes human goodness in such a way that (1) *excludes* the apparent sources of our happiness that aren't particularly good and (2) *includes* instances of goodness that don't appear to make us particularly happy.[62]

John Stuart Mill comes close to accomplishing both (1) and (2) in *Utilitarianism*. But before proceeding, I must admit that calling on Mill in this instance is at least counter-intuitive because Mill's status as a utilitarian is the subject of considerable disagreement. One one hand, Robert Audi expresses a popularly held view when he describes Mill as utilitarianism's "…best known proponent," and sees Mill's *Utilitarianism* as the doctrine's "classic defense."[63] Mill doesn't help us avoid this conclusion continually when avows many of the basic theses of utilitarianism and refers to himself as a 'utilitarian,' even late in life.[64]

Nevertheless, there are deeply un-utilitarian convictions in Mill's work. He believes equality and liberty are good *in themselves*, which seems plainly inconsistent with

---

[62] One way to do this, at least in the abstract, would be to show that we're simply wrong in judging that which makes us happy. One might argue that what *appears* to be happiness flowing from things which aren't good isn't *true* happiness at all and that the instances of good that strike us as neutral or even negative actually do make us happy and we just fail to see this. Such a strategy would work best with an objective standard of happiness, where a person is *not* assumed to have an authoritative judgment on what makes her happy. But utilitarianism avows a subjective standard of happiness, under which, generally speaking, people are in a good position to assess their own happiness. Hence, the utilitarian has the challenging task of showing how happiness as we subjectively experience it is the highest good. I will proceed to argue that Mill is, despite appearances, best interpreted as having an objective notion of happiness. This will resolve, I believe, the worry for him, but at the cost of any claim to being a classical utilitarian.

[63] The Cambridge Dictionary of Philosophy, ed. R. Audi, Cambridge, 1995

[64] Coope, pp. 35

happiness being the highest and singular good.[65] Mill seldom talks about *maximizing* happiness, which isn't in keeping with an essentially maximizing doctrine.[66] Also, Mill's account of justice (even in chapter five of *Utilitarianism*) suggests that an action's rightness or permissibility does *not* derive from it's consequences, which is tantamount to rejecting one of the basic precepts of consequentialism.[67] L.W. Summer, D.P. Dryer, and David Lyons take Mill to be neither an act nor a rule-utilitarian.[68] Anthony Quinton holds that Mill's *On Liberty* "…is only vestigially utilitarian."[69] John Plamanatz provocatively concludes that, "[t]here is not much left of Benthamite utilitarianism, when John Stuart Mill has completed his defense of it. What is left is, strictly speaking, not utilitarianism at all…"[70] These divergent interpretations of Mill's core commitments are not an accident, I believe, for there are considerable resources in his work suggesting both utilitarian and deontological commitments.

Understanding what Mill is doing in *Utilitarianism*, then, turns out to be something of a challenge. One can find inconsistencies and what appear to be plainly fallacious arguments in this work, but when this happens we would do well to remind ourselves that Mill was one of the great minds of the the nineteenth century and what might appear to be superficial issues are actually Mill's attempt to operate within some deep

---

[65] "Was Mill a Utilitarian?", Christopher Miles Coope, *Utilitas*, Volume 10, Issue 01, March 1998, pp 33 -67, (pg. 36-37 for the intrinsic evil of inequality and 38 for the inherent value of liberty.)

[66] Ibid, pp. 39-42

[67] "Mill's Deliberative Utilitarianism", pg. 69, David O. Brink, Philosophy & Public Affairs, Vol. 21, No. 1 (Winter, 1992), pp. 67-103

[68] "Justice, Liberty and the Principle of Utility in Mill", by D. P. Dryer, and "The Good and the Right' by L. W. Sumner both contained in *New Essays on John Stuart Mill and Utilitarianism*, ed. W E. Cooper et al., Canadian Journal of Philosophy, Supplementary Volume, v (1979) Dryer, p. 63; Sumner, p. Ill

[69] *Utilitarian Morality,* 2nd ed., London, 1989, p. 39.

[70] John Plamanatz, *The English Utilitarians*, Oxford, 1958, p. 144

conceptual constraints. One of these constraints is precisely the question before us; namely, how can we connect our experience of happiness with human good in a way that allows us to avoid the conclusion that the sources of our happiness aren't particularly good and the human good doesn't make us happy.

The reason to consider Mill in this moment, then, is that the problem he's confronting is fundamentally the same anyone, utilitarian or other, would encounter when trying to connect a somewhat fickle subjective experience with a normative state. Ultimately, I believe the most charitable interpretation of Mill's argument, the one that stands the best chance of connecting happiness and human goodness, is one that's inconsistent with utilitarianism. Examining Mill's argument is, thus, revealing because it illuminates the conceptual tensions at play for utilitarianism.

Before engaging the substance of Mill's defense of the doctrine in chapter two of *Utilitarianism*, it's instructive to understand his commitment to empiricism in chapter four both because it's an example of a tempting moment to superficially criticize Mill and because undertaking this commitment constrains the resources available to him to connect happiness and goodness. As Mill explains why happiness is worthy of desire, he famously appears to equivocate, between claiming happiness is universally *desired* to claiming happiness is universally *desirable*. He writes:

> No reason can be given why the general happiness is *desirable*, except that each person, so far as he believes it to be attainable, *desires* his own happiness. This, however, being a fact, we have not only all the proof which the case admits of, but all which it is possible to require, that happiness is a good: that each person's happiness is a good to that person, and the general happiness, therefore, a good to the aggregate of all persons.[71]

---

[71] Mill, *Utilitarianism* 4.3.4 (my italics for emphasis)

That statement that ‘x is desired’ is an empirical claim (i.e. that people *in fact* want x), whereas the statement that ‘x is desirable’ is a normative claim (i.e. that x is *worthy of desire*, or *good*). Thus, the inference appears to be that happiness is desirable, or good, *because* we desire it. If this inference were sound, Mill would have successfully bridged the empirical and the normative. We might agree with Mill that one cannot explain why people desire happiness, so we must accept our desire for happiness as a datum.[72] But it doesn’t seem as obvious that happiness (or anything else for that matter) is *desirable* because it is *desired*. Mill uses an analogy with sound, that it’s because we can hear x that x is audible, and seemingly by extension, it’s because we desire x that x is desirable.[73] But it seems clear that ‘audible’ means *able to be heard*, so, quite sensibly, because we can hear x, x is *able to be heard*. By contrast ‘desirable’ means *worthy of desire,* not *able to be desired*. While it follows that because we desire x, x is *able to be desired*, it certainly does not follow that because we desire x, x is *worthy of desire*. It appears that Mill in this passage is deriving an ought from an is, inferring that we *ought to* desire happiness from the fact that we *happen to* desire it. Along these lines, G.E. Moore famously argues that Mill in this argument is blatantly committing the naturalistic fallacy.[74]

[72] I say we might agree with Mill despite the fact that I have hitherto been arguing against the claim that people should desire happiness. Nevertheless, I think, there might be a formal interpretation of happiness as the object of desire that could render this claim trivially true. One does not see a similar justification for the equivocation between that which people actually desire and that which they *ought* to.

[73] Mill, *Utilitarianism* chapter 4.3.1- 4.3.2.

[74] G.E. Moore criticizes Mill of making an “artless use of the naturalistic fallacy” in equivocating between what is desirable and that which people actually desire. Moore writes, “Mill has, then, smuggled in, under cover of the word ‘desirable,’ the very notion about which he ought to be quite clear. ‘Desirable’ does indeed mean ‘what it is good to desire’; but when this is understood, it is no longer plausible to say that our only test of *that*, is what is actually desired.” See GE Moore *Principa Ethica*, (section 39-40)

While on a surface reading of the text, it might appear Mill is committing such a fallacy, historical context allows a more charitable interpretation. As Geoffrey Sayre-McCord noted, Mill cannot intend that desirable means *able to be desired*, for the claim *happiness is able to be desired* is unhelpful to Mill's larger argument (i.e. showing that happiness is merely the sort of thing that *can be* desired goes no distance toward showing it *is* good).[75] So, we must make some sense of Mill claiming that people's desiring happiness indicates that it is indeed *worthy* of desire.

Here, it helps to remember that as one of the pillars of British empiricism, Mill is committed to the notion that we learn everything we know through direct observation. Indeed, he reiterates this conviction at the beginning of the passage in question:

> To be incapable of proof by reasoning is common to all first principles; to the first premises of our knowledge, as well as to those of our conduct. But the former, being matters of fact, may be the subject of a direct appeal to the faculties which judge of fact- namely, our senses, and our internal consciousness. Can an appeal be made to the same faculties on questions of practical ends? [76]

Mill's intended answer here is 'yes,' that our knowledge of practical ends can be similarly derived from our perception of ourselves and others. The fact, on such a view, that x is desired by all would be strong evidence that x is worthy of desire, since the only possible sources of evidence are one's internal consciousness and perception of others. When starting from such a commitment, Mill's original argument begins to more clearly resemble a feature of his empiricism and less the equivocation Moore uncharitably suggests. But even if this evidentiary standard does follow from an epistemic commitment, our preferences still should not be regarded as *definitive* evidence. Being

---

[75] Geoffrey Sayre-McCord, "Mill's 'Proof' of the Principle of Utility: A More than Half-Hearted Defense", Social Philosophy and Policy 18 (2):330 (2001), (especially 335-336)

[76] Mill, *Utilitarianism* 4.1.2- 4.1.7

restricted to empirical sources of knowledge simply reinforces the need to scrutinize our perceptions and regard our conclusions as fundamentally tentative.

In addition to his empiricism, however, we must remember that Mill is *not* endeavoring to show happiness *as it occurs in all people* is the highest human good. He is rather trying to identify a happiness that can only be expected to reliably occur in persons endowed with the right mix of luck and culturally felicitous circumstances. He explains that people might sometimes fail to choose the path of greater happiness (in his sense) because of a lapse of reason, even though they know an activity would inspire a greater pleasure.[77] Such people might fail to prefer worthy activities because they are not thinking clearly or perhaps because they don't allow themselves to consider the matter at all. Others, Mill reasons, might have been so beaten down by life, that they are no longer capable of experiencing the pleasure of their higher faculties. "Capacity for the nobler feelings," Mill writes, "is in most natures a very tender plant, easily killed, not only by hostile influences, but by mere want of sustenance."[78] Mill believes that through their commitment to particular lives or occupations people can "… lose their high aspirations as they lose their intellectual tastes, because they have not time or opportunity for indulging them."[79] In such a case, he believes a person might 'addict' herself to a lower pleasure because that's all she's capable of enjoying. Mill

---

[77] *Utilitarianism*, 2.7.1-14

[78] *Utilitarianism*, 2.7.15

[79] *Utilitarianism*, 2.7.18-20

relatedly suggests that many who try to preserve an ability to appreciate more worthy pursuits fail to do so and ineffectually break down.[80]

For this reason, it's no counter-example to Mill that a damaged or aberrant person might fail to find happiness in that which is good. As I discuss later in this chapter, it wouldn't be evidence against Mill's ultimate view that such failure were common to most or even all persons in a society that didn't provide the cultural and educational resources required for fully functioning moral agency. But again, the purpose here is not to find an argument against Mill's larger project. Rather, I'm considering his argument here to show how it plausibly works for Mill's purposes and, yet, that those purposes turn out to be decidedly inconsistent with utilitarianism.

Let us, then, adopt this more charitable reading of Mill and, following his empiricism, determine whether we generally observe a connection between happiness and goodness in non-aberrant and undamaged persons in an environment that *did* provide the requisite resources. For brevity's sake, I will refer to a person so situated as an 'well-developed person'. This is important for our overarching argument because we're trying to establish that the sources of a well-developed person's *actual* happiness are compatible with the her good. If harming innocents would maximize the net-happiness of well-developed persons in a society, the second principle of utilitarianism might require that such harm was morally good and this would weigh against the doctrine's credibility. On the other hand, the perfect congruity of such persons' happiness and our intuitions about what's good for her would be evidence in favor of the theory. The

[80] It isn't immediately clear why simply trying to enjoy both higher and lower pleasures would cause one to break down, especially as it's scarcely imaginable that one could live on a diet of exclusively higher pleasures. But I leave this issue to the side as it is peripheral to the present issue.

concern, of course, is not that such a person or society would find happiness in utterly repugnant action, but that well-developed persons might tend to find happiness in mundane, neutral, or mildly vulgar behavior. If such pedestrian activities were frequently the source of *more* happiness than the use of their intellectual faculties (and if they didn't cause any pain to others), the second principle would seem to problematically require that well-developed persons' vulgar leanings would constitute their highest good.

Mill begins chapter two of *Utilitarianism* by framing what he takes to be the doctrine's core thesis, then proceeds to flesh out the concept of happiness that utilitarianism promotes, and ultimately defends it against what he takes to be some prominent criticisms. The first of these criticisms is the issue before us; whether happiness really is our highest good. Mill memorably opens chapter two by defining the premise on which he takes utilitarianism to be founded,

> Utility, or the Greatest Happiness Principle, holds the actions are right in proportion as they tend to promote happiness, wrong as they tend to produce the reverse of happiness. By happiness is intended pleasure, and the absence of pain; by unhappiness, pain, and the privation of pleasure.[81]

Here, 'utility' is used interchangeably with 'happiness', which is associated with pleasure and opposed to pain. Utilitarianism, for Mill, seeks to maximize happiness/pleasure and minimize pain and the privation of pleasure. He anticipates that he'll need to say more about what constitutes pleasure and pain, but none of this, for Mill, changes the doctrine's essential thesis:

---

[81] *Utilitarianism* 2.2.1-2.2.5

> …that pleasure, and freedom from pain, are the only things desirable as ends; and that all desirable things… are desirable either for the pleasure inherent in themselves, or as a means to the promotion of a pleasure and the prevention of pain.[82]

At that point, Mill notes that critics have tended to bristle at utilitarianism's veneration of happiness as the highest end of human life. According to Mill, these critics subject the doctrine to a charge also leveled against early Epicureans, that forsaking more noble pursuits than happiness renders utilitarianism "… a doctrine only worthy of swine."[83] The worry expressed here is that our happiness is no more noble than its sources. If our happiness were the product of empathy, intellectual pursuits, and psychological temperance, utilitarianism would not be vulnerable to this criticism. If alternatively, our happiness were the product of base pleasures and treated the experience of these as our highest end, the critics concern would find a target. Mill counters that just as Epicureans recognized that because human beings have more refined capacities than lower animals, the sources of their happiness are more elevated and multifarious than those of lower animals. Thus, he believes, utilitarianism recognizes some forms of happiness are superior *in kind* to others. Here, Mill introduces his *quality of pleasures* doctrine.

He explains that pleasures admit, not only of different *quantities*, but different *qualities*. Although we generally prefer more pleasure to less, Mill sees that a smaller amount of a higher pleasure could produce a greater happiness (i.e. a happiness of a greater *quality*) than a larger quantity of a lesser pleasure. Mill's test for the existence of "higher pleasures" displays his characteristic empiricism. He writes,

---

[82] *Utilitarianism* 2.2.10 - 2.2.14

[83] *Utilitarianism* 2.3.5

> Of two pleasures, if there be one to which all or almost all who have experience of both give a decided preference, irrespective of any feeling of moral obligation to prefer it, that is the more desirable pleasure. If one of the two is… placed so far above the other that they prefer it, even though knowing it to be attended with a greater amount of discontent, and would not resign it for any quantity of the other pleasure which their nature is capable of, we are justified in ascribing to the preferred enjoyment a superiority in quality…[84]

Although, for example, we find a particular pastry pleasurable to consume, we might agree that no amount of this pastry could possibly outweigh the pleasure we derive from our children, even if they are a source of greater discontent. So long as well-developed persons acquainted with both shared this view, Mill would contend that the happiness we derive from our children is of a *greater quality* than that which we derive from the pastry. Thus, Mill's doctrine suggests, while it might *seem* that the greatest sources of a well-developed person's happiness don't contribute to her good, we might be focusing the *quantity* of contentment they produce, as opposed to the *quality* of the happiness. At this point the claim is just that if well-developed persons prefer a smaller amount of given pleasure (even though it brings more discontentment) to another, the former is of a higher quality. The question, then, is whether the sources of a well-developed person's *higher quality* pleasures are more clearly consistent with her good than the baser sources of her lower pleasures. *Prima facia*, this notion seems promising, but it will depend on which these higher pleasures turn out to be.

Following the quality of pleasures test, Mill explains that pleasures of greater quality are produced when we use our higher faculties (hereafter the *higher faculties* assertion):

> Now it is an unquestionable fact that those who are equally acquainted with and equally capable of appreciating and enjoying both, do give a most marked preference to the manner of existence which employs their higher faculties.[85]

---

[84] *Utilitarianism*, 2.5.4

[85] *Utilitarianism*, 2.6.1

The strategy here is clear; associating higher pleasures with the higher faculties is intended to connect our higher pleasures and that which is good for us. The first part of the *higher faculties* assertion is certainly helpful and reflects Mill's concern is with well-developed persons. The requirement that those subject to the quality of pleasures test must be *equally acquainted with and equally capable of appreciating and enjoying both* would remove two groups who might otherwise be problematic; (1) duffers, who might sincerely believe a life focused on the lower faculties might yield greater happiness in part because they're not greatly endowed with higher faculties and (2) those with higher faculties but stricken with some malady that prevents their enjoyment of them (i.e. the mathematician who, due to great depression, cannot see any value in his groundbreaking work and the brilliant physicist who would otherwise enjoy his research were it not for the intense migraines he suffered whenever he concentrates).

The second part, that all those sufficiently acquainted *'…*give a most marked preference to *the manner of existence* which employs the higher faculties,' subtly clarifies the range of behavior Mill is concerned with. He is *not* here referring to singular actions, but entire ways of life. This gives him the flexibility to hold that even if particular actions produce far more happiness than others (e.g. a leisurely day at the beach might produce more happiness that a day conducting a physics experiment), such actions are not to be treated as higher quality pleasures unless the *lives* of which they are a part produce more happiness (e.g. a life of leisure might produce a lower quality pleasure than a life of scientific research). Mill is careful to attach the individual act to the manner of existence of which it is a part, but the subject of his judgment is the entire life, not the isolated action.

A striking facet of the higher faculties assertion is that it associates the pleasure derived from a manner of existence to the faculties employed by that manner. In one way, this is astute, as two activities of similar physical extension might draw on very different internal resources. A life spend playing in the dirt might be extensionally indistinguishable from a life spent cultivating a garden, but this distinction allows us to see why the latter could be more meaningful (e.g. it involves planning and nurturing in a way the former does not).

However, the assertion that manners of existence involving higher mental function will generate higher quality pleasures, might miss other sources of higher pleasures. It might be that solving complex mathematical problems would require higher mental function than parenting children, even if the latter were a more meaningful experience. A well-developed person's most profound experiences demand less intellectual function than calculus, but it doesn't seem that the happiness they produce is lesser as a result.

This concern about pedestrian yet meaningful behavior reveals a bifurcation in Mill's discussion of quality of pleasures and higher faculties. There's a clear distinction between higher and lower pleasures, where intellectually rigorous pursuits are the paradigm of the former and bodily pleasures are that of the latter. But it seems that many manners of living are somewhere in between and one worries that these might not sort in the right way. Although life-defining, parenting children does not, for the most part, involve rigorous intellectual activity. Much of such a life is mundane and draws primarily on our emotional capacities and patience, but it doesn't strike us as being therefore less meaningful. Even if a well-developed person would clearly prefer an

intellectual life over one spent on bodily impulses, it's fair to ask if an intellectual life would bring her a greater quality of happiness than one of relatively routine parenting. The question, then, is whether the 'higher faculties' are limited to advanced cognitive functions or could include things like patience, kindness, and love.

Mill distinguishes the higher faculties as consisting of those things of which *swine*, *beasts*, *animals*, *fools*, and *ignoramuses* are incapable. He often characterizes higher faculties as "…pleasures of the intellect," but allows that these *might* include "…feelings and imagination, and… the moral sentiments."[86] It is, thus, a bit unclear where a life of parenting would fall. To be clear, our concern is not about whether swine and ignoramuses would choose a vulgar life, as Mill clearly excludes them from the category of well-developed persons with which he's concerned. They are only relevant inasmuch as Mill defines 'higher faculties' as *things of which which swine and ignoramuses are incapable*. If such base creatures *were* capable of feelings, imagination, and the moral sentiments, then these capacities would *not* qualify as higher faculties. Since higher faculties are what produce higher pleasures, parenting would therefore *not* qualify as a higher pleasure. On the other hand, if feelings, imagination, and the moral sentiments *were* higher faculties, this category (and by consequence the category of higher pleasures) could increase so much that it might not sort cases in the way Mill intends.

Here, Mill seems to be on the horns of a dilemma. If the higher faculties include feelings, imagination and moral sentiments, he might no longer be able to use the higher pleasures doctrine to distinguish chess from checkers, literature from pulp fiction,

[86] *Utilitarianism*, 2.4-5

and mathematics from idle conversation. For even checkers, romance novels, and gossip with friends engage these capacities. If, on the other hand, the higher faculties do not include things such as feelings, imagination, and emotional connection, then number of worthy human experiences might not qualify as higher pleasures. Enjoying a sunset, experiencing friendship, gardening, cooking, and a variety of other human connections and experiences that aren't intellectually rigorous activities still strike us as embodying that which is good for us, and do so in part *because* they stimulate our feelings, imagination, and moral sentiments.

Mill might want these capacities to count as *middle faculties*, more elevated than sensory pleasures but less than intellectual pursuits. But if higher pleasures were still to be preferred over middle pleasures, Mill would then have to regard a life of mathematics as a higher order of pleasure than lives of parenting or gardening, and, thus, someone capable of each but only practically able to experience one, should choose the life of mathematics. Perhaps, then, both middle and higher faculties should count as equally capable of producing higher pleasures (or that higher faculties should include the middle faculties). Mill needn't, after all, intend a proportionality requirement, where, as mental rigor increases, so too does the quality of pleasure derived. One might hold that a threshold must be met which would exclude merely corporeal behavior, but would include both lives committed to parenting and performing calculus. But in such a case, the same issue returns: Mill would not be able to distinguish deep friendship from surface acquaintances, the experience of a great novel from a trashy romance, or a comic strip from a moving work of art.

But perhaps this doesn't matter? As one considers the issue, there does indeed seem to be a sort of antiquated elitism in this line of reasoning, where rarified intellectual activities are treated as better for us than more pedestrian ones. Perhaps a moving work of art needn't generate higher pleasure than a pretty picture. Perhaps the pleb is capable of happiness as great as the savant, even if the dullard is not. Perhaps there's nothing wrong with opening the gates to Mill's higher faculties and allowing that more of our experience should qualify.

However, as we free ourselves from the notion that intellectual pursuits must be prized above all others, one wonders if the experiences that engage the middle faculties might actually produce *more* happiness (and thus be more closely connected to human goodness) than those that draw on the higher faculties. It might not be absurd to believe that a life of parenting, deep friendship, and connection with the world around oneself is actually more choice-worthy than a life dedicated to intellectual pursuits.

Continuing along this line, we seem to have the intuition that the very best life for a person overall, especially the sort of well-developed person Mill envisages, would involve activities that draw on our higher, middle, *and* lower faculties in some proportion Some math, some parenting, a garden, friendships, and even a bit of bodily pleasure here or there might make for the best life. It's also not unthinkable that experiences utilizing multiple capacities generate a greater pleasure than those that use the higher faculties in isolation. Cartography draws on our intellectual capacities while hiking draws on physical ones. While both might produce happiness, orienteering might produce a greater happiness than either in isolation. Again, Mill's remarks about

happiness refer to *manners of experience* and not simply individual actions. But it seems plausible that a life integrating our different faculties might generate a higher quality of pleasure than one emphasizing the higher faculties. In the end, I think all this suggests that the sources of human goodness are so multifarious and their interactions are so complex that they cannot be sufficiently ranked by the faculties used to experience them.

As one rehearses the argument, one worries that the move from quality of pleasures to higher faculties simply pushes the recalcitrant data up a level. To assuage the apparent inconsistency between sources of our happiness and our good, we invoke the quality of pleasures doctrine, in the hope that the sources of our greater quality pleasures will be more consistent with our good. To generate this result, Mill specifies that we have a decided preference for our higher faculties, which means that our higher faculties will turn out to be the source of our higher pleasures. However, here the concern about happiness' elliptical arc becomes a concern. It isn't clear that we derive greater pleasure from increased use of these faculties.

At this point, it might be tempting to claim Mill fails in his effort to connect our actual experience of happiness to human goodness through the quality of pleasures doctrine. But, I think, we should resist this temptation and try to follow Sayre-McCord's more charitable interpretation in our reading of Mill. We should assume that Mill is not simply making questionable empirical claims about our preferences and rather be open to the possibility that something deeper is going on. If, as I've tried to show, it's difficult to connect our actual experience of happiness with human goodness, there appear to be

two clear routes to easing the tension, (1) changing the notion of happiness we're using or (2) more carefully circumscribing the group whose experience counts. Mill, I believe, does *both* and does so to an extent that allows him to easily underwrite the quality of pleasures doctrine, the higher faculties assertion, and the second principle of utilitarianism generally. The cost is just that the resulting view diverges significantly from the utilitarianism being discussed here.

With regard to the notion of happiness he's employing, Mill explicitly disavows the moment-to-moment subjective well-being ordinarily presupposed by utilitarians, almost ridiculing it when explaining that happiness is not an impossible aim:

> If by happiness be meant a continuity of highly pleasurable excitement, it is evident enough that this is impossible. The state of the exulted pleasure lasts only moments, or in some cases, and with some intermissions, hours or days, and is the occasional brilliant flash of enjoyment, not its permanent and steady flame. Of this the philosophers who have taught themselves that happiness is the end of life were as fully aware as those who taunt them. The happiness which they meant was not the life of rapture; but moments of such, in an existence made up of few and transitory pains, many and various pleasures, with the decided predominance of the active over the passive, and having as the foundation of the whole, not to expect more from life than it is capable of bestowing. A life thus composed, to those who have been fortunate enough to obtain it, has always appeared worthy of the name happiness.[87]

It's hard sufficiently emphasize what a departure this is from the notion of happiness on which classical utilitarianism relies. Instead of treating it as a linear phenomenon that increases without necessary limit, Mill sees that happiness is subject that limits and patterns. Moments of pleasure will inevitably be fleeting and one who experiences more intense pleasures will be likely to have the capacity for more intense pains.[88] Mill also thinks some tempos of life are preferable to others. That an active life is 'happier'

[87] Utilitarianism, 2.12.12-25

[88] Utilitarianism 2.6.14-17

in his sense than a passive one, *even if they include equivalent pleasures*, indicates that, for Mill, pleasure is *not* wholly constitutive of happiness.

Another departure from classical utilitarianism that's already been mentioned appears here as well: Mill's notion that happiness is a property that attaches to *lives,* not merely moments or actions. Rather than encouraging a person to maximize moment-to-moment pleasure or happiness, Mill's concern is with the entire structure of a life. Moreover, there is in Mill a notable absence of language encouraging the *maximization* of a life's happiness or the necessity of *positive balance* of pleasures and pains. He encourages the temperance to *not* expect more happiness of this life than is reasonable. He explicitly considers how people who believe that happiness is the end of life could be satisfied with "…such a modest share of it."[89] His answer is that satisfaction with the happiness in a life is determined *not* merely by the ratio of pleasurable to painful actions, but also by larger commitments that shape one's experience. A *tranquil* life with relatively little pleasure and an *exciting* life with relatively more pain might each be satisfying to live, according to Mill, so long as the choosing agent understood that the contours of her choice would delimit the the pleasure/pain ratio available to her.[90] This, again, indicates that happiness involves more than simply a balance of pleasures and pains for Mill. But, more generally, the concern with a life's balance of *tranquility*, *excitement*, and *activity* indicates an interest in the arc of a person's experience that exceeds that of standard utilitarianism. In these passages it seems that happiness, for Mill, is an achievement dependent on a series of

[89] Utilitarianism, 2.13.1-3

[90] Utilitarianism, 2.13.6-8

commitments and aspirations that attach to a life or meaningful parts of it (e.g. a meaningful career or a good marriage).

It's not merely that this notion of happiness is robust, it also places demands on the kind of person capable of experiencing it. As Mill fleshes out this conception of happiness, it becomes clear it presupposes not only a concept of person, but also of a citizen and a just society. It's a notion that makes happiness a moral and political achievement, one that closely connects human goodness, and one that may be unavailable to many. This relates to our present discussion because it constrains the set of persons whose pleasure is weighed in the quality of pleasures doctrine and the higher faculties assertion and, indeed, these constraints come to light when Mill discusses what's required for someone to be a 'competent judge' of pleasures.

One will recall that a pleasure is more desirable than another if *all or almost all well-developed persons who have experienced* both prefer one. But for a pleasure to be of a higher quality, Mill specifies that it must be those who are *competently acquainted* with both who would choose it even if it were accompanied by a greater discontent.[91] To be competently acquainted, or what Mill later describes as a *competent judge*, demands more than one might initially suppose.[92] Competent judges, for example, must develop both *public* and *private affections* assuring that they do not fall into excessive selfishness, which Mill believes to be the principle cause of dissatisfaction in life. Here we can see this notion of happiness imposing requirements on those who can be happy,

---

[91] Utilitarianism 2.5.3-11

[92] Utilitarianism 2.8.1

and therefore on those would be deemed competent judges of pleasures.[93] Private affections *require* the development of empathy and a network of related emotional capacities. Public affections and sincere interest in the public good *require* a concern for justice and the acceptance of certain duties of citizenship in a political community. The development of private affections requires an environment to nurture emotional development and the exercise of public affections requires a society with at least minimal justice and a political education.

In addition, happiness and competent judgment require, for Mill, a relatively broad mental cultivation. This needn't be philosophical acumen, he adds, but to be a competent judge of pleasures and intellect must be able to:

> …find sources of inexhaustible interest in all that surrounds it; in the objects of nature, the achievements of art, the imaginations of poetry, the incidents of history, the ways of mankind past and present, and their prospects in the future.[94]

Such a person would find beauty in world around her, would have interests in literature, history, and politics, and possess a genuine concern for the progress of human beings and the world at large. Endowing people generally with such attributes would require a meaningful system of socialization and formal education.

Lastly, for Mill, competent judges, must also be *self-conscious* and *self-observant* so that they may routinely monitor their own dispositions and propensities, assuring that they don't fall into the poor habits that can cause an otherwise high-functioning person to succumb to lower pleasures.[95] The self-consciousness to which Mill refers includes a

[93] Utilitarianism 2.13.18-19, 2.14.7-10

[94] 2.13.29-32

[95] 2.10.8

awareness that a person exists as one-among-others, subject to duties and responsibilities that follow from morality and social cooperation. Self-observation involves the practice of employing self-consciousness to impartially evaluate and correct one's own behavior. Together, self-consciousness and self-observation allow the intellectual and emotional qualities detailed above become manifest in a person's own thought and action.

The picture we're given is of a person who is (1) unmistakably moral and possesses the empathy needed for personal attachments and a commitment to public welfare, (2) emotionally developed enough to be actively involved in human affairs and living deliberately, (3) educated and intellectually curious, and (4) sufficiently self-aware to continually evolve and remain committed a moral life. The existence of such a person would, by itself, be a certain kind of political and cultural achievement, as these traits presuppose substantial cultural and political resources. Aware that these requirements are substantial, Mill nevertheless maintains that such mental cultivation should justly be "…the inheritance of everyone born in a civilized country."[96] This is much of what leads him to develop such deep criticisms of the educational and political institutions of his day.[97] Indeed, Mill believes that "imperfect social institutions" can prevent anyone in that society from developing into a competent judge.[98] It will follow that many will be incapable of experiencing Mill's happiness simply because they weren't afforded felicitous social conditions.

---

[96] 2.14.2-4

[97] 2.12.28

[98] 2.14.37-38

This entire notion, of course, is inconsistent with a surface reading of Mill's initial statements on happiness, that "[b]y happiness is intended pleasure, and the absence of pain; by unhappiness, pain, and the privation of pleasure…"[99] and his subsequent conclusion, "…that pleasure, and freedom from pain, are the only things desirable as ends; and that all desirable things are desirable for the pleasure inherent in themselves, or as a means to the promotion of pleasure and the prevention of pain."[100] On the more robust interpretation, happiness is the product of a myriad of intersecting factors involving the shape and tempo of a life and the development of variety of personal capacities and social institutions, not merely pleasure and pain. Elsewhere, Mill holds that equality and liberty are both inherently valuable, which similarly indicates that he doesn't believe that pleasure and pain are the only things with inherent value.[101]

There are undoubtedly resources in Mill to support a narrow interpretation of happiness as a function of pleasure and pain, but there are equally many that indicate a richer concept of happiness at work. Such resources allow us to better understand Mill's claim that whoever doubts a person of higher faculties (despite requiring more to make him happy) is not still happier than inferior one, "…confounds the two very different ideas, of happiness, and content."[102] Contentment for Mill is attainable by dullards and beasts as a simple function of pleasure and pain. But happiness develops

---

[99] Utilitarianism 2.2.3-5

[100] Utilitarianism 2.2.10-14

[101] For Mill's discussion of equality's inherent value, see "Was Mill a Utilitarian?", Christopher Miles Coope, Utilitas, Volume 10, Issue 01, March 1998, pp 33-67, especially pp. 36-37. For a discussion Mill's attempt to harmonize the inherent value of liberty with some version of Utilitarianism, see "Mill's Deliberative Utilitarianism" David O. Brink, Philosophy & Public Affairs, Vol. 21, No. 1 (Winter, 1992), pp. 67-103

[102] Utilitarianism 2.6.33-34

for Mill into a more robust notion, one that is properly moral precisely *because* it includes so much more than than a favorable balance of pleasure to pain.

If we reconsider Mill's quality of pleasures doctrine and higher faculties assertion in light of this more robust notion of happiness and the corresponding requirements for competent judgment, both would have a far better chance of working for his purpose. A competent judge, so described, would be more likely to prefer using the higher faculties and to deem certain satisfactions incomparably higher than others. Such a person would plausibly value friendship, liberty, and justice more than any amount of corporeal pleasures. But the point is *not* simply that when we shift our attention from people generally as we find them to well-developed people that they're likely to prefer the higher faculties.

It rather seems that adopting this expansive interpretation of happiness and competent judgment largely *obviates* the need for the quality of pleasures doctrine in the first instance. After all, the doctrine was supposed to ease a tension between happiness and human goodness, in order to connect a subjective experience with an objective state. But when we're no longer using pleasure as our proxy and instead treat happiness as the product of a life of deliberation and temperance, happiness ceases to be an exclusively, or even primarily, subjectively experienced state. What Mill instead gives us is a thoroughly developed *objective* notion of happiness that's predicated on

the shape of a life, the development of a variety of capacities, and a society that can support both.[103]

If we used this notion to establish the second principle of utilitarianism, we'd be trying to demonstrate that a largely objective state that we call 'happiness', connects with another objective notion, human goodness. This connection is much easier than that of subjective happiness with goodness, not merely because it isn't subject to variability among persons in the way our traditional subjective notion is, but because some non-trivial kernels of goodness are already built into this notion of objective happiness. The fact that this notion of happiness requires empathy, public affections, and the incorporation of the general good into one's own self-regarding preferences makes it far more likely that these sources of 'happiness' will be consistent with morality generally. In one of the more frequently cited passages in *Utilitarianism*, Mill goes so far as to suggest that a society should condition its citizens to find their own happiness inexorably intertwined with moral goodness generally. He writes,

> …utility would enjoin, first, that laws and social arrangements should place the happiness, or… the interest, of every individual, as nearly as possible in harmony with the interest of the whole; and secondly, that education and opinion, which have so vast a power over human character, should so use that power as to establish in the mind of every individual an indissoluble association between his own happiness and the good of the whole… so that not only he may be unable to conceive the possibility of happiness to himself, consistently with conduct opposed to the general good, but also that a direct impulse to promote the general good may be in every individual one of the habitual

[103] David Brink also believes that Mill has an objective notion of happiness and even of Mill's higher pleasures doctrine. (See Brink pp. 69-76) His notion differs from the account I provide above in a number of respects, but importantly when he locates the objectivity in Mill's higher pleasures doctrine in Mill's simply valuing certain "intellectual" mental states as being intrinsically more valuable than others (pg. 76). My view is that the objectivity in Mill's happiness isn't so blatant as to follow from simply from valuing more intellectual activity to less, but rather that Mill's concept of a worthwhile person and a life include substantial normative notions that make happiness and success in life an objective state. I believe these notions are not limited to simply valuing more intellectual activity to less, but rather extend to certain tempos and shapes of life more than others. On my account, for example, Mill might reasonably believe certain meaningful, but intellectually undemanding, activity is importantly preferable to some intellectually rigorous activity, whereas I believe this is not possible on Brink's interpretation.

> motives of action, and the sentiments connected therewith may fill a large and prominent place in every human beings sentient existence.[104]

If every person's happiness were dependent on promoting the good of all people, Mill would be able to connect happiness and goodness trivially. In fairness, Mill in this passage is referring to the "…ideal perfection of utilitarian morality" and the basis of the doctrine needn't depend on such a connection. Nevertheless, we see the seeds of such a vision in Mill's requirements for competent judgment of happiness.

We should again resist the temptation to accuse Mill of circularity here not only because it's uncharitable but also because it misses the real lesson taken from this moment in Mill's work.[105] It's uncharitable because it's likely that Mill is not simply trying to sneak moral qualities into a substantial connection between subjective happiness and human goodness. If indeed he intends the more robust, objective interpretation of happiness, he would likely hold that the connection between happiness and moral goodness is formal, or almost so. This objective notion differs enough from our ordinary usage that one who adopted it would be best served by treating 'happiness' as a term of art. We're not invoking that notion when we say that we smile when we're *happy* or that all people naturally seek *happiness*, or when we associate *happiness* with a wide variety of pleasures from lofty to base. Forcing an objective notion of happiness (as the flourishing that results from a life lived well) onto our *ordinary* use of 'happiness' would

---

[104] *Utilitarianism* 2.18.10-23

[105] The temptation is to believe that the competent judges requirement might be circular. If it's the fact *these* things are worthy of desire that makes those who prefer them competent judges, the proof that they are indeed worthy of desire, cannot be that competent judges prefer them. Or alternatively, if a prerequisite of a being competent judge is that one agrees that pleasures employing the higher faculties generate a greater happiness, then proof that the use of higher faculties does indeed generate greater happiness cannot be that competent judges say so.

inevitably result in the same shoehorning issues we've seen to this point. Mill might, then, intend 'happiness' as a term of art, in which case the charge of circularity would miss.

But leaving aside the exegetical issue of Mill's own intent, a largely objective notion of happiness that provided a formal connection to goodness would leave the strain of utilitarianism we've been considering in the following bind: while it would assure the second principle of succeeds trivially, it would virtually doom the first principle. If (objective) happiness contains essential elements of human goodness, then the claim that *satisfying a person's informed preferences, increases her (objective) happiness* would strain plausibility unless exceedingly robust constraints are built into the *informed* condition (which, as we have seen, is problematic). If one continued pushing the worry up the inferential chain, holding, for example, that the only informed preferences were those that promoted good generally, the first principle might survive. But at the point, one would have so throughly latticed the desired normative conclusion through the premises that circularity would then be a worry and the argument would appear pointless. The lesson to be drawn, I think, from utilizing Mill's arguments to defend utilitarianism is that there seems to be an unavoidable tension that arises in associating a subjective notion of happiness as it applies to people with an objective notion of human goodness; and it's not clear that any level of complexity can resolve this.

But even apart from this, for utilitarianism, a nagging issue remains. The modern data about happiness suggests that a person's hedonic arc doesn't ascend linearly as her activity becomes more rigorous. The curve is frustratingly elliptical, where we seem

to want what is just out of reach and we tend to cease desiring it once it is acquired. The concern this data presents is twofold. First, our happiness over the course of our lives tends not to rise in the way that happiness maximization, or even *sustained happiness improvement,* would require. Second, there's no data to suggest that among people generally, temporary increases in happiness follow from more sophisticated activities, or one's good generally. The data does indicate that boredom is something people try to escape through activity, but successful means of doing this may equally be elevated or mundane.

I think we often draw the wrong lesson from the many critiques of utilitarianism that have appeared in the last fifty years. We tend to conclude that the problem lies in the preferences we allow to count or the class of persons we're measuring. But the culprit might have been so obvious as to have escaped notice. When connecting preference satisfaction to happiness and happiness to goodness, the middle term might be that which fails us. Our happiness is fickle, often shallow, and ultimately caught in an interminable cycle. It neither tracks the value we assign to events in our lives nor the trajectory of virtue. Whether we're wicked or noble, wealthy or poor, we will all be both happy and sad, and mostly somewhere in between. Because goodness and happiness have clearly different arcs, showing that the greatest happiness actually does coincide with a robustly normative conception of the good will inevitably involve certain idealizations: either one must idealize happiness as we experience it or idealize the persons for whom this connection can be made.

This isn't a concern specific to Mill or even one confined to utilitarianism. It is rather an issue for any theory, moral or other, that treats a subjective notion of happiness as a singular end or as unreservedly good. Perhaps the lesson should *not* be that we need to severely restrict which, or whose, preferences should count. It might instead be that we should give up on happiness as an action-guiding motive and an unambiguous moral predicate. It might be no more obvious that happiness is the highest good than it's obvious that a rational person will always prefer that which makes her happiest. In fact, wouldn't it be strange if a person didn't always prefer the path of greater happiness because she somehow *knew* that happiness wasn't the highest good for her; that instead she was guided in her life-structuring decisions by another motive which, however inchoate, aimed at something finer and that she wanted far more?

Chapter Three
Kant: Moral Law, Disclosed by Rationality

In the last chapter, I tried to show that assigning happiness a central role in moral theory is difficult in part because its trajectory differs from that of human goodness. In the present chapter, I explore the less central role happiness has in Kant's ethics. I'll provide some exegesis of the core elements of Kant's ethics and then I'll consider various avenues through which happiness might be expected to enter his moral theory, his *imperfect duties*, in particular. I'll argue that even in an expansive interpretation, happiness plays a minor role in Kant's ethics because he is, more generally, trying to wean us from the tendency to treat happiness as a reliable metric of goodness. I think this discussion will depict Kant's ethics as a paradigmatic example of a morality relatively unconcerned with happiness, where, by consequence, moral obligation can stand in tension with it. Though generally Kant's ethics are, for this reason, treated as austere, I think the aforementioned difficulties that arise when happiness is central to a moral theory will help assuage this initial impression of Kant's ethics. In fact, in this light I think Kant's ethics leave a promising space that might be filled by something more deserving of moral protection than happiness. My hope is that reviewing these features of Kant's ethics puts us in a better position to see what that thing is.

One of the hallmarks of Kant's philosophy is his lifelong effort to identify veritable foundations for our knowledge.[106] In his epistemology, this takes the form of locating the limits of our possible knowledge and revealing the structure to which all of our experience must conform. Though he believes that our knowledge is necessarily limited and that we can make no claim to understand things-in-themselves, he painstakingly distinguishes between the content and form of our thought as he explores the limits of what we can know.

In his ethics Kant is arguably more successful in finding an indubitable foundation for our understanding. Whereas the actual structure of the world need not conform to the limits of human understanding, the ethical question Kant concerns himself with, how ought *we* behave, is intrinsically anchored to *our* nature as persons. The *facts* of ethics, so far as Kant is concerned, are determined by distinctive features that constitute us as persons and, unlike the actual structure of the external world, we have direct access to these.[107] In this way, Kant avails himself of a richer set of premises in his ethics than he allows himself in the first *Critique*.

For Kant, the question of how we ought to act is determined by our rational nature.[108] The practical reason that governs our daily lives consists of numerous conditional reasons, or hypothetical imperatives; if I want x, then I ought to perform

---

[106] All page references to Kant's works in ethics are taken from the *Practical Philosophy, The Cambridge Edition of the Works of Immanuel Kant*, edited by Mary J. Gregor. As Paul Guyer and Allen Wood note in the volume's general editors' preface (pg. viii), this text maintains pagination in the standard German edition of Kant's works, *Gesammelte Schriften* by means of marginal numbers. In accordance with the prevailing convention, I adopt this practice below. References beginning with a '4' refer to Kant's *Groundwork of the Metaphysics of Morals*, references beginning with a '5' refer to Kant's *Critique of Practical Reason*, and references beginning with a '6' refer to Kant's *Metaphysics of Morals*.

[107] 4:452

[108] 4:412

action Y. But such a system of reasons is worryingly incomplete, since such imperatives could issue from an indiscriminately large number of antecedent conditions, some of which might issue from our appetites.[109] If the sole purpose of reason is to satisfy our desires and impulses, a system of hypothetical imperatives might merely serve to connect our unthought impulses to the rational means that will satisfy them. One of the aims of Kant's ethics is to refute Hume's notion that reason is slave to the passions and if hypothetical imperatives accounted for all of our deliberation, Hume's view would seem stronger. In such as case, the role played by reason in our moral agency would be quite small indeed.

In wanting to find a deeper foundation for the reasons that should motivate us to act, Kant's ambition was to locate a categorical imperative, or a demand that issues on us all, irrespective of desire or inclination.[110] If such a principle could engage us at a level prior to our various desires, in virtue of some defining property we share as persons, it could escape the capriciousness of our hypothetical reasons.[111] The first formulation of Kant's categorical imperative, that one should *act only in accordance with that maxim through which one can at the same time will that it become a universal law*, aims to do precisely this.[112] When one imagines oneself before any contingent characteristics or dispositions have been assigned, when one could literally be any person, the principle's justification becomes clear.[113] The only permissible maxim is that

[109] 4:414-416

[110] 4:416, 4:420

[111] 4:389-390, 4:392

[112] 4:421

[113] 4:454

which all or any of us of us could intelligibly will, and the test is to imagine that the maxim is willed universally. The lie, the theft, and the assault are unacceptable actions because they follow from maxims that are only intelligible when they are *not* universally willed.[114] These attempts to treat ourselves as unrestricted by the maxims that we presuppose in others indicates a failure to respect the autonomy and equality among persons and, for Kant, this is morally unacceptable.[115] Kant's categorical imperative procedure, thus, generates a class of duties that follow from our status as free and equal moral agents, irrespective of our various desires or dispositions.[116]

If the general thesis of Kant's ethics is that the realm of permissible action is defined by what we can will in accordance with universal law, then a persistent theme of his moral philosophy is that *our exercise of rationality* will disclose duties that bear on us in virtue of our humanity alone. There is little concession in Kant to what we contingently desire or want, as our desires don't endow particular courses of action with greater legitimacy than they would otherwise have. Rather, the broadest contour of his

---

[114] 4:402-403, 4:422-423.

[115] In fact, Kant offers a far more nuanced discussion of problems that arise when trying to universally will particular maxims than my discussion here reflects. Some maxims, such as those required for a false promise, are incoherent because they presuppose a convention (namely, that promises are issued with the expectation that they be upheld) that couldn't plausibly exist if the maxim involved in making a false promise were universally willed (for in such a case, persons could not trust one another enough for any convention of promising to take hold). Kant notes that other maxims, such as neglecting one's talents and dedicating one's life to mere amusement, could be imagined to apply universally without contradiction or incoherence. In such a case, the world would simply consist of indolent persons (a circumstance that Kant regrettably associates with 'the South Sea Islanders'). The problem with such a maxim, however, is that actually *willing* it is inconsistent with a person's nature as a rational agent, for "... as a rational being he necessarily wills that all the capacities in him be developed, since they serve him and are given to him for all sorts of purposes." In this way, the falsifying criteria for the first formulation of Kant's Categorical Imperative, the Formula of Universal Law (FUL), are more complicated than this discussion about it suggests. I gloss over these important details because my intention in discussing Kant's Categorical Imperative is merely to establish that his principle method of identifying moral duty attaches to rational properties of moral agents, irrespective of any dispositions or preferences they might contingently possess. (4:422-423)

[116] 4:438

morality is characterized by duties that attach to our status as persons, where our moral obligation is not merely to act in a certain way, but to act *because* we recognize that moral law commands us to.[117]

Kant's insistence that the only intrinsically good thing is a good will reflects this broader conviction that the determining criterion for moral action is *intentional* more than it is *extensional*.[118] He writes:

> [T]he moral worth of an action does not lie in the effect expected from it... For, all these effects (agreeableness of one's condition, indeed even promotion of others' happiness) could also have been brought about by other causes.[119]

It is maxims that are permissible or impermissible; the actions that follow from them inherit their status from the intention than drives them.[120] One and the same physical act may be impermissible, neutral, or praiseworthy depending on its motivating intention. For Kant, the 'moral moment,' so to speak, has occurred once the intention has been fixed and this *precedes* the performance of the action.[121] This distinguishes him from consequentialists according to whom the moral worth of an action isn't determined until *after* its performance, when its effects are felt. Though Kant's ethics are principally concerned with action, his concern is more profoundly rooted in the intentions which drive these actions, and specifically with our ability to allow ourselves to be guided by moral law as disclosed by our rational minds.

---

[117] 4:437 Kant writes here that it is not sufficient to act "...*in conformity with duty* but not *from duty*."

[118] 4:393-394

[119] 4:401

[120] 6:225, 4:399-400

[121] 4:410, "Thus the moral worth of an action does not lie in the expected effect from it and so too does not lie in any principle of action that needs to borrow its motive from this expected effect."

There is little concession in Kant's ethics to the notion that the morally correct action should *feel* right. He famously holds that beneficial actions motivated by a pleasant disposition or an interest in happiness are morally neutral at best.[122] The person moved to charity because it makes her happy isn't praiseworthy for Kant, because to the extent she is guided by her interest in happiness, she might have performed terrible actions if those things made her happy. For Kant, the difficulty lies not with the inclination itself so much as with its *contingency*. Contrasting these are actions motivated by the recognition of duty and it is only such actions, for Kant, that can be truly praiseworthy.[123] The crucial difference is that the duties required by moral law don't depend on an agent's inclinations.[124]

Nevertheless, it caricatures Kant's view to claim that he believes taking joy in an act removes its moral worth. Rather, he believes that one can derive pleasure from a praiseworthy act and even from doing one's moral duty, so long as the principle (or maxim) motivating the action is not to produce pleasure but to do what duty commands.[125] The crucial question for Kant is what motive guides our action. "[I]n the case of what is to be morally good," Kant writes, "it is not enough that it *conform* with moral law but it must also be done *for the sake of the law*..."[126] So long as we're guided by moral law alone, our actions are equally praiseworthy whether our inclinations

---

[122] 4:398

[123] 4:398

[124] Here I am referring to Kant's *perfect duties*. As I note below some scholarly disagreement exists over whether a lack of inclination might affect our obligation to perform our *imperfect duties* in particular cases.

[125] 6:377, 4:399

[126] 4:390, *editor's* italics (Mary Gregor)

encourage us or nag us to act otherwise. In this regard, it's not so much that Kant's duty is hostile to the notion of happiness as it is indifferent to it.[127]

But in the prototypical case, Kant sees our inclinations as seductions to be overcome, or at the very least restrained by reason.[128] This is, in part, because Kant believes 'happiness' in common parlance is little more than a general term we assign the satisfaction of our inclinations, and that it determines nothing about their content.[129] He holds that the conditions that would satisfy these inclinations vary considerably from person to person and from time to time. He pointedly notes that even if the conditions of our happiness didn't vary among different people or over time, they would still not be adequate grounds for moral obligation because they are empirical and not sufficiently connected to our nature as persons.[130] This is a persistent theme regarding our subjectivity in Kant's work: that our inclinations, desires, and perceptions are too varied and unsystematic to reliably guide us. Whatever truth we're capable of understanding lies in the world outside of us and our tool for accessing this is our rationality, not our direct perception.

Interestingly, Kant seems not to have always held that our subjectivity was cut off from our objective experience of morality. Very early in Kant's career (around 1762) his notes make reference to a *direct apprehension of goodness* and this appears to have

---

[127] See "On the Value of Acting from the Motive of Duty," pg. 21, in *The Practice of Moral Judgment*, Barbara Herman. "One need not be indifferent to the possible satisfactions that a dutiful action many produce. It is just that the presence of such possibilities should not be the ground of the agent's commitment to acting morally. Overdetermined actions can have moral worth so long as the moral motive is the determining ground of action- the motive on which the agent acts."

[128] 4:396, 4:399

[129] 4:399, 5:26

[130] 5:26-27

been the influence of Kant's familiarity with the moral sense theory of Francis Hutcheson.[131] The notion is that we directly experience the necessity of our duty and the goodness of our ends. Kant appears to entertain the notion that these feelings might "...contain the foundations for all other practical principles...,"[132] and that a metaphysics of morals might be applied to an empirical system of happiness in persons.[133] If successful, such a system might ground duty in a feeling of goodness persons directly experienced. We would, in such a case, reliably feel pleasure when performing our duty and experience joy when securing our ends. This would connect our apprehension of objective moral law with a particular subjective experience and provide a link between a metaphysics of morals and a normative psychology.

However, as Allen Wood notes, these remarks coincide with Kant's 'silent decade.' When he emerges from this period to write the mature works that contain his moral philosophy, Kant's discussion of a reliable moral sense is gone. It is replaced by Kant grounding moral worth in the freedom of the will and rejecting empirical principles as being possible grounds for moral law due to their unsystematic nature.[134]

This ground becomes perhaps the most distinctive aspect of Kant's ethics. Morality is not to be felt or intuited, nor is it to be pursued as a mechanism that unifies our subjective experience with objective obligation. Instead, Kant tells us that our rational nature offers us a chance to recognize what is required of us in virtue of our

---

[131] *Practical Philosophy*, T*he Cambridge Edition of the Works of Immanuel Kant,* general introduction by Allen Wood, pg xiv-xvi.

[132] Wood, pg. xiv, Kant, 2:300

[133] Wood, pg. xv

[134] 4:442

constitution alone, which connects with Kant's discussion our will's autonomy. Our will isn't free in the sense that it's beyond the influence of any forces; it's rather free in the sense that it's through a person's will that one allows herself to be guided by reason. When guided by her passions or inclinations, a person is not free in the relevant sense. Kant notes that while reason is poorly suited to satisfy our inclinations, it's perfectly suited to produce a good will, and thus this must be "...the true vocation of reason..."[135] The process of revealing our duty by applying reason to our will is, according to Kant, a gift with which nature has uniquely endowed us and thus our essential purpose. Cultivating and maintaining a good will, which Kant takes to be the only intrinsically good thing, requires that we act from duty as disclosed by rational deliberation. Kant's freedom is, in large part, a freedom *from* happiness-seeking behavior. Rather than accommodating our intuition that we should be happy, Kant tries to persuade us that we shouldn't allow ourselves to be led by something so capricious.

Imperfect Duties: A Different Happiness

While it is a central piece of his view, the categorical imperative does not constitute the whole of Kant's ethics. Although *this part* of his delineation of duties doesn't find a prominent role for happiness, Kant does locate a duty to promote the happiness of others in his *imperfect duties*. One might think this is the place where happiness can find an important place in Kant's ethics.

[135] 4:396

To be sure, Kant is *not* blind to the force imposed by our desire for happiness. He admits that being "...happy is necessarily the demand of every rational but finite being and therefore an unavoidable determining ground of its faculty of desire."[136] Kant is perfectly aware that our desire for happiness will be a persistent motive throughout our lives; he simply holds that the subjective and variable nature of our inclinations prevents them from grounding something as universal as moral law. But if this is true, we must understand how we could have an imperfect duty for the happiness of others. If the sources of happiness are too contingent to ground duty, one wonders what the substance of this obligation could be.

Kant's discussion of our imperfect duties is given in *The Metaphysics of Morals*, which is divided into two sections, *the doctrine of right* and *the doctrine of virtue*. The first of these is concerned with constraints on our action that follow from "the *formal* condition of outer freedom."[137] These are principally restrictions on various actions inconsistent with the categorical imperative. But Kant notes that ethics can extend beyond setting restrictions on our permissible ends and actually establish positive ends that we ought to pursue. *The doctrine of virtue* is concerned with establishing these positive ends that Kant hopes will counter our natural inclination toward hedonistic behavior.

The imperfect duties are Kant's "duties of virtue" and he initially defines virtue as simply the real opposite of vice.[138] Kant notes that while the ends we adopt are objects

---

[136] 5:25

[137] 6:379-380

[138] 6:384

of free choice, application of the categorical imperative shows that there are some ends we *ought* to set for ourselves.[139] When ends are prescribed by a law of morality, they are also duties, thus his duties of virtue. Kant holds that there are only two of these, the *duty for self-perfection* and the *duty for the happiness of others*.

These are *imperfect* duties because unlike Kant's perfect duties, these are limited in their range of application. In addition, whereas perfect duties are often of a *negative* sort (requiring that one refrain from engaging in a certain mode of behavior) the imperfect duties are *positive* in that they ask agents to actually perform certain actions. These ends, however, to promote one's self-perfection and the happiness of others, are so broad that their application could be almost limitless. Scholarly debate exists over the extent of these obligations, as a maximalist interpretation of these duties could easily absorb all of an agent's resources.[140] Some believe that our imperfect duties are so slight that we can choose when, and even if, we honor them.[141] Others hold that our duty for self-perfection and our duty for the happiness of others stand in a limiting relationship.[142] But without even investigating the scope of our duty to promote the happiness of others, it's fairly clear that Kant didn't intend this duty in a way that would place happiness near the heart of his moral theory.

---

[139] 6:382-384

[140] See Thomas Hill's "Kant on Imperfect Duty and Supererogation," from *Dignity and Practical Reason,* "...the rigoristic interpretation leads to the conclusion that there is virtually nothing morally indifferent, contrary to what Kant says," pg. 151

[141] See Hill, "…one may avoid [acting on an imperfect duty] at *any* time (though not at all times) that one feels inclined" pg. 159 (Hill's italics and parenthetical).

[142] Barbara Herman's "Obligation and Performance" *from The Practice of Moral Judgment (pp. 159-183)*

While we do have a duty to promote the happiness of others, this is *not* happiness in the ordinary sense discussed earlier. This duty follows from our obligation to take their permissible ends (and *only* these) as our own, and try to promote them for their own sake.[143] But just as Kant notes that pursuing our own permissible ends might not result in greater happiness, he admits that promoting another's permissible ends might increase their moral fitness, but *not* their happiness. Moreover, Kant leaves it to us to determine what would make another person actually happy, allowing that we might disagree with the person in question. He notes that if we disagree, we have no obligation to give a person that which she believes would increase her happiness, but are only obligated to provide for their "true needs."[144] Kant also states, "[t]he happiness of others includes their *moral well-being* and we have a duty, but only a negative one, to promote this."[145] We are thus to refrain from encouraging behavior in another that might later be the cause for her "inner-reproach" and assist her instead only toward ends we believe are compatible with her well-being. The happiness we're obligated to promote in others is thus narrowly circumscribed.

One might wonder why we don't have a duty to promote our *own* happiness or the perfection of *others*, and Kant's answer turns out to follow trivially. We don't have a duty for our *own* happiness, according to Kant, because we must inevitably will it and it would be contradictory to be required to will something we cannot avoid wanting. We don't have a duty for the perfection of *others* because, for Kant, a person's perfection is

---

[143] 6:388

[144] 6:388, 6:394

[145] 6:394

constituted by having *her* ends perfectly accord with *her* duties and thus it would be contradictory for me to be responsible to set someone else's ends.[146]

The remarks above suggest that our *duty for the happiness of others* is close to *a duty to promote the self-perfection of others*, but one that escapes the incoherence of such a duty. Because we are only obligated to promote (1) permissible ends that we believe would make this person happy and that express (2) her true needs, where (3) this includes a concern for her moral well-being, we are *not* simply obligated to satisfy her desires for happiness in an ordinary sense. We're *only* duty-bound to assist her in pursuing ends that we believe she would be morally permitted to will for herself. Since this class is circumscribed by her own duty for self-perfection, where this involves making her ends consistent with her duties, our promotion of her 'happiness' might consist in just assisting her pursuit of her permissible ends. Our *duty to promote the happiness of others*, thus, does not seem to be directed to improving their subjective well-being, or pleasure. Rather, this duty seems to be directed toward improving their *objective* well-being, which, as Kant admits, might be independent of their pleasure.

Kant's examples of promoting the happiness of others include instances of beneficence and helping the poor.[147] As discussed earlier, even for our new understanding of happiness, improving the objective circumstances of persons in poverty, chronic pain, or under existential threat does sustainably improve their happiness and objective well-being, unlike material improvements for those not so afflicted. Thus, in these cases, improving others' circumstances coincides with

[146] 6:386

[147] 6:448, 454, 457

improving their subjective well-being. But nowhere in *The Metaphysics of Morals* do we find Kant encouraging us to improve the subjective well-being of others where this does not simultaneously improve their objective well-being because he is centrally concerned with the latter. As such, those seeking a central place for our modern notion of happiness in Kant's ethics will not find it in our duty to promote the happiness of others.

One might expect that one's imperfect duty for self-perfection could serve as a grounding for happiness in Kant's ethics. We've already seen that one cannot have a duty for one's *own* happiness, since it would be incoherent to force someone to will that which she could not avoid willing. But to the extent that Kant believes that we must ever strive toward our own perfection, we might wonder if happiness is part of this perfection. More pointedly, since Kant repeatedly urges us that the role of reason and moral law is to check the influence of our baser inclinations, one might wonder if part of our self-perfection lies in inculcating this process so that our inclinations are harmonized with our recognition of moral law. One might think that such a person would be closer to perfect than one who continually struggled against wayward inclinations in her moral life.

For Kant, our duty of self-perfection is quite broad. It requires that we cultivate a variety of our faculties, including our capacity for benevolence, respect, and sympathy.[148] For Kant, however, "...[t]he greatest perfection of a human being is to do his duty *from duty*..."[149] The cultivation of one's will and moral feeling lie at the center of one's duty for self-perfection. Kant notes that for the "purest virtuous disposition... the

[148] 6:401, 6:402-403, 6:451-452

[149] 6:392, original italics

law becomes also the incentive to his actions that conform with duty."[150] Here, Kant is as close as he will be to including one's own happiness in her self-perfection. He notes that an element of perfection is actually feeling happiness at having done one's duty. He writes:

> When a thoughtful human being has overcome incentives to vice and is aware of having done his often bitter duty, he finds himself in a state that could well be called happiness, a state of contentment and peace of soul in which virtue is its own reward.[151]

But he immediately warns against this feeling being an acceptable motive for action.[152] For reasons already familiar, acting from a hedonic motive would strip an action of its moral worth and so a person motivated by the euphoria she experienced when performing good deeds would not be a moral agent at all. Moreover, even here, when Kant is urging that a nearly perfect person would feel happiness when doing her duty, one must note that he still concedes that doing one's duty is nevertheless "often bitter." This is noteworthy, because even here Kant does *not* go so far as to say that stomping out our non-moral inclinations is an element of our perfection. His view is rather that our inclinations can be expected to persistently oppose moral law and our self-perfection lies in developing a powerful will to overcome these desires. Kant goes so far as to claim that a person may confuse her lack of temptation with the virtue required to resist immoral action. "[H]ow many people who have lived long and guiltless lives may not be merely *fortunate* in having escaped so many temptations?" Kant asks.[153] One might even infer from this passage that the person who successfully

[150] 6:387

[151] 6:377

[152] 6:377

[153] 6:392, original italics

struggled against powerful inclinations is morally *fitter* than an agent whose lack of desire allowed her weaker will to direct her action, though Kant himself doesn't appear to take this view. It rather seems, for Kant, these two agents are morally equivalent because both are motivated to action by their good will, however easy or difficult their struggle was.

In either case, Kant's imperfect duties paint a moral landscape as starkly rational as his perfect duties. Moral action doesn't promote a subject's happiness and Kant allows that in the prototypical case the two often lie in tension. Morality is found in the development of a good will, which acts from duty as revealed by a person's rational nature. Her perfection lies not in stomping out "...the crude state of [her] nature"[154] but rather using her higher faculties to *overcome* it. Kant isn't trying to underwrite our intuition that we should be happy as much as he's trying to reorient us so that we want something else. Much of what makes Kant's ethics demanding is that he to refuses underwrite our intuition that our happiness is important and deserved. Instead, he asks us to reconsider what we want for ourselves and, by consequence, from a moral theory.

One might defend Kant's view by holding that he gives us a protected sphere where, after we've satisfied our moral obligations, we have a limited freedom to pursue our own ends. This, after all, cannot be said for versions of utilitarianism that leave us infinitely obligated to relieve the world's suffering. But we miss the point of Kant's ethics if we assume Kant would have us pursuing happiness when we are free of moral obligation. He rather wants us to embrace what is most fundamental in us and eschew

[154] 6:387

the vestiges of our animal nature, however urgent their claim feels.[155] This is the theme that connects his ethics with his epistemology; that the role of philosophy is to identify the limits of what we can know and, if possible, connect us with what's real. If we accept this is an important goal for moral theory and we can come to believe that our desire for happiness is in some sense mere appearance, we won't mind relinquishing our expectation that moral theories provide a space for it.

But if we don't start off with the concern that appearances can be deceptive and that the role of philosophy is to release us from their hold, Kant's picture won't get the traction it needs. In particular, if we believe that the objective world must somehow lie down with our subjective experience, that actions which *feel* good likely *are* good, and that our desires point us toward our proper ends, we will tend to prefer a moral theory that is more solicitous of our dispositions.

However, when we apply the new hedonic picture discussed in chapter one, that our happiness is cyclical and bound within fairly narrow limits, the Kantian picture is largely unaffected. It's as if Kant somehow anticipates such a revelation, since he always held that the sources of our happiness were too unsystematic to ground moral law. He also grants that when we aim at happiness we will tend to miss, in writing, "… the more a cultivated reason purposely occupies itself with the enjoyment of life and with happiness, so much the further does one get away from true satisfaction." He

[155] He grants that all persons will naturally seek happiness, but believes this inclination should be moderated by our rational nature. "There is, however, *one* end that can be presupposed as actual in the case of all rational beings (insofar as imperatives apply to them, namely as dependent beings), and therefore one purpose that they not merely *could* have but that we can safely presuppose they all actually *do have* by a natural necessity, and that purpose is *happiness*."

explains this is in part because the sources of happiness are empirical and it would be impossible to predict the conditions that would facilitate our future happiness:

> …the problem of determining surely and universally which action would promote the happiness of a rational being is completely insoluble, so that there can be no imperative with respect to it that would, in the strict sense, command him to do what would make him happy; for happiness is not an ideal of reason but of imagination.[156]

In some sense, the new picture of happiness appears to be an endorsement of Kant's urging that we cannot be guided by our fickle dispositions.

The relative indifference to our subjective experience that is traditionally treated as a deficiency of Kant's ethics, rather seems, in light of the new psychology of happiness, to be something of a virtue. If we accept this picture, or even some limited version of it, we should not treat the ability to underwrite our expectation of happiness as a criterion of adequacy of moral theories. Instead, such a picture prompts us to ask if we should be aiming at something more clearly anchored to a successful life than happiness and that perhaps our morality should be providing for whatever this is.

What to do with Happiness

How then should we view happiness, morally speaking? If Kant's ethics are to be praised for their relative indifference to our happiness and utilitarianism is doomed because of the central role it assigns happiness, one wonders if the conclusion is that we should stop being concerned with happiness altogether. If so, that seems problematic. For however strong the arguments are to ignore happiness, our impulse

---

[156] 4:71

toward it is irrepressible. Even if we were to discipline ourselves, one imagines that our desire for happiness would continue to make itself known.

But we needn't and shouldn't ignore happiness altogether, and not just because it would be impossible for us to do so. Here, we can make a helpful analogy to hunger. Satiating one's hunger will typically feel pleasant, but this can be done well or poorly and there's certainly no linear relationship between hunger and health. At the extremes, hunger is relevant, as both malnutrition and gluttony are bad for us. But in the middle spectrum, hunger itself is neither healthy nor unhealthy. Sometimes we should not satisfy our hunger and at other times we absolutely should. We have a difficult time ignoring our hunger because it is, like our desire for happiness, irrepressible. But the proper response is to manage it; not allowing our bodies to become so famished we gorge ourselves, but consuming just enough that our activity is not diminished. Satisfying our hunger shouldn't be regarded as an end, but as a necessary means for our activity. What we shouldn't do is treat our hunger as an unqualified indication of what's good for us.

Similarly, satiating our desire for happiness will feel pleasant, but this doesn't indicate whether an action is good or bad for us. At the extremes, happiness absolutely matters. If, for instance, our misery becomes so encompassing it eclipses our concern for others and we're unable to act well, it takes on a moral significance. But, conversely, if excessive contentment were to blind us to others' suffering or to discourage an even temperament, it too might be morally proscribed. Prolonged unhappiness might result in a bad life, but it isn't itself *constitutive* of a life going badly. Unhappiness is often

appropriate, even laudable, and happiness can be inappropriate and condemnable. They are, in this way, morally underdetermined predicates whose significance is fixed contextually. If unchecked, our hedonic balance can run to extremes and interfere with morality. But within ordinary, managed bounds, happiness and unhappiness needn't have any particular moral significance.

The result for utilitarianism is clear enough, but the effects are felt through moral theory generally. We will continue to have a persistent sense that we want happiness for ourselves and those we care about, and seeing the cyclical nature of our desire is unlikely to diminish this. But, where before we treated this impulse as a moral intuition, even a portal into human goodness, we might reclassify it now. Our desire for happiness might be as simple as other hungers we choose to satisfy or not. We rein in our desires for food and reproduction because we know their satisfaction is not always good for us, even though fundamentally both impulses have a purpose. Likewise, we might see our desire for happiness as something akin to an evolutionary vestige; vital once and even now at the extremes, but less significant within a broad spectrum of normalcy. We might have outgrown happiness in the same way we've outgrown other drives that once protected our flourishing. It's important that we don't starve and that we're not utterly miserable, but our lives should not be spent chasing bliss or gluttony. When our basic well-being has been secured, happiness has served its function. Perhaps then we should move on, for there are higher hills to climb.

Relegating happiness to this status undoubtedly strikes one as counterintuitive, but there are reasons for this reaction. One is that we have been inculcated with a

moral education and linguistic unclarity that treats happiness as our highest or even singular end. The prevalence of this notion might plausibly generate a false consciousness that distorts our intuitions regarding its importance. Another reason for our reaction might be that our desire for happiness is persistent and we have a general tendency to believe our desires are veridical. But neither of those are sufficient reasons to assign it a central place in our life-structuring decisions. When we correct for these social and biological influences, we might find that we've been captivated by a picture whose utility is at an end. If, however, we are to look somewhere beyond happiness, we must identify a more worthy end. It is to that subject that we now turn.

Chapter Four
What we want for ourselves
A discussion of subjective, objective, and hybrid theories of well-being

We commonly see morality as a *constraint* on our action, as the authority that regulates our behavior, *requiring* us to perform certain actions and *prohibiting* us from performing others. Collectively, these restrictions can narrow the paths open to us. But while this view holds that morality limits the scope of permissible behavior, it generally provides for some surplus freedom of action that remains after the demands of morality have been satisfied.

There certainly are *fully-directive* moral theories that purport to guide all of one's behavior, leaving little or no such surplus. But even assuming such conceptions could be defended, fully-directive moral theories would seem to remove a good deal of the self-determination and autonomy that we normally associate with a human life. In such a case, a person's deliberation and life-planning would appear to simply involve identifying the moral course of action and then pursuing it by the most efficient means. While we have the conviction that other-regarding morality should play a prominent role in a person's life, we nevertheless have some intuition that it shouldn't fully determine *all* of her choices. Indeed, when a moral theory demands too much of a person we tend to treat that as evidence against the theory because we believe that, *ceteris paribus*, an ordinary life should include some freedom from moral constraint where self-interest can operate.[157]

---

[157] See Scheffler, Samuel (2003) *The Rejection of Consequentialism*. Oxford University Press, Oxford

For the purpose of this discussion, I'd like to adopt the view that other-regarding morality should limit the permissible paths open to a person, but that it also should leave *some* space for a person to choose how to lead her life.[158] Assuming this, there's a question as to how one should use the freedom to act that remains after the demands of other-regarding morality have been met. This is essentially a question of what we should want for ourselves after we have given all that we owe to others.

An answer would largely be informed by *self-regarding* considerations, or those things that would tend to somehow benefit oneself. Accordingly, I'd like to inquire into which self-regarding reasons should guide the decisions that structure one's life. Because any such motives would be circumscribed by our moral obligations, it makes sense to restrict the self-regarding reasons we consider to those allowed under various moral theories (as opposed to considering reasons that might exceed what morality permits). These would be reasons for action that contribute to one's own well-being, in some sense or other.

To address this admittedly broad question, I'd like to consider certain theories of well-being and do so by using a fairly intuitive classification provided by Richard Arneson: (i) those that hold a person's self-regarding reasons principally involve improving her subjective well-being; (ii) those that hold a person's self-regarding reasons are rather to improve her 'objective' well-being; and (iii) those theories that

[158] Here I am assuming a circumstance like our own, of just moderate scarcity. There are imaginable, and even actual, circumstances of extreme scarcity in which the observance of even a minimal moral system might leave no remaining bandwidth for personal choice. The issues I discuss here would not find purchase in such environments and, accordingly my conclusions only apply to circumstances in which some freedom of action remains after all moral obligations have been satisfied.

maintain a person has self-regarding reasons to achieve subjective well-being through her pursuit of objective well-being.[159]

My intention in doing so is to explore the sort of self-regarding reasons that issue from each type of theory. I'd like to show how each of these groups of theories remedies a difficulty with the other, but does so at the cost of engendering a new concern. More particularly, I'll try to show that subjective well-being theories resolve a motivational concern present in objective well-being theories, but are plagued with happiness-machine and normative concerns that objective theories escape. I'll also try to show that while hybrid well-being theories avoid the respective worries present in the former two, they make demands on a person's emotional constitution that might be unreasonable, especially in non-ideal cases. I'll conclude by proposing that non-hedonic, self-regarding reasons resolve the concerns with subjective and objective well-being theories without engendering the demands of hybrid views. My overarching purpose in doing so is to explain why this notion is a better explanation of our life-structuring decisions than those allowed by traditional subjective, objective, and hybrid well-being views.

[159] Richard Arneson, "Desire Formation and Human Good" *Royal Institute of Philosophy Supplement*, 59, (2006) pp 9-32. While I adopt Arneson's distinction between subjective, objective, and hybrid well-being views, I direct it toward a different purpose.

I.

Perhaps the theories of well-being closest to our ordinary notion are those that seek to maximize, or at least promote, a person's *subjective* well-being as she identifies it. Such views assert that, after the requirements of other-regarding morality have been met, a person should use whatever latitude remains to do that which makes her happiest, what she desires most, or what will have the best consequences as she sees it.[160] These views posit that a person's principal, self-regarding motive is for her own happiness *and* that a person is in an especially good epistemic position to know what would be best for her. The notion is that after we correct for complicating factors such as imperfect information and limited cognitive ability, a person will desire things that improve her well-being. This self-regarding motivation is posited by utilitarian and consequentialist views (and hedonic internalism, generally) as well as being a precept of social and economic rational choice theories and the public policy that follows from them.[161] Even when a person is mistaken about what would make her happier, such views often respect her autonomy to pursue whatever end she'd choose, in part because they posit that granting her preferences respects her autonomy and that this

---

[160] This group includes hedonistic theories such as utilitarianism and other varieties of consequentialism, rational choice theories, and autonomy maximizing theories. These views differ in significant and complicated ways. Some advocate maximizing happiness (utilitarianism) where others seek to maximize autonomy. They are unified only by their assertion that what's good for a person must improve her subjective well-being, whether this is by allowing her to experience happiness or experience freedom.

[161] For foundational economic applications see Jevons, William Stanley "A Brief Account of a General Mathematical Theory of Political Economy" Journal of the Royal Statistical Society, London, XXIX June 1866, pp. 282-87. "A true theory of economy can only be attained by going back to the great springs of human action -- the feelings of pleasure and pain." (pp. 282) See also Roy Weintraub. (2007). "Neoclassical Economics". *The Concise Encyclopedia Of Economics*.

For uses of utility functions to resolve the St. Petersburg Paradox see Bernoulli, Daniel; originally published in 1738; translated by Dr. Louise Sommer (January 1954). "Exposition of a New Theory on the Measurement of Risk". *Econometrica*. 22 (1): 22–36

will *itself* tend to increase her happiness.[162] For the purpose of the present discussion, what's most relevant is that such theories hold one's self-regarding reasons should be directed toward ends that increase her *perceived* well-being, or happiness as she identifies it.

A concern with such views is that a person might desire things that don't provide for what we would generally consider a good life. There is the worry that a person's pleasure might be maximized by artificial means (e.g. pleasure machine cases)[163] or that a person might prefer meaningless activities (e.g. Rawls' grass counting man)[164] or, more plausibly, that a person might be drawn to base pursuits (e.g. excessive pecuniary gain or bodily pleasure). The issue is not primarily that a person might be *wrong* about what makes her happy, though this would be of some concern for rational choice and autonomy maximizing theories even if less so for consequentialist theories.[165] The deeper concern is that such a person might be *right*, that she might be made happiest by a mundane or vulgar life that isn't best for her. In such a case,

---

[162] This fails to distinguish between consequentialist theories with autonomy-maximizing theories. The former would not advocate pursuing ends that a person believed would not maximize her happiness whereas the latter would. While this is an important distinction, it doesn't bear on the present discussion.

[163] Nozick Robert *Anarchy, State, and Utopia*, Basic Books, 1974, (pp. 42-45)

[164] Rawls, John *A Theory of Justice*, Belknap Press, 1971 (pp. 434)

[165] Consequentialist theories would be less vulnerable to the concern that a person might desire things that are not best for her because these theories ultimately advocate *whatever* state of affairs would have the best outcome. They assume that a well-informed person would be in the best epistemic position to determine what's best for her, but if she were not, such theories would simply prescribe whatever course of action would be most likely to have the best outcome. This could include surrendering to the authority of a third party, assuming the negative utility of such surrender would not override the benefits of the prescribed course of action.

Autonomy maximizing and rational choice theories could not advocate such surrender to a third party's prescriptions because they seek to maximize the autonomy or choice of the subject. As such they would be more imperiled than consequentialist theories by a circumstance in which people regularly desire ends that are not good for them.

respecting her preferences might not be a reliable method for promoting what we tend to think of as her well-being.

One response, as Richard Arneson notes, is that this worry simply rejects a defining precept of theories that seek to maximize subjective well-being.[166] In asserting that some choices are objectively *better* than others, this concern denies that whatever a person desires is, *ipso facto*, best for her when this is precisely what subjective well-being theories posit. While a fair point, Arneson's characterization of these theories' defining thesis, that *what a person desires is best for her*, is open to two different interpretations.

The first is the more innocuous notion that because a person is in a better position to know what's best for her, absent some stronger authority, we should accede to her belief. The second is a more tendentious interpretation; that a person's desires themselves *determine* what's best for her. The first interpretation posits that there's some independent fact of the matter, determined not just by a person's preferences but perhaps also by her capacities and the circumstances in the world around her, and that she is simply in the best epistemic position to identify this. The second interpretation does not assert that there is any such independent fact, but rather that a person's desiring something above all else *makes* it best for her; either because there is nothing more to something's being *best for oneself* than one believing it is or because this belief is sufficient to outweigh any other benefits an alternative might bring (or because

---

[166] Arneson, 2006, (pp. 12) "This implication of the desire satisfaction view might strike some of us as counterintuitive, but this sense of unease arises from the belief that the satisfaction of some basic desires is inherently less valuable than the satisfaction of others. This way of thinking presupposes that some things we might desire to do or get are objectively more valuable than others. This just asserts what subjectivism denies, so the subjectivist should not attempt to tinker with the desire satisfaction view in order to render the view less counterintuitive in this respect"

this belief is necessary for something to be of great benefit to oneself in the first instance). What's at issue between these two interpretations is the ontology of what's best for a person and her life. In the first case, a subject might be wrong and thus an omniscient (or even well-informed) observer might know know better, but in the second case, such a circumstance is definitionally impossible.[167]

The first is less demanding, as we can more easily accept that (1) *a person generally has better knowledge of what would be good for her* and that (2) *accommodating her beliefs about what's best for her will itself promote her welfare* than we can accept the notion that *a person's own belief that something is best of her is sufficient to outweigh all other benefits*. In the latter case, one worries about a person who isn't thinking clearly and a person who is uninformed. One also worries about the somewhat severe subjectivity of value the second interpretation requires and the way in which this might not respect a person's obligations to her future self. In honoring a person's present beliefs about what's best for her we might be disenfranchising her future self in a way that's inappropriate. For these reasons, I'd like to leave the second, more austere interpretation to the side.

If we adopt the more benign, first interpretation of these theories, as merely asserting that, *ceteris paribus*, a person is in the best epistemic position to determine what's best for her, then the objection that a person might desire things that are *not*, in fact, best for her is *not* a straightforward denial of the thesis as much as it is an acknowledgement that things might not always be *ceteris paribus*. It expresses the

[167] The former possibility is what gives rise to the 'ideal-observer' views popularized originally by Henry Sidgwick and later by Richard Brandt and Peter Railton. I discuss these views and the challenges they engender extensively in a prior chapter.

concern that even under moderately idealized conditions (e.g. reasonably good information about oneself and generally good cognitive function) a person could be wrong about what's best for her.

This concern could be resolved by positing *extremely* idealized conditions (e.g. perfect information about oneself and perfect cognitive function). But such an assumption pushes toward implausibility, as it would posit that a person is in the best position to know what would maximize her own well-being because she has something akin to perfect self-knowledge. In such a case, it seems reasonable that whatever she desired would therefore maximize her well-being, but the informational demand on the antecedent condition seems so high as to make the claim almost trivial (i.e. a person would know what's best for her, if she knew all things true of her and her circumstance).[168] The less we insist upon perfect information and mental function, the more reasonable the concern that a person might not always be in the best position to know what's best for her. Thus, if we resist the unpalatable supposition that a subject's desires themselves formally determine what's best for her, the possibility will remain that a person's preferences could fail to sufficiently direct her and, moreover, this possibility will grow larger as we make less idealized assumptions about her possession of complete information and perfect mental function.

In addition to the worry that a person might desire the wrong things, views assigning a central importance to subjective well-being face another concern. When a view asserts that a person has a self-regarding interest in happiness, there is a question about whether that person seeks happiness itself *or* the state of affairs from

[168] Again, I argue for this conclusion extensively in a prior chapter.

which this happiness arises. This 'hedonic paradox' first identified in Sidgwick's *Methods of Ethics* notes that happiness is a state best achieved indirectly because the pursuit of happiness itself tends to be self-defeating.[169] A surer route to happiness tends to be desiring another end, the attainment of which brings happiness as a result. If a person derives happiness from a project, tending a garden for example, subjective well-being theories will tend to posit that her ultimate *end* is happiness and gardening is merely a *means* toward that end. This doesn't match one's first-person experience because such a person typically believes that happiness isn't her goal, but rather that she really cares about gardening itself and happiness is merely a byproduct. Moreover, one who does consciously seek happiness as an end and chooses a means that has reliably made others happy (such as gardening) will tend to be unsuccessful in her pursuit. As Sidgwick puts it, to find happiness one should, "…to some extent put it out of sight and not directly aim at it."[170]

This raises the question of whether subjective well-being theories contend that a person desires happiness or the states of affairs that give rise to it. If a person truly desires the states of affairs, she should desire them independently, even if they are not accompanied by happiness. But the notion that happiness is *not* a person's primary self-regarding motive in acting is inconsistent with the basic premise of utilitarianism and other subjective well-being theories. Indeed, the unifying precept of theories in this first group is that they contend that a person's self-regarding motive in acting is to achieve some positive subjective state. The notion that people are motivated to

---

[169] Sidgwick, Henry (1874). *The Methods of Ethics*. Thoemmes Press., (pg. 3)

[170] Sidgwick, Henry (1874), (pg. 3)

achieve goals apart from the experience they expect to follow is alien to the hedonic internalism on which these theories are predicated.

On the other hand, theories that contend a person ultimately desires happiness, and not the states of affairs expected to give rise to it, are vulnerable to happiness machine challenges. If happiness of greater intensity, duration, and quality can be achieved through artificial means *and* if subjective well-being theories hold that happiness is indeed the ultimate end toward which one strives, such theories would have to prefer a maximally happy artificial state to less happy actual experience.[171] Theories that presuppose it is happiness we seek certainly have responses to this concern.[172] It's sometimes contended that even if artificially induced pleasure were of a greater intensity or duration, it would necessarily be of a lower quality than actual experience. But, if *ex hypothesi*, the experiences are indistinguishable, it isn't entirely clear why the pleasure would be of lower quality.

A more nuanced version of this critique holds that a person doesn't actually want happiness from *any* source, but rather desires the happiness that follows authentically from actual experience. Certainly this is our intuition, but there's a question as to whether subjective well-being theories have the latitude to accommodate this notion. The difficulty comes if we assume that, *ex hypothesi*, (1) there's nothing in one's subjective experience *itself* that distinguishes happiness derived from actual as opposed to artificial means and that, according these theories, (2) the scope of what could possibly be good for a person is limited to her positive affective states. In such a

---

[171] Nozick, Robert (1974). *Anarchy, State, and Utopia*. New York: Basic Books. pp. 42–45.

[172] See Matthew Silverstein's "In Defense of Happiness: A Response to the Experience Machine" *Social Theory and Practice*, Vol. 26, No.2 (Summer 2000) (pp. 279-300)

case, a person would have no reason to prefer authentically derived experience to the artificial variety (especially if she mistakenly believed her artificial experience was authentic).

The suggestion that authentically derived happiness somehow *feels* better doesn't just seem untenable, it also misses the point of the underlying intuition. The concern isn't that an artificially-induced subjective experience is somehow lacking, it's rather that we want our moods to be *connected* to events in the world in a rational way. In the same way that we don't want to experience pleasure from artificial stimulation, we'd recoil at the prospect of deriving pleasure from hideous events or pain from wonderful ones. Our discomfort with artificially induced happiness is an instance of a more general interest we have in reacting *appropriately* to the world. Just as joy is an inappropriate reaction to suffering, happiness is an inappropriate reaction to non-experience. Underlying both is our interest in having a fitting response to the circumstances around us and, to the extent we are able, experiencing the world as it is. The difficulty with subjective well-being theories is that they lack the resources to reach beyond our experience and accommodate these sorts of reasons for acting. So long as one couldn't tell the difference, such theories don't have the resources to explain why an experience authentically connected to the world would be preferable to a more pleasurable alternative.

These issues, though distinct, are permutations of a single concern. We seem to have an inextinguishable conviction that what's good for a person must involve more than her own experience of happiness or desire-satisfaction. It must, we believe, involve properties she possesses, relations she stands in to others, and some rational

connection to the world. One might, then, hold that while it's clear that a person has self-regarding reasons to improve her subjective well-being, such reasons do not exhaust what is good for her and thus are not *by themselves* sufficient to guide her life planning.

II.

A second group of views is responsive to this concern and, accordingly, posit that a person has reason to improve her *objective* well-being. Such views hold that a person should do that which is *best* for her, even if this is not what would make her happiest or what she herself would want. Derek Parfit exemplifies this view in writing, "…certain things are good or bad for people, whether or not these people would want to have the good things, or to avoid the bad things."[173] These views clearly do *not* claim that a person has privileged access to that which is best for her, in part because they hold that certain things are independently valuable for a person irrespective of a subject's endorsement.[174] Objective well-being views tend to survey the various dimensions of a human life (e.g. social, intellectual, physical, professional) and advocate the value of attaining certain states of development across these dimensions (e.g. the development of friendships, at least minimal education, freedom from pain and illness, an opportunity for gainful employment and financial improvement). These states collectively form a list and their achievement is taken to constitute objective well-being. Accordingly, such views have come to be known as 'objective-list' theories.

[173] *Reasons and Persons* Derek Parfit (Oxford: Clarendon Press, 1984) especially (pp. 499)

[174] Arneson 2006, (pg. 21)

Some worry such objective-list theories are so heterogenous as to appear *ad hoc*.[175] We're disposed to think, the objection goes, that the items on an objective-list should be tied to features of personhood or a concept of a person's nature to justify their inclusion. But objective-list theories often don't start with a metaphysical conception of a person and derive features of well-being from it. They rather tend to list qualities that would enhance various aspects of a person's life, but don't show why these particular qualities are necessary (or, as a set, are sufficient) for well-being. One could imagine that particular items on the list might be replaced with other items without manifestly altering a person's well being. Absent such grounding, some argue objective-list theories can properly be seen as *lists*, but fall short of being *theories*.

But even if true, this challenge would be a defect of *particular* objective-list theories, not a problem inherent to this sort of theory. Heterogenous ends, by themselves, are not incompatible with grounding a theory in a person's nature, since this nature need not be monistic, after all. If a person's nature were plural, as Plato's tripartite theory of the soul holds for example, the different parts of one's self might justify heterogeneous ends.[176] If, alternatively, a person's nature derives from her particular capacities, and these capacities vary meaningfully across persons, then the preconditions of well-being might be different for different people. Moreover, it's possible that a person's nature might *not* fully dictate what's good for her. To the extent that a person is engaged in a process of self-construction, she might seek to moderate some parts of

[175] See Rice, Christopher (2013), Defending the Objective List Theory of Well-Being. *Ratio*, 26: 196-211. "Since this theory does not identify any underlying feature common to all list items, it might be argued, it has no principled method for identifying basic objective goods." See pg. 210.

[176] Ibid. The three parts of the soul Plato envisages ((i.e. reason, spirit, and appetite) might each dictate different ends a person should pursue. See Rice, pg. 204.

her nature and encourage others. In such a case one couldn't infer the sources of her well-being from her nature alone. Relatedly, if a person is an evolving entity, her well-being might depend on different things at different stages of her development. So, the fact that an objective-list consists of heterogeneous, fundamentally plural, or even shifting ends, need not, by itself, be an irremediable defect.

But this concern is at least obliquely related to the primary concern with objective-list theories for the purpose of this discussion; namely, that objective-list theories don't appear to provide for happiness *as we generally know it*. A person might secure all the items on such a list and still fail to be 'happy' in the sense in which we ordinarily use the term. Attaining these items might provide a person with a meaningful or even a good life, but it seems possible that such person might nevertheless not experience the positive emotions we associate with happiness.

One response (an analogue of that which Arneson offered to subjective well-being theories) could go as follows; the fact these theories don't provide for subjective well-being shouldn't be seen as a deficiency, if indeed objective-list theories hold that human well-being derives primarily from objective factors. Whereas subjective well-being theories seek to ensure the primary component of a good life *as viewed from the inside*, objective-list theories strive to ensure the objective components of a good life *as they appear from the outside*. Because they're providing for two different notions of what constitutes well-being (i.e. the elements that constitute a life that feels best from the inside as opposed to those that constitute the best life when viewed from the outside), one shouldn't expect the prescriptions of this type of theory to assuage the intuitions that motivate the other. Put this way, there's some sense in which the

problem with an objective notion of well-being is simply that it *isn't* a subjective notion of well-being.

Rather than objecting to what they don't provide, the thought goes, we should rather interpret these theories as challenging our intuitions about what makes a life worth living and as inviting us to think further about what sort of well-being we think a good life must include. If, upon reflection, we conclude that a good life requires certain objective components and not any particular positive feeling, then the concern that objective-list theories don't provide for our subjective well-being will be less troubling. If we thought of well-being as analogous to nutrition, for example, the issue would seem less dire, as something is nutritious even if a person doesn't enjoy eating it.

But even if we do conclude that a good life doesn't require much happiness at all, objective-list theories still face a concern from the motivational standpoint. Such theories might prescribe a life that one simply doesn't value enough to pursue, even if it is attractive from the outside. One will recall that these theories are supposed to provide self-regarding reasons that form the *motivational* basis for a person to select among a plurality of lives that are, *ex hypothesi*, allowed within moral constraints. If objective-list theories prescribe ends that a person isn't moved to pursue, such theories would fail to serve this motivational function. After all, a person can claim that she's given others all that's required of her by operating within the constraints of a moral life. It would seem that the freedom to pursue a life allowed therein should be directed toward ends that person is actually inclined to pursue. She might be motivated to pursue some ends on an objective list, particular those that bring some level of enjoyment, and yet not be inclined to pursue others.

We have the intuition that if something enhances our well-being, we should be able to recognize and endorse this fact.[177] We can endure bad tasting, nutritious food because it creates some tangible benefit we can identify. It might give us more energy, greater strength, and a longer life. But it would seem more difficult to pursue the goals on an objective list that are supposed to benefit us, if one does not find them compelling. Those that bring about some measure of enjoyment offer an additional incentive that might spur is to action. Even if the pleasure they bring distorts the proper motivation, that we pursue friendships, for example, for the pleasure they bring rather than the myriad ways they develop our agency, we'd still stand a chance of getting the objective benefit (i.e. even if we seek friendships for pleasure we still might, as an unexpected benefit, develop our agency through them). The trouble comes when a person cannot be moved to action by items on an objective-list.

Another response to this challenge is simply to include happiness (the subjectively experienced variety) as an item on the objective list. One might reasonably hold that subjective well-being itself has some measure of objective value in a life. But while this might remedy the issue of attaining everything on the list and not experiencing happiness, it doesn't assure that the happiness would be appropriately connected to the other objectively valuable things on the list. It might be that a person found other items on the list to be labors for which she wasn't motivated, such that they competed with happiness for her attention. In such a case, other items on an objective list might be seen as obligations to herself that, like her obligations to others, restricted her freedom to pursue her own ends. So even if including happiness on an objective list

---

[177] Dworkin, Ronald (2000). *Sovereign virtue: the theory and practice of equality*. Cambridge, Mass: Harvard University Press. (pp 283-284)

assured that a life wasn't miserable, it wouldn't provide a clear connection between happiness and the other sources of objective value in a person's life.

This leaves the proponent of objective-list theories in the awkward position of having to argue that a person *should be* motivated to pursue the items on the list, even if she is not, because the prescribed life would benefit her, even though she doesn't believe it would. One could certainly argue that the agent's judgment might be impaired and that if informational deficits were repaired that she *would be* motivated to attain the items on the list for self-regarding reasons. This is akin to thinking that she has a duty to herself, even if she doesn't realize it. But it's difficult to argue that a duty exists when neither the *obligator* nor the *obligee* believe it does, and this is the case with an unacknowledged duty to self. There is an essential tension with claiming that a life is *valuable* for a person if that person doesn't *value* it. In such a case, a person's obligation to herself would appear similar to other-regarding obligations to the extent that it could conflict with, and override, her own self-regarding preferences. Not only would this leave little space for a person's autonomy, it could quite plausibly fail to provide action-generating basis for her action.[178] Absent an agent's endorsement, the fact that some things make a life objectively better is enough to support the claim that they *should* provide compelling, self-regarding reasons for action, but not that they are, in fact, reasons that move *her* to act.

---

[178] One might argue that a view needn't technically deny a person's autonomy to direct her own life. It need only claim that a person should choose a certain course of action, even if she doesn't prefer it. The actual choice is left to the person. It's perfectly plausible that she could retain the authority to choose vulgar pleasures, even if she shouldn't or if those are not what's best for her. One's having the autonomy to make a choice needn't imply that whatever she chooses is the right course of action.

III.

A third group of views seek to avoid prescribing a life that a person cannot endorse while at the same time not surrendering itself to the capaciousness of individual preference. Such views, sometimes referred to as *hybrid well-being theories*, hold that one should do that which is best for her *and* that which she enjoys or desires.[179] Robert Adams epitomizes these views in writing, "…two criteria (perhaps not the only criteria) for a life being a good one for a person are that she should enjoy it, and that what she enjoys should be, in some objective sense, excellent."[180] Attention should be given to the latter claim, as this is what differentiates these views from objective lists that simply include happiness as an end. Hybrid well-being views do not simply hold that a person should pursue some objectively valuable ends and *other* pursuits that bring her enjoyment, but rather that she should derive some level of subjective well-being from her objective pursuits. As Adam's writes, "[i]t is in the enjoyment of excellence that a person's good is primarily to be sought."[181]

Since hybrid well-being theories require that a given action has to be good for a person to contribute to her well-being, they avoid the possibility that a prescribed end could be vulgar or damaging for a person. But since they require that a person must desire or enjoy the end, they avoid worries of paternalism and a lack of respect for autonomy. There is something undeniably attractive about the manner in which hybrid well-being theories provide for both subjective and objective well-being, as they seem

---

[179] See Richard Arneson's "Desire Formation and Human Good" Royal Institute of Philosophy Supplement / Volume 59 / December 2006, pp 9 - 32

[180] Adams, Robert Merrihew (1999). *Finite and Infinite Goods: A Framework for Ethics*. Oxford University Press. (ch. 3, pg. 93)

[181] Adams, Robert (1999) (ch. 3, pg. 101)

to avoid certain pitfalls attendant to each of the prior views. Because vulgar preferences and pleasure machines would not provide moral agents with objective well-being, hybrid theories would not advocate such practices (thus avoiding the worries associated with subjective well-being theories). Because these theories require that a person derives some degree of subjective well-being from her pursuit of objective ends, there is no chance they would advocate a life that appears objectively valuable from the outside but which do not move a person to action (thus avoiding the motivational issues faced by objective well-being theories). *Prima facia*, hybrid well-being views would seem to bring together both what's subjectivity valued in a life with what is objectively valuable, and, in this way, they represent an attractive alternative to both prior groups.

However, this virtuous union doesn't come without a cost. Requiring that moral agents have an emotional constitution that enables them to experience happiness in their pursuance of an objectively valuable life might place an unreasonable demand on them. As a practical matter, it's plausible that a person might *not* enjoy increased (subjectively experienced) happiness as a result of pursuing what she believes to be an objectively valuable life. Even apart from familiar worries about unusual misfortune and psychological defects, the sources of an ordinary person's pleasure could intractably diverge from that which she believes is objectively valuable. Again the analogy to nutrition is apt; we can easily agree that a diet composed of vegetables and the occasional piece of fish might be the most healthy for us. Those with a measure of fortitude can even motivate themselves to live on such a diet for most of their lives. But the additional expectation that this diet will be enjoyable, to say nothing of being *more*

enjoyable than less-healthy alternatives, seems to be a different matter. One might appeal to the benefits of the healthier diet (longer life and more energy, for instance) or the disadvantages of the less healthy one, but while these might motivate us to adopt the healthier one, they don't appear to affect our affinity for it. Knowing that something is the best choice can certainly motivate us to pursue it, but it's less clear how that knowledge enables us to find enjoyment in it. A thing's excellence and our enjoyment of it don't seem to be clearly connected. We sometimes think of ourselves as lucky when we enjoy that which is good for us (i.e. those who genuinely enjoy daily exercise) or when we do not experience the pleasure that frequently attends to that which is bad for us (i.e. those people who don't enjoy sweets or rich foods).

But while there might be some element of luck in the sources of our pleasure, social forces plays likely play a larger role. We can see this in the most prominent of hybrid well-being views, Aristotle's ethics, which does indeed find that a person should find enjoyment in that which is excellent or best for her. This is apparent in his numerous claims in the *Nichomachean Ethics* that "…enjoying and hating the right things seems to be the most important virtue of character," and that "…actions expressing the virtues are pleasant in themselves."[182, 183]

However, first it's important to note that Aristotle's is an aspirational theory, which aims to offer an *ideal* of action measured by the excellent person. So there needn't be

---

[182] In citing Terence Irwin's translation of the *Nicomachean Ethics*, I adopt the traditional Latin abbreviation of the work 'EN', referring to *Ethica Nicomachea*. I utilize the conventional Bekker numbers for citations (i.e. "EN VII.1, 1145b 12" refers to *Ethica Nicomachea*, book VII, chapter 1, Bekker page 1145, Bekker column b, line 12)

[183] EN I.8, 1099a 18-22, EN X.1, 1172a 22-23, See also I.8 1099a 12-15

any presumption in Aristotle that most people will be able to achieve this standard.[184] Second, and crucial for present purposes, one must note that the claims above are abstracted from a broader view that presupposes a wealth of social resources and moral education that many in our society do not enjoy. This is reflected in the way Aristotle repeatedly reminds us in the *Nicomachean Ethics* that a person would need a good upbringing, at least minimal luck, and extensive habituation to have any chance at achieving virtue and happiness.[185] Aristotle also notes that the state has an obligation to develop moral character and virtue in its citizens.[186] The extent of this obligation, for Aristotle, goes beyond merely compelling good behavior to educating citizens and helping them develop emotionally.[187]

This socialization and moral education might plausibly foster sufficiently healthy preferences that a person raised in such an environment might not see such a divergence between the sources of her enjoyment and her objective good. This would be akin to a person raised in a society in which the diet consisted entirely of healthy portions of vegetables and fish coming to enjoy such a diet. Finding pleasure in such foods seems more reasonable to expect of such a person than of one raised on a diet of delicious but unhealthy foods. Thus, when in the *Nicomachean Ethics* we find claims that appear at first glance to impose staunch demands on a person's emotional constitution, such as "[s]omeone who does not enjoy fine actions is not good," the

---

[184] EN I.2, 1094a 11-12

[185] On virtue and happiness requiring favorable circumstances see 1100b 23-34, 1096a 1-2, and 1100a 6-10. On acquiring virtue through practice see 1103a 25-26

[186] EN X.9, 1180a 29-30

[187] EN X.9, 1180a 19-22, 22-23, especially 1097b 32-33, "…it will be hard for someone to be trained correctly for virtue from his youth if he has not been brought up under correct laws…"

demand might actually be more more reasonably met by those to whom Aristotle is actually referring (i.e. those raised in the socially felicitous circumstances he advocates).[188]

Unlike Aristotle's case, the self-regarding reasons presently under consideration must meet a motivational requirement for persons situated in circumstances like our own. Because the reasons posited by these views must have some purchase on people generally, it matters whether a small minority or a great majority of people find they can derive enjoyment from a life they find objectively worthwhile. A few unfortunate persons afflicted with dissonant constitutions needn't cast doubt on a moral or motivational theory. But if hybrid theories issue a standard an ordinary person cannot reasonably meet, it would be provisional evidence against their application to people like ourselves.

To be fair, this assertion requires a couple of emendations. First, it's reasonable to ask a person to stretch a bit to meet a standard. Even if *presently* one cannot derive happiness from the pursuit of objectively valuable ends but some effort, conditioning, and reflection would allow such convergence, a theory that demanded such need not be seen as unreasonable. In such a case, it would be appropriate to gradually reduce one's cognitive dissonance and make the achievement of happiness a life-long goal. Because this kind of emotional unity is so clearly preferable to alternatives, it must be clear that ordinary persons are not reasonably capable of it before we consider it an excessively ambitious end. Second, even if such constitutional unity is ultimately

[188] EN I.8, 1099a 18-22

unachievable for us, it would not evidence against the worthiness of the end or the lofty picture of human agency from which it derives. It would merely suggest such a moral or motivational theory is not appropriate for *us* in our present circumstances, even if it might be entirely reasonable for those differently situated.

The question, then, is whether finding enjoyment in that which is objectively valuable for us would merely require a bit of stretching and self-improvement or whether absent the social and institutional support Aristotle envisages, it would be unattainable for most persons in our circumstance. Certainly, there will be some for whom the convergence of subjective and objective well-being will require little stretching at all. These will be people who enjoy the taste of broccoli more than french fries and those experience of exercise over passively watching television. Conversely, there will be those who find so little overlap between that which feels good to them and that which is good for them that they cannot motivate themselves toward objectively valuable ends. In between fall the vast majority of us and the plausibility of hybrid theories depends on whether this middle group can meet their demands.

It's here that our stubborn psychology becomes a concern. Our experience of subjective well-being seems to elliptically track activity itself rather than favoring the pursuit of objectively valuable ends. But this needn't come as a surprise. After all, if most of us did find enjoyment in that which is most valuable, we would find it easier to lead excellent lives than we typically do. The frequent tension in an ordinary life between that which feels good and that which is good for us forms part of the reason we find it morally praiseworthy to *will* oneself toward a better choices. Those lucky enough

to find a great overlap between the sources of their enjoyment and that which is best for them will actually have less need for a strong will, as even inclination would lead them toward objectively valuable ends.

But while it is fair to expect our rational wills to identify and even push us toward the best life, they have considerably less power to make us enjoy it. Absent the social support Aristotle envisages, the expectation that a person should not only pursue the best life but find also pleasure in it puts a demand on one's volitional capacity that it might be powerless to meet. This risks placing well-being beyond the reach of ordinary persons in our circumstance, through no fault of their own. For this reason, it seems that hybrid well-being views would be better suited for those in environments that promoted a greater convergence between their desires and their objective good.

IV.

This would seem to leave us in a quandary in which none of the three options seem sufficient. I believe, however, there is a fourth that would secure the desired elements of a hybrid theory without the unreasonable demand. This alternative involves meaningfully lowering the psychological demand found in hybrid theories by allowing that happiness need *not* follow from a person's pursuit of objectively valuable ends and requiring only that one *endorses* their pursuit. Such a modification would release moral and motivational theories from the grip of a psychology that seems largely unsuited for

extended contentment. But, at the same time, it would avoid the aforementioned issues that plague both subjective and objective well-being views. Because this view requires that ends be objectively valuable, it avoids worries of base preferences and pleasure machines. Because it requires an agent's endorsement, it eliminates the motivational worries associated with prescribing ends a person cannot endorse. One simply endorses a valuable end and pursues it.

This sounds generally promising, but it prompts one to wonder what shape such endorsement would take. I believe that self-regarding, non-hedonic reasons are uniquely well suited for this task. Such reasons issue from a person's recognition that she is (1) a special kind of object in the world and (2) one that is in a state of becoming. One's connection with this object and awareness that it is an evolving entity generate an enduring interest in its future growth. Neither non-ideal social circumstances nor a psychology that promotes activity over satisfaction interfere with a person's ability to become cognizant of one's status as an evolving object, which makes it more reasonably attainable than the emotional unity hybrid theories require. At the same time, the plethora of dimensions in which a person could flourish as an object allow one's self-determination and autonomy to be expressed by her choice (i.e. one could seek to become patient, wise, generous, etc.). Though circumscribed by the bounds of other-regarding morality, the ends toward which a person strives become uniquely self-defining.

A person acting on a non-hedonic, self-regarding reason is moved to action not by the prospect of happiness, but by the hope it will bring her closer to an end she affirms.

This motivation is entirely compatible with the expectation that the course of action, even if successfully performed, may make her *less* happy than she would be otherwise. Such reasons can motivate actions both small and large. The person who helps her elderly neighbor take out her garbage every week or tends to the plants in an otherwise neglected public field might neither derive greater happiness from the action nor feel obligated to perform it. She might rather want to be a neighborly or nurturing person and if so, it would be wrong to describe her action as having hedonic or other-regarding motives.

Larger, life-structuring decisions are also commonly driven by non-hedonic, self-regarding reasons. This is the thought of the expectant parent and the aspiring scientist. There is surely a hope that these life-defining choices will be accompanied by *some* measure of happiness. But there's also a recognition that their value exceeds whatever happiness they bring and that this value would endure without it. The parent and the scientist shouldn't embark upon their respective paths expecting them to be enjoyable or even hedonically neutral. Instead, they should recognize that parenthood is exhausting and scientific research can be tedious. But when acting on a non-hedonic, self-regarding reason, even if they learned that on balance there would be considerably more exhaustion and less joy than an alternative, they could proceed undeterred. The parent is motivated by *becoming* a parent, by all the ways it will develop her capacities for love and empathy. The scientist is moved by the practice of science, either devoting her life to solving an important problem or giving herself to a process she finds noble.

Not only are they both defined by these choices, their motive in making them could be such definition. It might be that the scientist could have chosen another problem to solve or another discipline to join. If the parent was unable to have children, she might have found another avenue to develop selflessness and empathy. In this way, it's possible that the particular path through which one's self is defined might be less crucial than the definition sought. Alternatively, a particular path might be so uniquely suited to a person's ultimate aim that other alternatives might not suffice. But in either case, when one acts on a non-hedonic, self-regarding reason, one is moved to choose a certain path not by expected happiness, but by some feature of the choice important to defining the person as an object.

The fact that a person must be *internally moved* to act on such reasons prevents the worry that an objective end, although valuable, might be too alien for her to endorse. Indeed, the clearest indication of a person's investment in such an end comes when she pursues it without the expectation of increased happiness and without being compelled by other-regarding obligation. Such choices are defining of character precisely because they peel away the forces that direct much of our behavior. When one takes away the how something feels and what she owes to others, all that's left is what she wants for her *self*. The content of this desire defines her character, as does the absence of any desire for her self beyond a state of feeling.[189]

At the same time, the fact that the consequences of such choices will structure much of her life offers some normative constraint on the ends one would choose. While it's

---

[189] For a further discussion of the manner in which action guiding motives determine ones character, see chapter five.

unlikely that a person would be moved to define her character by pursuing all and only Aristotelian virtues, it's also unlikely that a person could choose to pursue a wicked end. When one detaches subjective motivations, there seems little reason a person could rationally want to become something ugly. Admittedly, what different people choose to make of themselves will vary more than an objective-list might allow, but I think this is a virtue of acting on non-hedonic, self-regarding reasons. There are many dimensions of flourishing and part of the richness of a community is found in the different ends toward which people strive.

It is natural to hope that progress toward such self-avowed ends would generate a sense of fulfillment and satisfaction deserving of the name *happiness*. But we should resist this tendency. It's not only plausible that one might make progress towards worthy ends of her own choosing and yet *not* feel prolonged contentment, it is expected. Either by design or cultural upbringing, it seems we are not built to attain a goal and then rest in our accomplishment, but rather to ceaselessly deploy new, more lofty goals. Our satisfaction is likely to retreat even more quickly than it came as we recognize that we're capable of still greater feats of becoming. Thus, expecting an affective reward for the pursuit of a valuable life might lead to disappointment and frustration.

If one found this line of thought compelling, one might feel a related temptation to redefine our notion of happiness. After all, the successful pursuit of self-chosen ends should at least be sufficient to stave off harsh regret and a sense of worthlessness, and so there's some inclination to think that it is their *absence* we should call happiness. But such a redefinition seems not only disingenuous, it co-opts our referent for the

moments of joy, however fleeting, that we are equally likely to experience. There is enough value to subjectively experienced well-being that we shouldn't simply overwrite this notion with another. It would be more prudent to allow both concepts to persist and instead consider that maybe we should be reaching beyond happiness in structuring our lives; that there are greater hills to climb.

One must acknowledge that this solution will not meet with universal success. There will, no doubt, be those who cannot, or will not, take an interest in themselves as self-determining objects, preferring instead to be led by their baser, subjective impulses. But as I have argued in earlier chapters, one's inability or unwillingness to be moved by one's status as an object does not change this fundamental fact, it simply informs the sort of object she is. Moreover, the recognition that, as well as being subjects, we are objects of experience is sufficiently available that the ordinary person is capable of non-hedonic, objective reasons for acting.

My contention, then, is that non-hedonic self-regarding reasons offer a compelling alternative to subjective, objective, and hybrid well-being views. I believe they avoid worries of hedonic paradoxes and base desires faced by subjective well-being theories, but at the same time avoid the motivational issues attendant to objective-list theories. Moreover, I believe they do this without incurring the demands on a person's emotional constitution imposed by hybrid views. In closing, I should admit that it's reasonable to carry a lingering wish that we were built in such a way that what *feels* best to us *is* best for us. But this might be a moment in which accepting certain incongruities in ourselves can allow us to fashion the best life available to the kind of thing we are.

Chapter Five
A Non-Hedonic, Self-Regarding Motive

I've argued that it's difficult for happiness to function as a central element in moral theory because its doesn't track the goods in a person's life. Presently, I want to focus on the way in which many of the decisions that determine the trajectory of our lives do not make us particularly happy and yet we are glad that we made them. Professional success and the decision to have children, for example, are *not* accompanied by the increases in happiness a rational hedonist would predict. Tenured professors are not notably happier than those who failed to attain tenure, and parenthood actually brings a *depreciation* in a parent's happiness that isn't relieved until children are fully grown.[190] If such outcomes were reasonably foreseeable, a hedonist would have to claim that people acted *irrationally* in choosing to pursue advanced professional status or have children.[191] Yet, those who have received tenure or become parents tend to be emphatically pleased with these outcomes. The tendency, both for the rational hedonist

[190] For the hedonic depreciation of parenthood see "Evaluative and Hedonic Well-Being Among Those With and Without Children at Home" Angus Deaton and Arthur A. Stone, 1328–1333, *PNAS*, January 28, 2014, Vol. 111, No. 4.
For tenured professorship not increasing happiness see Gilbert, D.T.; E.C. Pinel; T.D. Wilson; S.J. Blumberg; T.P. Wheatley (1998). "Immune neglect: A Source of Durability Bias in Affective Forecasting". *Journal of Personality and Social Psychology*. **75** (3): 617–638, (esp. 624)

[191] To the extent that the higher professional status or child rearing carried *equivalent* happiness, it might not be deemed irrational.

and in our ordinary thinking, is to interpret the reported lack of regret as evidence that these choices must have increased our happiness over alternative options.[192]

But we can understand straightaway why these choices might *not* increase happiness: they each cause greater stress than alternatives. Because children are objects of concern, parents tend to worry about their well-being. Higher professional posts will typically carry the expectation of greater performance. A more successful life, across personal and professional dimensions, might be more taxing than a less successful one. While these pursuits provide us with joys we wouldn't otherwise experience, they also bring disappointments. The data shows that the net experience is not positive; these choices don't generate more net-satisfaction than alternatives. Nevertheless, we spend our lives in their pursuit and often without regret.

Set point and adaptation theories simply augment these findings by suggesting that even if these statuses *didn't* carry greater stresses, we would internally generate dissatisfaction. In order to spur our continued action, an insatiability seems to be built into our faculty of desire. The tenured professor will soon aspire to full professorship, and the parent who initially hoped for a healthy baby will soon want her child to excel as maturing person. The epiphany, then, has two parts: (1) although we regard these decisions as pivotal for our lives, from a hedonic standpoint they're just collections of events with net satisfaction levels close to neutral and, more unsettlingly, (2) even if

---

[192] Another familiar response is that such an analysis fails to take a sufficiently broad time slice to appreciate the happiness produced. The notion is that agents believe these choices will increase their happiness *all-things-considered*, even if this requires balancing short term worries against delayed gratification in these pursuits. But there is empirical reason to believe that even a longer time slice doesn't assure sufficient net-happiness to justify such choices. Moreover, this broader analysis tempts us to treat non-hedonic goods as long-term hedonic justifications for these choices. In doing so, such an analysis equivocates between a happiness producing event and one that produces a non-hedonic good.

they turned out distinctly positive or negative, our minds would internally adjust our satisfaction to return to a baseline. The very notion that a life-structuring decision will result in increased happiness over the long term is not merely unlikely because different paths will generally consist of a roughly equal number of happy and sad events, it's almost *incoherent* because our psychology is structured in such a way as to prevent any such sustained increase.[193]

While this explains how these choices fail to increase our happiness, it doesn't explain why we not only don't regret them, but actually tend to emphatically endorse them. We aggressively pursue these ends and then continually reinvest ourselves in the paths they send us down when this would be unjustified if our motive was happiness-seeking. The easy answer I'd like to avoid is that we're simply irrational pursuing these ends in the first place. This explanation holds that we might be aiming for happiness but continually missing because we don't understand its nature. We might think that having children and achieving professional success will make us happier and we're repeatedly disappointed when they don't. If we recognized the pattern of adopting a goal, attaining it, counter-adapting, and then adopting a new goal, we should think we'd wasted our lives. But, the thought goes, our propensity for impressionistic, dissonance-reducing memory serves to mitigate our regret and even leave us with a forgiving fondness for our decisions.[194] On this hypothesis, we act

---

[193] I note that it's *almost* incoherent because, as I've noted throughout this discussion, material improvements will matter at the extremes. Improving, for example, the often dangerous and brutal working conditions of the world's poorest people will indeed result in a sustained increase in their happiness. But those of us not suffering from poverty, poor health, or existential threat are less likely to see a meaningful increase in happiness following life-structuring decisions.

[194] For an explanation of how rationalization can prevent regret when recalling past events see Wilson, T., Meyers, J., & Gilbert, D. (2003). "How Happy Was I, Anyway? A Retrospective Impact Bias." *Social Cognition*, 21, 421-446.

irrationally in seeking sustained happiness and either fail to notice the pattern or forgive ourselves for chasing a chimera.

In what follows, I propose an alternate interpretation of our behavior that avoids the disquieting conclusion that many of our life-structuring decisions are irrational. Broadly speaking, I want to show is that our agency encompasses another perspective, one that derives from our recognition that we're self-defining objects in the world and that we have in interest in the thing we become. I think this perspective gives us special kind of reason to act, what I call a *self-making reason,* and that acting on such reasons in our life-structuring decisions facilitates the self-constitution essential for defining a person's character. On this account, we are not acting irrationally in our pursuit of these ends. We're just not chasing happiness.

I want to begin by first trying to show how our common way of thinking about happiness and morality collapses the space that self-regarding, non-hedonic reasons could occupy, which is why such reasons are frequently overlooked. I want to show that just as our impulse toward happiness generates hedonic reasons for action, our recognition that we're self-determining objects can generate non-hedonic, but nevertheless self-regarding reasons for action. I'll suggest that being objects (1) with which we identify ourselves and (2) which are continually subject to our own judgment, generates reasons to act for their sake. I'll describe how *character*, as I employ the notion, is the part of one's self determined by one's choices, as delimited by her capacities and circumstances. Here, I'm advancing the view that a person's concern for herself as an object int he world creates an interest in her character, and that this

interest generates reasons for acting in a particular manner. With those pieces in place, I'll suggest that the life-structuring decisions that might appear to be irrationally motivated by hedonic reasons are more plausibly explained as issuing from self-making reasons. In closing, I'll argue that this view is not merely to be preferred because it generates a more reasonable explanation of our behavior, but because it portrays us as more interesting creatures than the traditional rational hedonistic view. I think a good way of beginning this account is by explaining why we fail in the first instance to correctly identify the reasons that motivate much of our life-structuring action.

One's Self as an Object

Our motives are often characterized, in the broadest sense, by two sets of distinctions that we take to be aligned. First, we see our actions as being guided by either *self* or *other-regarding* motives. One acts on a *self-regarding* motive if her intention in acting is to benefit herself, and she acts on an *other-regarding* motive if she's moved to action by a concern for others. A tendency in our everyday thought (and a precept of rational egoism) has been to align this with a second distinction between *hedonic* and *non-hedonic* motives. Self-regarding motives are generally interpreted as being hedonic in nature while other-regarding motives are taken to be non-hedonic, and frequently moral.[195] When we act for ourselves, we're thought to be trying to increase our happiness. When we act for others, we're intending to satisfy our obligations or

[195] Other-regarding motives might increase our happiness as an unintended consequence, but their direct, motivating intention is to honor our moral obligations to others, or help them in some supererogatory fashion.

help others in some way, but any happiness that might follow is ancillary. So ingrained are these associations that we sometimes use the notions almost interchangeably. To be selfish is to be excessively concerned with one's own happiness and to be selfless, or other-regarding, is to be more concerned with the welfare of others than one's own joys.

But the alignment of these distinctions can give rise to certain confusions. The assumption that self-regarding and hedonic motives are largely co-extensive tempts some to interpret patently other-regarding behavior as self-regarding, if, for example, an agent takes pleasure in the act. Thus, because the famine volunteer and Abe Lincoln rescuing the pigs seem to take pleasure in their respective actions, they're sometimes interpreted as being selfish.[196] The alignment also invites the fallacious inference that actions performed with no expectation of increased happiness must be driven by an other-regarding motive. Decoupling these pairs allows us to see some non-hedonic behavior may be entirely self-regarding and happiness-generating actions may be driven by other-regarding motives.

For present purposes, the more pertinent confusion follows from assuming that self-regarding behavior *must* be driven by a hedonic motive. This issues from an underlying intuition that if one selfishly prefers x, she must believe x (if attained) will increase her happiness. Whatever aspect of the outcome she prefers is presumed to be the source of some happiness, not only because happiness is taken to be her highest self-

---

[196] In his discussion of psychological egoism, James Rachels imagines Abe Lincoln freeing wild pigs caught in underbrush because 'he could not sleep later if he did not' as behavior that is improperly thought of as selfish. See James Rachels, *The Elements of Moral Philosophy*, 2nd ed, (McGraw-Hill, 1993).

regarding end, but also because it's expected to follow from the attainment of all her other self-regarding ends. Not only will she want *whatever* would make her happiest, but attaining an outcome she desires should *itself* generate happiness. Thus, according to this view, it's incoherent to suggest that one could have an entirely self-regarding end, the attainment of which is not even expected to bring increased happiness.

But I think self-regarding, but nevertheless non-hedonic, reasons abound and that we actually rely on these heavily when making decisions that structure our lives.[197] A reason for action is self-regarding and yet non-hedonic if it moves a person to act for her own-sake, independent of any expected happiness. Such reasons could move someone to perform actions that do increase her happiness, and even actions that she *expects* will increase her happiness, so long as the intention that drives the action is non-hedonic. The concern with such cases is that because hedonic motives are so familiar, there's a temptation to mischaracterize the driving intention as being happiness-seeking when in fact is wasn't. Counterfactual analysis could help disambiguate the action-driving motive (e.g. where the person would have acted in the same fashion even if she didn't expected greater happiness to follow), but the clearest demarcating case of a self-regarding, non-hedonic reason is when a person is moved to perform an action that she believes won't increase her happiness, and does so for her own sake.

---

[197] While I believe life-structuring decisions are best explained by non-hedonic, self-regarding reasons, I do *not* believe these reasons are limited to influencing major, life-structuring choices. As I will try to show, a concern for the object one is motivates a variety of behavior, much of it involving smaller, less pivotal decisions that nevertheless do a great deal to define one's character.

This prompts one to wonder what things or events one could want for her own sake when she doesn't expect they'll increase her happiness (and might even decrease it). Indeed, this will be a null set if we assume that when considering a self-regarding action, we do little more than weigh the expected affective consequences. But I think we can, and do, reasonably want things for ourselves even when we have no expectation they'll increase our happiness because we also can have an interest in the objects we become.

Our affective states involve the feelings one has, on the inside, so to speak, interacting with the world. It is here that one *feels* motivated or tired, starved or full, happy or sad. Such feelings manifest themselves in moods that pervade much of our experience and, in this way, impose themselves on us. But we also have access to an awareness of ourselves as objects in the world. This involves a recognition of our properties and behavior as they appear from the outside.[198] Here, one is *aware* that

---

[198] We admittedly don't immediately see ourselves as we appear in the world and this makes observing ourselves more challenging. When we imagine how we appeared in the world on a given occasion our memory is often clouded and perhaps distorted by the feelings we experienced at the time. Contrastingly, when we recall the actions of others, we just see the action uncluttered by the person's inner-experience. Nevertheless, we can imagine how we appear in the world and try to distill this from the our affective states at the time.

Relatedly, one might believe that talk of an "awareness of oneself as an object" doesn't get beyond our affective experience in the way I'd hope. If awareness belongs to our subjectivity, the thought goes, this awareness is every bit as subjective as our affective experience. While this is a fair point, I'm intending something more modest with this manner of speaking. There is a sense in which seeing oneself through the lens of our affective responses and focusing predominately on these states of feeling obscures our awareness (our admittedly subjective awareness) that we're objects that constitute the world. My point here is that we can have some interest in the kind of object we *are*, independent of how this object *feels*. I'm actually friendly to the rejoinder that our awareness of ourselves as objects is itself part of our subjectivity, so long as one doesn't make the further inference that our self-making reasons, therefore, reduce to hedonic reasons.

she's hardworking or lazy, thin or rotund, happy or sad.[199] Whereas the internal, affective perspective concerns what it feels like to live, the external perspective of ourselves concerns an awareness and assessment of what sort of thing we are in the world. This is admittedly less immediately accessible than our affective states, but we nevertheless have access to vantage point.

This external perspective is most familiar to us as the way we experience and assess others. We commonly make judgments about another person on the basis of their actions and the motives that we believe drive them. These judgments do not simply fall on a spectrum from laudatory to critical, as they will often involve noting distinctive aspects of a person that transcend good and evil (e.g. that a person is devoted, cautious, or reserved). Through interaction, we sometimes form enough judgments to make an overall assessment of a person. I take this awareness of others and the judgment that follows to be part of our everyday practice of *coming to know someone*.

My suggestion is simply that we can and often do make judgments of ourselves in the same fashion. An angry person not only has the affective experience of rage, she also can be aware that she *is* an angry person and she might have any number of judgments about this. She might be proud, ashamed, or want to sublimate it in a more productive fashion, but each of these attitudes requires (1) an awareness of herself as an object and (2) an ability to make judgments about this object. Self-loathing, for

[199] The value of being a happy person is not limited to the subject that feels joy. My point here is that one can be aware of her affective states and develop attitudes about these, independent of their subjective significance for her. A person might, for instance, experience guilt when she observes how happy she is at the funeral of a loved one.

example, requires both awareness *and* judgment of oneself as an object, for one who loathes herself has a poor opinion of the object she takes herself to be. Even if she loathes herself because of the way she *feels*, the loathing derives from a negative assessment of the object that experiences such a feeling (e.g. she loathes herself because she enjoyed the suffering of others).

When I refer to a person as an *object*, I don't mean that reduce her significance to that a table or a chair, for a person is a profoundly special sort of object. While most objects, as we think about them, are simply inanimate, some are animate but incapable of forming intentions, and a relatively few are animate and capable of intentional action. A person, however, is the only object capable of intentional action that is guided by a concept she has of herself. A person is not only aware that she's a part of the world, she can form a notion of the kind of thing she'd like to be and mold herself accordingly. This makes her a self-constituting object, a thing of her own determination, and this distinguishes her from all other objects as we normally think about them.[200] Even though this unique status might depend on her subjectivity, it is not an aspect of it. It is rather a unique feature she has as an object in the world. When I refer to a person's 'awareness of herself as an object,' I am referring to an awareness she has of herself as an object that acts on reasons and the judgments herself in light of this behavior.

It might help to contrast the distinction I'm making from Thomas Nagel's distinction between the subjective and objective perspective of the world in *The View From*

[200] There some evidence that certain animals might be capable of this level of awareness and self-construction. If this were the case, it would certainly count in favor extending them the status of personhood.

Nowhere.[201] Nagel takes the *subjective* perspective to involve one's particular point of view and experience as determined by her position in the world, filtered through her perceptual capacities and consciousness. A perspective becomes more *objective*, for Nagel, as it abstracts away from the capacities and dispositions of a particular person and instead relies on mind-independent features of the world. "An objective standpoint," he writes, "is created by leaving a more subjective, individual, or even just human perspective behind."[202]

Nagel sees the spectrum of subjective to objective perspectives as series of concentric circles where the maximally subjective lies at the center. The benefits of increasing objectivity in achieving a greater understanding of the world justify what Nagel sees as a lifelong project to abstract away from a tendency to see the world from our idiosyncratic and socially bound subjectivity to achieve a more detached perspective. Nevertheless, he sees that a maximally objective perspective is not just impossible for creatures like us but that it's also hollow and valueless at its core, as it treats a person as "…a small, contingent, exceedingly temporary organic bubble in the universal soup."[203] Moral life, for Nagel, accordingly involves a perpetual effort to balance these objective and subjective perspectives (i.e. the reasonable commitment one has to the importance of her own life and ends against the detached perspective in which they are almost inconsequential.)

---

[201] I am indebted to Seana Shiffrin for suggesting Nagel's view as a helpful contrast to the distinction I'm drawing above.

[202] Nagel pp. 22

[203] Nagel, 377

Nagel's notion of the subjective and objective standpoints differs from the distinction I want to draw, as the hedonic and self-making based reasons I want to differentiate would *both* belong to what Nagel calls the subjective view. My discussion is not meant to push our epistemic limits as subjects to transcend our capacities and historical circumstance. On the contrary, my interest is in revealing a perspective one should have *in virtue* of her specific capacities and circumstance that would be devoid of content from Nagel's objective perspective. Whereas Nagel's objectivity is a bird's eye view so high that one's self is indistinguishable from other persons and things, the self-making perspective I'm trying to describe is profoundly terrestrial. It differs from our affective perspective only in that instead of focusing on the way to the world looks from the inside looking out, it asks us to consider what we look like from just outside ourselves looking in. It is, in this way, not a view from nowhere, but more akin to a look in the mirror.

There is some rough analogy to could be made between Nagel's concentric circles and the perspectives I'm trying describe. What I call the hedonic or affective or hedonic perspective lies at the center of the circles Nagel envisions, as it's more immediately available and imposing in our conscious thought. Our pains, pleasures, hungers and longings are omnipresent and in some ways, irrepressible. I want to distinguish these from a self-regarding concern we can have for our choices *apart* from any affective impact they will have. Such a concern issues from our recognition that we exist in the world not only as things that feel, but also as objects in the process of self-creation. For most of us, this interest is not as immediately pressing as our affective states, and our struggle is to engage it rather than escape its grasp.

The reason this affective perspective is less immediately available is evident enough. The vast majority of our perceptual apparatus is outward facing as its function is to interact with the world.[204] If we had another set of eyes that extended away from ourselves but were trained back on us, we could more easily understand how we appear in the world. If, at the end of every day, a person had to review immersive footage of how she appeared in the world on that day, it would tend to evoke a different reaction than her maximally subjective perspective. Such a perspective would allow her to make judgments about herself, like a sculptor standing back and evaluating her work. She could, for example, see how she sounded when she spoke harshly to her child or how she appeared when she sheepishly approached her employer. Whereas the affective perspective is consumed by feelings of frustration for her child and fear toward her employer, the view looking back at herself allows her to see that she might lack patience at home or be excessively timid at work.

One reason we should distinguish between these perspectives is because they could generate opposing reasons for action. The view looking back at herself might encourage this person to be more compassionate with her child and bolder with her employer even if these responses could cause her more frustration and anxiety. Acting on the affective perspective could motive her to be sterner with her child and more subservient at work, which might move her further from the kind of person she aspires to be. Although we reflectively hope that the adoption of such attitudes would ultimately

[204] Here, what I intend is that we generally see, hear, smell, feel, and taste things in the world *around us*. All of those senses can be turned back on oneself (e.g. one can try to smell one's own skin). But there is an extent to which our senses are more naturally trained on the outside world and one must redirect them to oneself. One can hear oneself talk, for instance. But until one consciously tries to do so, this sound generally doesn't register in one's consciousness. Proprioception is clearly a limiting case of this, as it involves a direct awareness of one's body. Many thanks to Gavin Lawrence for this clarification.

result in greater internal peace, there's no assurance this would be the case. The behavior that contributes to the self-making ends she has for herself might come at a cost to her overall hedonic state, and yet, she might have been rational to choose the former. Distinguishing these perspectives allows us to express her reasons for such a choice.

Because the external view I'm describing is a kind of abstraction and therefore less natural for us, it would occupy a ring on Nagel's scale further from the center. But while it's less immediately accessible than our affective states and the direct perceptions that engender them, it isn't beyond our reach. It is not coming from a place of detached, impartiality in the way Nagel's objective view does, but rather from a standpoint deeply invested in and concerned with the thing we are. Here, we are stretching a little to assure that we don't plunge into solipsism, but we are still comfortably within our own experience.

That said, being aware of oneself as an object and even forming judgments about it doesn't establish a concern for it, let alone one that would rival our concern for happiness. If I'm proposing that a class of life-structuring decision is better explained by a person's interest in self-definition than by a hedonistic motive, then it's incumbent on me to show that she is not only aware of herself as an object, but that she cares about the kind of object she is. But I think we can find the source of this concern in the relation between one's affective experience and the object she is. A person's recognition that she's a self-creating object capable of reason-driven action allows her to see that this thing is uniquely connected to her affective experience. There is some

independence, since a person can be noble or wicked irrespective of whether she feels happy or sad. Yet this object is the singular thing to which her experience is tied. She can find new company or move to a new place, but she cannot escape what she is. It is, in this way, the inextinguishable counterpart to her subjectivity and, as such, it will be continually subject to her judgment.

Moreover, this object isn't merely the twin to which her affective experience is paired or even the capsule that encases it. Rather, a living connection exists between the two. If we can see a person as a web of reasons and dispositions motivating her action, this web has two products: a collection of affective states a subject experiences and the object she is.[205] They are, then, two effects of a single cause.[206] In this sense, a person *feels* the affective component of what she *is*. When she looks in the mirror she *sees* that she's thin, but she *feels* the hunger, the tautness of her skin, the languor of her limbs. The choices whose pleasure is felt or pain is suffered are written on the object she is. Because her affective experience and the object she is have a common source, her judgment of this object is not disconnected from the feelings she experiences. She is, rather, aware that her actions and reasons that drive them will have two sorts of self-regarding consequences, effects on how she feels and on what she is.

---

[205] This is not to say that an agent's motives in action are the *singular* cause of her affective states or the object she is because both are influenced by other factors in the world. A person's actions and reasons for action both present themselves as external influences on the affective states and character of others. Here, I simply mean to say that a person's reasons-in-action exert all the influence of which she as an agent of change is capable.

[206] Strictly speak, this is excessively simplified. Her affective states and the object she is are not *only* effects of her actions, they're obviously causes of subsequent actions. My point here is just to describe the connection between her affective states and the object she is.

It's natural to ask which of the two products of our action we more closely identify with, affective consequences or the object that we progressively mold. We have a tendency in moral philosophy to identify ourselves more closely with our subjectivity, perhaps because historically having consciousness has been seen as a demarcating property of personhood, whereas being an object is something we share with inanimate things. But I'd suggest that the fact that a person feels, or even that she's conscious of feeling, shouldn't be the anchor of her identity. Instead, I think one can locate her identity as the chooser of actions that have these two kinds of two effects, where her response, in the aggregate, defines the kind of creature she is. Just as she is motivated to *feel* a certain way, she can be motivated to *be* a certain thing. She might want to be good, or strong, or pure, or kind, but she can have some interest in what she will become. She, most essentially, is the thing that decides in any given case whether to be led by her hedonic reasons or her interest in self-creation (or some other-regarding interest).[207] If she is guided in her life-structuring decisions by the desire to achieve certain pleasant affect of states, this will be partially determinative of the object she is.[208] It's equally defining of character if she allows her self-regarding, non-hedonic reasons to guide her life-structuring decisions.[209]

---

[207] This is not to say that there is only one sort of self-making reason on which a person could act. We can have a multitude of heterogenous self-making reasons, each of which might define our character differently.

[208] In such a case, a person might have a deficient character or in an extreme case, lack any character whatsoever. But the fact that a person isn't concerned with the object she is, doesn't mean that it ceases to be appropriately subject to the judgment of others. See chapter six for a further discussion of the manner in which an agent's motives in acting determine part of her character.

[209] A person's attitudes (which belong to her subjectivity) can additionally be relevant to our assessment of her character in a less direct manner than directly motivating her action. We have expectations about how a person should be disposed to her circumstances, toward others, and toward herself. In this sense, it will matter whether she feels grateful for her relatively easy life or if she lament the remaining burdens, whether she found her deepest joys in acquisitiveness or the well-being of others, and whether she was unhesitatingly guided by inclination or tried to moderate its influence.

In this way, a person can have an interest in the object she is because *this very interest* is defining of character. Rather than merely being creatures who experience pleasures and pains, we're rather creatures who decide how heavily to weigh our affective states in our deliberations. Ideally, we'd most prefer to love and desire the right things. But it's likely that we'll have to periodically choose between what will feel best and what will make us best and these choices will determine our character. We are not, in this way, the sum of our experiences, but of our choices.

What I think we should surrender is the notion that a symmetry must exist between the two products of a person's action, the feelings she experiences as subject and the qualities that define as an object. Acting on a praiseworthy, self-making reason will often engender an affective state, but it might *not* be a positive one. An ugly choice might be accompanied by the feeling of pleasure and a noble choice might be accompanied by pain. The relation also isn't unidirectional because a person's affective states will also have effects on the object she is. The experience of pleasure itself might actually constitute poor character (e.g. schadenfreude) and decency might be found in the experience of pain (e.g. empathy). Thus, although our affective experience and the object we become are intricately connected, they needn't covary.

I think, then, we locate self-regarding, non-hedonic reasons in our concern for the object we are and want to become. When we act out of concern for this object, we act so that it will *be* a certain thing, not so that we will feel a certain way. Because this often gets lost in our rational reconstruction of self-defining actions that also generate increased happiness, we tend to collapse the self-making into the hedonic. It's when

these motives come apart that we're reminded of our interest in self-definition. The nobility of bravery, for instance, lies in it feeling awful.[210] The brave person acts in a noble or self-sacrificing manner despite her fear and in doing so weighs her self-making over her hedonic reasons. We want to *be* brave enough that we're willing to endure the unpleasant feelings that accompany it. Our motive is self-regarding, but it's importantly not hedonic. In such a case, we're aspiring to a state of being even at the cost of a state of feeling.

In what follows, I try to explain in more concrete terms what one's concern for her being entails. I'll suggest that a person has an interest in constructing a character that is appropriately responsive to the world in which she finds herself.

Self & Character

Even if all this is accepted, it remains to be seen how being an object in the world, even an object in which one has a special interest, gives a person a reason to behave in a *particular* manner. The answer I'd like to advance is that the part of our self subject to our own judgment is what I'll call *character* and the reasons motivating a certain class of our behavior determine our character. I can begin this discussion by saying a bit more about is aspect of ourselves.

We distinguish generally between characteristics a person inherits and those that are, at least in part, a product of her choice. We might think of a person's lung capacity

---

[210] I'm treating bravery here as the ability to act even in the presence of substantial and justified fear, as distinguished from *intrepidity* where one might act without experiencing fear. While these qualities might be both good for an object to possess, the latter will tend to feel less unpleasant than the former. One might believe that bravery is a more heroic quality to possess than intrepidity *because* it feels worse and is therefore more difficult to sustain.

and her predisposition for anger as inherited, whereas her cardiovascular condition as an adult and her temperament are both influenced by her choices. There is a sense in which a person's inherited properties and existential circumstances delimit the possibilities open to her and subsequent choices influence the trajectory of her life within these limits. These subsequent choices constitute one's *character*, as I'm employing the notion.

A person's character, then, will be defined by her reasons for a class of action against a backdrop of a particular set of capacities and circumstances.[211] We might think of inherited intelligence as a gift and a propensity for weight gain as a challenge, but one's character is concerned with how one responds to these respective conditions (e.g. a person might have *squandered* her intelligence, whereas another might have been *vigilant* in maintaining her good health). It is in this sense that the inherited parts of a person's life can be seen as posing a question to which her character is an answer. Because character is determined by the imposition of a set of resources on the circumstances of an actual life, it is particular to both the set of resources and to the

[211] Since character is intrinsically concerned with a person's choices, it will have some moral resonance. However, a person's character will also include extra-moral properties, such as her *timidity* or *perseverance*. For the purpose of this discussion, I'm interested in both moral and extra-moral character traits. For a further discussion of the intersection of character and other-regarding morality, see chapter seven.

life.[212] Thus, a land of plenty and one of scarcity will not forge the same character, and the challenge of moving a boulder will not be equally character-defining for the powerful and the frail. To understand the character implications of such an action, we must know both the circumstances in the world and the capacities of the agent.

However, after external circumstances and capacities are fixed, the manner in which character is determined by choices can be fairly transparent. One's choices and reasons lie alongside the circumstances to which they are a response and *together* become the subject of judgment. When one is presented with the fact of human suffering and chooses to become a doctor to relieve this suffering, she becomes *someone who was guided by her desire to help people in her choice of vocation*. Alternatively, when in the same circumstances she is motivated by desire for prestige to become a doctor, she becomes *someone who was guided by prestige in her choice of vocation*. Such a terse characterization tempts one to call the former *noble* and the latter *vulgar*, though more elaborate characterizations will generate more nuanced judgments. But in both cases, we judge character, in ourselves and others, by holding one's reasons-for-action against the circumstances to which they are a response.

---

[212] Character must concern the *actual* interaction between a person's resources and the world (as opposed to a hypothetical interaction) because in a character-defining event, not only are a person's resources imposed on the world, but the world imposes itself on these resources. The world will frequently change a character in such contests and thus the history of a person's character is not merely the imposition of a static set of talents on a series of challenges, but a dynamic narrative. When we claim that, counterfactually, a certain person could have made a good trauma surgeon or war-photographer because of a unique set of talents, we're speculating that this set of talents would have successfully met those challenges. What we're typically not tracking is the effect such challenges would have had on the person if she actually took this path. Perhaps her talents would have sufficed, but her temperament would fail her. Or perhaps she would have overcome these challenges, but the stress of these events eroded her characteristic sensitivity. The impact of large decisions on a life is too complex for an isolated hypothetical to fully represent. For this reason, trying to assess a person's character by the way she *might have* acted in counterfactual circumstances will tend to yield an incomplete picture. Our judgment of a person's character will be more reliable if we restrict it to what she actually did and didn't do, as opposed to what she might have done.

One's character, then, can be seen an all-things-considered assessment of her choices in light of the world in which she finds herself.

Not all choices equally define character. Choices can be so isolated that our evaluation of a particular action should not apply to a person generally; thus we recognize some distinction between what a person has done and who a person is. But there are patterns in our choices and at some point one can have chosen a thing often enough that it becomes constitutive of her. A single choice cannot make a *drunk* or a *miser*, but a long series of choices will. Nevertheless, some actions are so deeply connected to one's network of values that even one decision might be defining of one's character; a *hero* or an *infidel* can be made by a single choice. In some trivial sense, every intentional action one performs will form a part of one's character, but much of what a person does is too routine to reveal any distinctive character. Additionally, much of our behavior is coextensive with that of others, and when we refer to someone's character, we're referring to the perhaps few, but distinctive, points of divergence from others.[213] However, when our choices connect with one another, when general motives pervade much of what we do, or when our choices diverge from the ordinary path of behavior, they become telling of character.

While both large and small decisions can be motivated by self-making reasons and therefore determine one's character, larger life-structuring decisions almost inevitably do. While a single response to an elderly person having difficulty crossing a street will demonstrate compassion or a lack thereof, we might be thoughtlessly rushing on a

[213] The fact that she is *someone who was guided by appetite in a sandwich*, tells us little about her character unless exacerbating circumstances existed (i.e. if, in eating this sandwich, she was losing a fight against life-hampering obesity or if it was the last food available and others were starving).

given day and therefore act poorly. Our life-structuring decisions tend to be different because we're expected to prepare for them and weigh alternatives. They're also typically made in the context of constraint where a decision will express the relative weight one assigns to various goods in life. When we act on reasons that express such hierarchies of value, they're written on our character and thus on the world.[214] My suggestion, then, is that a person's concern for the kind of thing she is generates reasons that can guide character-defining decisions, both large and small.[215]

The parent and the professor seeking tenure, for example, might have been motivated to become a certain thing, as opposed to being motivated to feel a certain way. They might have realized these pursuits are likely to be less pleasurable and require greater sacrifices than alternatives. Being a better parent or scholar might

---

[214] It might be unclear why the action plays *any* pivotal role in the determination of character, if so much is determined by merely possessing a motive or reason to act. It's certainly true that inaction is as telling of character as action is, so both one who is moved to action by her reason and one who is not are both defined by their behavior. Assuming two people in the same circumstances held the same set of values and one acted, it might not be clear why the performance of action itself should somehow determine the actor's character differently than the person who didn't act. Imagine two identical disenfranchised young people consider going to a protest and acting on precisely the same reasons, but one goes and the other does not. We're tempted to say that if the circumstances and the values truly were *exactly* the same, both people would have behaved similarly.

But even allowing that such a situation is possible, the performance of the action, even apart from the motive that drives it, still informs a person's character to the extent that it changes her existential circumstances. The options and information furnished to the person who attended the protest will likely differ from the person who stayed home (i.e. she might have been moved to future action by a rousing speech or met similarly minded people that influence her future choices). Thus, their paths diverge, at least temporarily, in virtue of the performance of the action itself despite having, *ex hypothesi*, identical motives.

Nevertheless, there is still some component of our life's trajectory that is left to chance. Imagine, for example, that one of our two otherwise identical protesters decided to stay home because she had a cold when the other did not and attended the protest. The event might prove to be pivotal in determining each person's character, even though a seemingly minor difference in circumstances prevented one of them from attending. But while moral luck surely plays some role and might occasionally play a large role, our internal convictions should nevertheless pull us in a particular direction. The young person who remained home with a cold will likely find her way into civil protest eventually if her feelings remain strong.

[215] I'm focusing on life-structuring decisions in this chapter just because they are most clearly explained as issuing from self-making reasons. But, as I discuss in chapter seven, self-making reasons also motivate smaller, everyday decisions.

require that one is a poorer golfer or friend, but one might reasonably make these choices despite anticipating the consequences. These choices determine one's character in part *because* they limit one's flourishing in some areas in order to increase it in others. In so doing, they express the relative value a person assigns to the options open to her. Because our lives issue each of us a limited currency, the extent of our pursuit and the alternatives we refuse for our chosen ends express what we take to be valuable in the world. In addition, a choice made for self-making reasons, especially when these come at a hedonic cost, is defining of character because it assigns this a greater importance to what a self *is* than how it *feels*. This begins to explain why a person might emphatically endorse a decision that resulted in less happiness than an alternative. She might have committed herself to a certain project that defined her and in so doing become a thing she respects.

We must, though, resist the temptation to believe that this respect must somehow manifest itself in any measure of happiness. Our hedonic and self-making reasons have different trajectories. Constructing a self resembles a jagged line; there are regressions and corrections, but we can reasonably hope the broad shape over a lifetime is linear and increasing. The pattern of our happiness is elliptical and simply doesn't admit of such progress. Correspondence between our character growth and happiness is, therefore, contingent. Moral education can help us find comfort in our choices and one might even be fortunate enough to inherit a disposition that finds happiness in self-making action. But the facts of our hedonic psychology suggest that moral education can only bring us so far and that we're unlikely to inherit a psychology in which the sources of our happiness closely match the ends we have for ourselves.

We must, I believe, assume there is no proportionate hedonic compensation for developing a fine character and, therefore, that our deliberation could be marked by an persistent tension between our desire to feel a certain way and our interest in being a certain thing. It's reasonable for us to find this tension regrettable and to wish there were a harmony between happiness and virtues we aspire towards, however unlikely it is. Nevertheless, we should also regard this tension as an enduring feature of personhood.

We must relatedly resist our impulse to reduce what I'm calling self-making reasons to a hedonic motive. The tendency is understandable because it's implicit in our language and presupposed by a utilitarian doctrine that's deeply ingrained in our ordinary way of thinking about ourselves. It's also a deeply comforting thought, where the alternative is disquieting. It would be far nicer if becoming a finer thing brought a corresponding pleasure. It's also intuitive to think that if we do indeed have an interest in constructing a certain kind of character, then we should be happier if we succeed. At the very least, one would think that we should have a motive to cultivate a psychology that accommodates our self-making reasons, since cognitive dissonance reduction would dictate that we make our hedonic and character-based motives consistent with one another.

A familiar form of the reduction-to-hedonistic-motive argument is to suggest that the present discussion fails to locate the motivating pleasure in these cases because it takes an excessively narrow time-slice. The notion is that if we just extended the scope of analysis we'd find sufficient pleasure to justify parenthood or tenure. We have a

tendency to think that when we finally arrive at our long sought end, all of our suffering will have been justified. It's a reasonable thought, but it rests on the assumption that securing the ambitions we have for ourselves as objects will bring about a measure of happiness sufficient to justify the effort involved. It is, admittedly, likely that securing these ambitions will bring about *some* measure of happiness. We're happy when our children grow up and when we secure tenure, but the happiness is ephemeral and nothing in the order of what would offset the stress involved in securing these ends. We simply wouldn't raise children or pursue tenure if a net increase in happiness is what we were after. But if we were driven to make ourselves into a certain sort of thing, we might act in just this way.

I believe one advantage this view has over the traditional rational hedonistic account is that it yields a less distorted view of our agency. To explain our self-regarding behavior as deriving from solely hedonic reasons, one must go through gesticulations that render some of our life-structuring behavior irrational. On such a model, a devoted runner exercises because she believes that being stronger will make her happier, even when there's good reason to believe it won't. On the view I'm putting forth, it's important we recognize that she's exercising even though she thinks it will impose a certain degree of suffering. It's significant because it indicates the relative weight she assigns to her hedonic and non-hedonic motives, which is itself telling of her character. It matters that she's knowingly choosing the more difficult, less happy path. It is for this reason that the scene in *Lawrence of Arabia* introducing T.E. Lawrence (played by Peter

O'Toole) is so revealing of his character. There, after lightning another's cigarette, Lawrence allows a match to burn his fingers as he very slowly puts it out. When the other fellow tries it, he remarks that it hurts, and asks what the trick is. "The trick," Lawrence replies, "…is not minding that it hurts."

One must also note that although rational hedonism is attractive because it supposes a greater coextension of virtue and happiness, it does paint us as somewhat vulgar creatures. If to be self-regarding is to be pleasure-seeking, a self is appetitive and simply restrained by the imposition of external morality. There is no place in rational hedonism to suggest that happiness is one of many distinct ends she pursues *for* herself. Talk of character might enter, on such a view, when one considers the relative weight an agent assigns to her self-regarding interests as opposed to her other-regarding concerns. A person might be seen as generous if she assigns more weight to other people's pleasures than her own, but she nevertheless possesses a single type self-regarding reason. However, that would reduce character to simply occupying a place within the realm other-regarding morality, when there is value in allowing character to connect with a larger normative concern. Two persons might discharge their other-regarding duties equally and yet meaningfully differ in the extent to which they flourish as persons. Character and non-hedonic, self-regarding reasons give space to describe this difference. I think helping ourselves to a new explanatory tool might render the world, and its occupants, more interesting. Moreover, even if our happiness *were* more closely connected to the good in our lives, it would still make sense to distinguish between acting for the sake of increased happiness and for the sake of self-construction.

In summary, I've tried to argue that the interest in self-definition I've outlined here is distinct from both a person's desire for happiness and her moral inclinations, as the former issues from her affective perspective and the latter is other-regarding. Amidst the cacophony of competing interests, this motive can easily get lost, but it frequently influences our decision making. Self-making reasons get overlooked because they often dispose us toward actions that coincide with *either* our hedonic motives or our concern for impartial morality. Because these latter concerns are more easily recognized, there is a tendency to characterize our behavior as being guided by happiness or by moral concerns when the picture is actually more complex. To isolate our motive for self-construction we have to locate behavior that *neither* contributes to our happiness nor is encouraged by our moral impulse, which often leaves us at the periphery of our experience.

Nevertheless, I think it is worth seeking out such reasons, not only because they render our behavior more intelligible, but because they show we're finer creatures than an exclusively hedonic/other-regarding model would allow. On this view, we're not irrational creatures chasing happiness, we're rational creatures trying to create ourselves. In what follows, I explore how particular life-structuring choices define one's character and offer a principle that I think should guide such decisions.

Chapter Six
Capacities: Fodder for Character

I hope I've established that our interest in ourselves as objects generates a species of reason that can guide our behavior in certain life-structuring decisions. At this point, I'd like to explain *how* the particular reasons on which a person acts define her character and thus the kind of object she is.

To address this admittedly large question, I'd like to consider a species of decision that largely structures our lives: the choice of which capacities to develop in ourselves. In choosing which talents to refine, we are, intentionally or unintentionally, structuring the people we will become and the lives we will lead. It is, in this way, one of the most profound choices we'll ever make. I want to start by considering some familiar motives that might be thought to form the basis for developing our talents and try to show that they fail to sufficiently explain these choices. I propose that our behavior in these cases is better explained by our motive to define ourselves and our character. I argue that the development of a capacity is character-defining because it recasts a person's existential circumstance and becomes the basis for the adoption of related values. For this reason, I propose that a person should develop capacities that she believes are (1) valuable and (2) responsive to the world in which she finds herself. My hope is that this discussion will yield a clearer sense of what it means to act on a self-making reason and how this action relates the object we become.

I've already discussed how the connection with the object we are invests us with a special concern for it. While we're invested in the success of those we care about, this

generally takes the form of *wanting what's best for them* where this involves providing the resources they require to develop themselves. But the development is nevertheless left to them; if they've been given sufficient resources, their success or failure is recorded on their character, not ours. Our relationship to ourselves is the opposite; the ultimate responsibility to develop the object one will become falls to each of us. We treat the resources open to us as our starting point and sculpt our character in response.[216]

Just as our hedonic impulse disposes us toward pleasure and away from pain, our connection to the object we are disposes us to promote its flourishing. A person wants to be making progress toward some end, or ends, she takes to be valuable, though this can admittedly take many forms. One might want to be kinder, better at her chosen profession, or more empathetic. It's quite likely, in fact, that a person will be pursuing multiple self-defining ends simultaneously, which might be unified under a single concept she has of herself or multiple different ambitions she has for her character. But whatever their internal relation, a person has an interest in making progress toward the ends she has for her self.

Both Kant and Rawls share an intuition along these lines, though for Kant it takes the form of a duty and for Rawls it takes the form of a hedonic reason. Kant's duty for

---

[216] It's for this reason that the question of whether *we make ourselves* or *the world makes us* misses the essential relation between persons and worlds. It is only against the constraints imposed by circumstance that we are able to define our character. Because a person's character simply *is* the response to the challenges posed by her circumstances, in a different world she'd have a different character. It is the *actual* choices we make and our reasons for these that define us (as opposed to the choices that we would hypothetically make in imagined circumstances) and these are framed by the world in which we find ourselves. But the object we are could not exist without the circumstances that impose choices on us.

self-perfection holds that, to the extent possible, we are obligated to maximize our talents and develop the capacities we have.[217] Rawls' Aristotelian principle holds that, as a psychological fact, persons will naturally seek the cultivation of their talents and will prefer more complex activities to less.[218] Put very roughly, Rawls tells us that we will want to develop our capacities and Kant adds that we're right in so doing. The question is which of a person's many capacities should she devote herself to cultivating?

Two Familiar Intuitions

We often think that developing our talents is, *ceteris paribus*, a good thing, but of course this development is subject to limits. First, our choice of talents to develop seems subject to moral constraint. We wouldn't, for instance, encourage Jack the Ripper's talent for murdering innocents without getting caught. We're also ambivalent, at the very least, about developing our own talents at a cost to others, such that we would not have a father perfect his poetry while his family starved.

But even if we are not harming others, we don't see ourselves as entirely free to develop our talents as we wish. Instead, we have two seemingly inconsistent intuitions that guide our deliberations in this area. The first is that we have reason to develop whichever morally permissible talent is *strongest* in us, such that it would be a shame if the tall, coordinated fellow pursued cooking instead of basketball or if the child-prodigy

[217] As I discussed in chapter three, because the duty of self-perfection is imperfect for Kant, its scope is contentious. This duty is, however, uncontroversially limited by our perfect duties, so that, for example, one could never use an imperfect duty to justify harming others.

[218] Rawls, John *A Theory of Justice*, pgs. 426-428

gave up piano to become a mediocre soccer player. Sometimes this is cast as a moral obligation, though it isn't always obvious to whom the obligation is owed (i.e. whether it's owed to one's self or the world generally). When one possesses a great talent, this reason, or obligation, seems to gain greater force. My reason to pursue piano lessons in childhood, if I have one at all, seems importantly weaker than Mozart's. This suggests that we're at least partially interested in the result, and not merely the process of developing oneself. The guiding notion seems to hold that since the distribution of these talents is arbitrary from a moral point of view, a person has a reason (and perhaps even an obligation), proportionate to her talents, to develop them so they benefit others.

The second intuition is that such a requirement is unduly burdensome and instead a person should develop whichever talent makes her happiest. It is in this sense that we commend Bobby Fischer for giving up chess and trying to live a private life and Potter Stewart for retiring from our highest court relatively young to spend more time with his family. The notion that merely possessing a talent incurs an obligation for its development seems to violate a person's autonomy. Those whose level of talent is closer to Salieri than Mozart (a group which, of course, includes the vast majority of talented persons) will need to spend much of their lives nurturing this talent before they can know whether it will bear extraordinary fruit. Even if this dedication is a justifiable choice, it seems an excessive demand. There is a sense that one's life is one's own and so long as one is not interfering with others, one should be free to develop or

neglect one's talents as her happiness dictates. If the inheritance of a talent incurred an obligation to nurture it, we would seem to be servants to a genetic lottery.[219]

It's possible that these intuitions could pull us in the same direction. Developing a talent that is rarer or stronger than those possessed by others could plausibly provide a competitive advantage that might provide security or social advantage, thus leading one to believe it might increase her happiness as well.[220] The tall, coordinated fellow doesn't merely benefit society at large by developing his talent for basketball. In a market economy, one might believe the surest method to securing happiness might be to develop her rarest or strongest talent.

But the concern is not that these intuitions could go together as much as the reason they might come apart. The intuitions about whether to develop talents that make one happy or those that are rare or strong might be proxies for a deeper question about whether one should develop her talents for her *own* benefit or for the benefit of *others*. Developing a talent that would promote one's own happiness seems transparently self-regarding and the paradigmatic cases of developing a rare or strong talent (curing diseases or performing artistic or athletic feats) seem to benefit others more clearly.

---

[219] In this light, even the more modest notion that possessing a talent gives one a *reason* for its development seems contentious. A talent for theft doesn't give one a reason to thieve, so the reason cannot simply be rooted in the possession or rarity of the talent. Mozart's reason to pursue music must be rooted in the value of music itself or its value to others, and perhaps even in his seeing it as so. But if the mere capacity to 'x' does not itself give someone a reason to 'x,' as surely it does not, then the fact that a person is one of the only people who can 'x' does not *itself* provide reason to 'x.' This intuition suggests the reason for acting must be rooted in some value the action has for the agent and that, although this value can take many forms, it must contribute at some level to the agent's happiness.

[220] It will not escape notice that to this point I've argued extensively that such development is actually unlikely to increase her happiness. The idea here is simply that a person might plausibly *believe* that developing a rare talent might bring external validation and therefore happiness, even if this is not the case. One might then act on the expectation that developing her strongest talent is a promising route to increasing her happiness, even when this strategy is unlikely to succeed.

There is an open question about whether, either in part or whole, our talents belong entirely to us or if they're owed to the service of others. Such inquiry surely begets further questions. If, for instance, developing a talent gives a person additional ability, the possession of which would obligate her to serve others, one can ask whether a person should be required to develop such a talent (i.e. it would be an obligation to incur further obligation). It is in this sense that in Michael Ondaatje's *The English Patient* Katharine Clifton declares, "A woman should never learn to sew, and if she can, she shouldn't admit to it."[221] Conversely, if such talents benefit one's self, one wonders if a self could be obligated to her future self to develop a talent, especially if that very talent is part of what will differentiate the future self from the present one.

But instead of probing these issues for the moment, I'd like to whittle away at the question of what should guide the development of our talents because, in its present form, it tends toward a false dichotomy; namely that (1) the pursuit of happiness and (2) our obligation to others exhaust our possible reasons for developing ourselves or acting generally. I'd like to show how a third option identifies the relation between our actions and ourselves as objects and in doing so blurs the line between what we ordinarily think of as selfish and selfless behavior.

En route to doing this, I'd like to provide some initial argument that the intuitions expressed above, though *prima facie* compelling, would be poor reasons for developing ourselves. While we might feel otherwise, our contemporary understanding of happiness makes it fairly clear that developing a particular talent is unlikely to lead to

---

[221] Ondaatje, M. (1992). *The English Patient: A novel*. New York: Knopf.

greater happiness than developing another because any increases in happiness are only contingently associated with particular talents. As I've argued earlier, it isn't at all clear that a person would be happier developing her artistic talents than her talents for finance (or vice versa). Small features and unforeseeable aspects of these different paths will typically influence the resultant happiness more than the development of the talent itself.[222] If an aspiring artist, or banker for that matter, is able to locate a similarly minded community, she will generally be happier than if she works in isolation. Both will tend to experience less stress and unhappiness if they are professionally successful than if they are not. Both will tend to be happier if their commutes are shorter and if they remain healthy.[223]

But these contingencies are unforeseeable from the forward-looking perspective, and more to the point, none of them attach directly to the development of the respective talents. There is nothing about the development of artistic skill that *itself* generates more or less happiness than the analytic skill used in banking. When we bracket contingencies of the sort described above, our current understanding of happiness would leave us with two predictions regarding the development of these talents: (1) neither is substantially more likely than the other to result in happiness and (2) neither will contribute to or detract from one's happiness nearly as much as other factors contingently associated with the development of each. Thus, the intuition that we

---

[222] Gilbert, Daniel, "The Peculiar Longevity of Things Not So Bad," *Psychological Science* 2004 Jan; 15(1): 14-9

[223] See "Developments in the Measurement of Subjective Well-Being" Daniel Kahneman and Alan B. Krueger, *Journal of Economic Perspectives*, Volume 20, Number 1—Winter 2006—(pp. 3–24), (esp. 12-13)

should develop *whichever talent makes us happiest* isn't fully coherent. One's being happy or unhappy isn't primarily to do with which talent she develops.[224]

The intuition that we should develop our strongest morally permissible talent presents a different sort of problem, namely that the *strength* of the talent doesn't seem nearly as important as its *content*. We certainly feel that a person who possessed a talent that would allow her to cure cancer should surely pursue this over a lesser talent for gardening. But we're less certain that the person who happens to be endowed with the amalgam of features that make a great bowler should therefore spend her life bowling. More pointedly, when we imagine a person who possessed both an unusual talent that would give her a moderate chance of curing a terrible disease and a much stronger talent that would make her the best of bowlers, we have the sense that the agent should develop her talent for medicine and not bowling, and the fact that her talent for bowling is rarer or stronger seems largely irrelevant. The content of the talent is what seems to guide our intuition about whether it should be developed, whereas its strength seems to be a peripheral consideration. More precisely, our intuition seems to be that the content of the talent should be treated as a multiplier on its strength, such that one would need a massive talent for a relatively trivial capacity to outweigh a modest talent for a noble end. Thus, the notion that one ought to develop whatever capacity is *strongest* or *rarest* seems to fix on the wrong property.

---

[224] This claim might be importantly false at the extremes. Fostering a talent for patently self-destructive behavior (gambling, mutilation, or addiction, for example) *can* markedly decrease one's happiness. However, even here we can note that such destructive behavior might not be accompanied by a decrease in happiness that matches the decrease in human function that would result. One might continue to be a relatively happy gambler or addict even as the essential architecture of one's life falls away. This is another illustration of the way in which happiness and flourishing can diverge. Learning to love the right things is a precondition of flourishing, but not necessarily of happiness.

Character and Choice

Even if we agree that we shouldn't develop a talent because we believe it will make us happy or because it's rare, we do *not* believe that within this realm it doesn't matter which talent we develop because they're all permissible and equally adequate choices. We believe the capacities a person chooses to develop matter, as do her reasons for choosing these. The present questions regard (1) how the decision of which talents to develop is related to a person's character and (2) what should regulate this choice.

I've already detailed the manner in which character, as I'm using the term, is defined by particular choices a person makes in light of her basic circumstances. I've also argued that a person has a particular interest in herself as an object and that her action can be guided by an evolving concept of the object she wants to be. At this point, I want to suggest that the talents or capacities one chooses to develop are profoundly determinative of character in two different ways: by (1) recasting the question to which her character is an answer and (2) generating further associated values. Broadly speaking, the notion is that because character is predicated on a person's response to her circumstance, the development of a capacity actually changes her (and thus reframes her existential question) and disposes her to make further associated choices. The development of talents is, in this way, a point of departure for one's character. I'll suggest that in choosing which we want to develop in ourselves, one should be guided by a notion of the object she wants to become.

We generally think developing talents and capacities is beneficial for a person. But this development will tend to alter a person's basic circumstance by expanding the

options available to her. Someone who develops a banker's quantitative skills, for example, will likely be presented with employment alternatives she wouldn't otherwise have. She'll have to decide whether to seek employment at an investment bank or a charitable organization, or alternatively, to work in a capacity that doesn't use these skills at all. If she does use these skills, she'll have to decide whether to defraud her employers or increase their efficiency.[225] Each of these decisions will be defining of character and they plausibly wouldn't arise without the development of this capacity. A person with very few developed capacities (e.g. a very young child) will not be in the position to make as many character-revealing choices because her horizon of possibility is much narrower.

Since character is defined by the imposition of a *particular* set of capacities on the challenges of a life, both the choice and the development of a new capacity will alter someone's character. Because a new capacity will allow different responses and generate new challenges, in developing one, a person is expanding (or at least redefining) the set of choices she'll face later. In choosing *not* to develop a capacity, she is choosing to restrict the options available to her in the future and thus the

---

[225] It's fairly uncontroversial that defrauding one's employer would be prohibited on moral grounds and so one might think that this choice needn't have a self-making motive. It might be that one's hedonic, self-regarding motive is simply limited by her other-regarding duty and a self-making reason never comes into such a deliberation. This is certainly possible, but it's also possible that one could have a self-making reason to not defraud her employer as well (i.e. one might not want to be the sort of thing that acted in that fashion). This might simply be coextensive with one's moral obligation, but it could exist even if one didn't have an other-regarding prohibition on the act. Imagine instead that one's employer had unjustly enriched herself at the expense of her employees and that morality would dictate that such money should be redistributed. In such a case, even though her other-regarding reasons would allow or even require her to defraud her employer, the agent might have self-making reasons not to do so (i.e. even though she might agree that such redistribution should occur and even that defrauding might be an appropriate means, she might not want to be the object that did it.) This is the sense in which even if one agreed it would be morally imperative to kill the future despot who would later cause profound suffering, one might not be able to bring herself to actually do it.

expectations that can be imposed on her. In both cases, the choice of which talents a person will develop and which she will not alters her existential circumstance and thus recast the question to which her character is an answer. Since we cannot choose the world in which we find ourselves, deciding which resources we use to engage the world is the most fundamental thing we can do to determine our the path our lives will take and our character.

Relatedly, different particular talents will tend to encourage different kinds of subsequent choices and values. While some choices will have analogues in other lives (i.e. the banker and the poet might both face the question of whether to marry for love), some will not (i.e. the poet chooses her vocation knowing it will likely result in relative poverty, whereas the banker does not face this choice). To reduce cognitive dissonance, our preferences adapt so that we tend to stop desiring things we believe to be beyond our reach. For this reason, when a person chooses to pursue a path that will make certain outcomes unavailable to her, she will have a tendency to stop desiring those outcomes. In this way, the choice to become a poet will tend to discourage certain values incompatible with this choice, such as the desire for wealth. Conversely, a person who develops a banker's quantitative skill is more likely to come to value pecuniary gain and will be less likely to value the vocational freedom that the poet might prize.[226]

---

[226] Certainly, the antecedent preference for acquisitiveness could motivate one to make pecuniary gain a central goal, which then would make banking an appropriate choice of profession. In such a case, the presence of a value leads to the development of a certain capacity, so the relationship is surely not unidirectional.

One will note that whereas developing a talent will almost certainly impose subsequent character-defining choices, it will not *necessarily* require the adoption of related values. But the fact that their adoption isn't necessary doesn't imply that these values are simply chosen or rejected by a person without consequence. Because the values are non-consciously adopted to reduce cognitive dissonance, a person will find that developing a talent will tend to alter her deliberative landscape. If the poet's choice of talent and vocation will relegate her to relative poverty, then maintaining her preference for finery will impose a certain degree of misery on her. If she somehow consciously prevented her non-conscious tendency to adapt, she would, in so doing, be consigning herself to ongoing discomfort. The development of a talent, then, might require either the non-conscious adoption of a related set of values or, in the rarer case, the conscious decision to endure continuing discomfort to maintain one's prior set of values. But in both cases the development of a talent challenges one's hierarchy of values.

It's possible that a development of a particular set of capacities, values, and their resultant choices may lead a person far enough that certain lives are unattainable for her.[227] In other cases, a person might break from a pattern of choices and recreate herself to some extent, resulting in two different characters and a fragmented life. But in each of these cases, the choice to develop a capacity has rippling consequences throughout one's character. Because the development of talents changes one's

---

[227] This is not to deny that external forces alone could lead a person to a morally blind alley. The choices made in such a circumstance might well be deeply determinative of character, but the fact that the choice was forced upon a person and not the product of her choosing (e.g. Sophie's choice) is relevant to the determination of character. We would have a tendency to see such a situation as tragic where we'd be more disposed to locate responsibility with the person whose knowing action led her to such an impossible decision.

circumstance and because developing particular capacities will encourage associated values and preferences, one's choices in this area are deeply determinative of character.

Value as a Criterion

Having some preliminary grasp on how one's choice of talents impacts one's character, I want to address the question of *what principle should regulate this choice*. Because character concerns a person's response to her circumstance, it would appear difficult to specify a universal rule irrespective of the different environments in which a person might find herself. In a land of plenty, one might appropriately devote herself to poetry or music, but in the midst of great scarcity, such a choice would be neglectful. I think, however, when we consider how different circumstances on the continuum from abundance to scarcity affect the range of possible character, a general principle emerges.

First, we can narrow the continuum straightaway, as both excessive abundance and extreme scarcity attenuate the range of possible character. Circumstances of extreme scarcity and suffering are so bleak that we expect little of a person, and this profoundly restricts the range of character available to such persons. It might, in such a case, only be possible to be unexceptional (e.g. to wither under the weight of such privation) or heroic (e.g. to somehow prevail or sacrifice oneself). Such environments, in fact, are terrible not simply because they cause a great deal of pain, but because they so dramatically reduce the self-making possibilities open to those who suffer them. The

nuanced aspects of character requiring social interaction and a modicum of existential stability may well be unavailable in such circumstances. When one's deliberative landscape is dominated by a constant need to provide for basic human functions (e.g. continued survival and the reduction of pain), our ordinary moral predicates don't find purchase and there will not be resources for any but the most rudimentary aspects of character. The savagery of such environments lies in their lacking the requisite features for morality and character to take hold.

Conversely, a land of excessive abundance, in which all desires were easily satiated, would ask so little of a person that it would be difficult to locate a motive for acting in the first instance. It would seem likely, in such a circumstance, that our propensity for counter-adaptation would search for an end that required pursuit and that we'd fall into despair if we couldn't locate one. More importantly, acting on a choice wouldn't indicate relative value because options would be unrestricted. If we assumed that in such an environment a person had unrestricted time and funds, one could sample every profession, relationship, and object without consequence.[228] For a person's action to indicate some hierarchy of value, at least *some* resource available to her must be restricted. In this way, the development of character is curtailed by both extreme abundance and scarcity. Character is determined by choices made among constrained options, where choice indicates a hierarchy of value. Extreme scarcity

---

[228] I'm assuming unrestricted time because even the ordinary length of a life imposes some restrictions on a person's choices. In the normal case, for example, one cannot choose an unlimited number or professions or partners to pursue. Nevertheless, some limitation would nevertheless be imposed order in which options were sampled (e.g. the Mona Lisa being the first painting one saw as opposed to the thousanth or Harold being one's first partner as opposed to her seventh.) I'm indebted to Gavin Lawrence for pointing out the limitations imposed by option-order.

removes choice and extreme surplus removes constraint. In neither case can character find an initial footing.

Character, instead, finds purchase where there's some mixture of choice and constraint. It will be instructive to consider an environment of moderate abundance and one of moderate scarcity, because each illuminates a part of a principle to guide the development of our talents. First, consider a land of moderate abundance, where the basic needs of all are met (such that there is no extreme suffering), but where there's not such overwhelming abundance that any possible desire is immediately satiated. If we posit that available resources alleviate the need to provide certain kinds of help to others, such a circumstance would ease *some*, and even many, of the other-regarding obligations we typically labor under.[229] If we were able to accept the notion that our happiness is largely self-regulating and doesn't require our attention (especially in a land of moderate abundance), we could free ourselves from another end that preoccupies us. A moderate abundance thus poses a deceptively vexing question; if one weren't to act for happiness or for the needs of others, what then should guide her action?

---

[229] Clearly a sufficiency or even abundance of distributable resources wouldn't relieve all our other-regarding obligations, as we'd plausibly still have a variety of positive and negative duties we owe to others. A plethora of resources wouldn't, for example, remove our obligation to not harm those around us. However, an abundance of resources would relieve us of taking on daunting burdens associated with things such as poverty, famine, and human subjugation. But such resources would also ease more common obligations, even if the scope of such relief is unclear. In such a case, a parent would not have to sacrifice to afford a sports program for her child, for example. (What's less clear is whether a parent who would otherwise spend a great deal of time taking her child to weekend sports could equally honor her self-making reason or obligation to nurture this child's interest by hiring a surrogate for such purposes. At the extremes, using such resources to satisfy as many of one's obligations as possible would be neglectful.) But it's nevertheless clear that the availability of such resources would alleviate some other-regarding burdens, even if the extent such removal isn't clear.

One might have the view that in such a circumstance a person's behavior wouldn't matter, that it would be perfectly acceptable for a person to act according to whim or indolently pass one's time. If so, that would lend support to the notion that our own happiness and the needs of others are our only two principal ends. But I rather think we have some intuition that a person who behaved in that manner would be lesser for it and, even in a land of moderate abundance, other ends persist. If that's true, then by consciously peeling back our familiar sources of obligation, this envisioned circumstance can help reveal other ends we might overlook.

If the world is a complex whole, then a person is a part, even if an infinitesimal one. When a circumstance of moderate abundance removes many of the other demands that would command her attention, she can consider what sort of thing she'd like to be. While a single part cannot hope to manifestly alter the trajectory of such a large whole, she can nudge it, even if just a bit. She can join in a tradition she endorses or oppose one she doesn't. She can add to progress made in a certain pursuit or alter the direction of another. Whatever she chooses, she'll also be an instance of a certain assortment of values in the world. She will be a bigot or a suffragist, a participant or a bystander, a banker or an artist. But her choices, even in some small way, will help determine the whole and this power endows her with the potential for character.

In virtue of this potential, I believe the development of one's talents should be guided by that which a person takes to be non-instrumentally valuable. The reason for this lies in the relationship between a person's reasons for action and the world. Because these reasons are constitutive of us as objects (of our character), they form a part of the

world, and not merely in the way other mental events do. Our passive beliefs and unreflective preferences are largely imposed on us, but we *choose* the reasons that guide our actions. The reasons we become what we are, in the aggregate, are the reasons the world is what it is. If these reasons define us and we make the world, in some sense, these reasons make the world. If we choose to become an instance of that which we value, the world will be structured by these values.

Suppose, for example, a person believes poetry is beautiful. Choosing to become a poet for this reason determines her character, and especially so when pursuing this end requires a significant portion of her life. She'll spend her life learning to create something beautiful *because* she values its beauty. She will begin to develop a poet's sensibilities and her reasons for choosing this path will be constitutive of her character (she will be the person who acted on these reasons). The possibilities excluded by her choice also define her character, as they indicate the cost of her choice and its relative significance to her. When her choice to become a certain thing expresses what she takes to be valuable in the world, she becomes, as an object, a manifestation of this value. While this might be insignificant in isolation, it's palpable in the aggregate. One person's choice might make just a single poet, but a multitude of such choices creates a literary tradition. One person's protest might be lost in the tumult of the day, but enough people's protests will change the world. In the same way a literary tradition or an intolerance of racism will help define a society's distinctive culture on the macro level, on the individual level such commitments define a person's character.

The result is less productive when someone develops a talent because she believes it will make her happiest or because it is strongest. In addition to being ineffectual for the reasons discussed earlier, being guided by a hedonic motive makes a person, and by extension a world, appetitive. Because happiness doesn't track our achievement, especially in the tedious stages of our pursuits, a world structured by this motive would likely be less accomplished than ours. Our psychology might allow us to be perfectly content with pale versions of ourselves and, by consequence, a lesser world. One who acted on the motive to simply develop whatever talent was strongest in her would allow an arbitrary distribution of endowments to determine her character. Creatures that choose to refine whatever capacity seems strongest in them, leave their character to chance.

Absent from such a circumstance is a world which makes itself in response to the challenges before it, patterned by the endorsed values of its inhabitants. When we think of developing ourselves toward a certain end, we have a tendency to become preoccupied with how it will feel but fail to consider what we will become. This is

unfortunately because we're not only creatures that can feel, we're objects defined by the reasons we choose to act on, which, taken in the aggregate, constitute the world.[230]

The preceding analysis doesn't narrowly circumscribe the values that should guide a person's self-development, but part of this follows from assuming a moderate abundance of resources. When instead we assume a *moderate scarcity*, another demand of character is felt more clearly. In such a circumstance, one has alternatives from which to choose, but nevertheless must mediate between competing obligations. In a circumstance of moderate scarcity, any number of options might be possible, but not *all* of them together. Here, one chooses to flourish as a parent or a professional, but perhaps not both. One can be an accomplished athlete or a devoted friend, but one will take precedence and that decision will be defining of character. Here, one genuinely chooses what she will become and what she will *not*.

A circumstance of moderate scarcity abides by the initial principle governing a moderate abundance, namely that one's choices of talents should be guided by a person's sense of what is valuable. However, it illustrates the importance of these

---

[230] This language of being *one part among many* and imagining the effect of *other parts acting on the same motive* might lead one to believe character decisions involve some sort of universalization test. But this is to overstate the matter. Motives do not derive their character significance from their universalization being impossible, unthinkable, or praiseworthy. One is a beautiful or terrible part irrespective of others. The motives of others only become relevant to the extent that they materially determine the subject's basic circumstance. In a land where strangers regarded each other with fear, offering assistance might indicate a generous character. But in a less guarded environment, such an action might not be character defining at all. Here, the character significance of generosity only takes hold *because* that motive is not universally (or even commonly) willed. The self-making reasons that define a character in the most distinguishing way will tend to be those that are *not* commonly acted on by others.

The most familiar universalization tests are used to determine whether an action is permissible. But in the case of character, the appeal to the whole is a way of reminding someone of their status as an object that is part of a much larger object. In this way, it can indicate that acting on a given reason might have a broad range of character implications, not merely that such action is permissible or impermissible.

values to being *responsive* to the world in which one finds oneself. As scarcity increases, certain actions become more imperatival, until a single permissible action leaves no choice at all. In a very slight scarcity, for example, the amount of her own resources a parent devotes to her children might lie on a spectrum from heroic to generous to normal to selfish to neglectful. In an extreme scarcity, the parent in a starving family faces an imperative to forgo her own interests for her children's, where her choice will either be heroic or neglectful. The point is not merely that in cases of both extreme scarcity and abundance the distinctive hues of character evaporate, but rather that the character significance of a choice is determined by the actor's circumstance and that her choice indicates what she takes to be valuable in the world. In some sense her choice, taken as a time-slice, constitutes what she is in the world.[231] It is one small vote she will cast in the referendum of what the world will be.

There is a temptation to think that one who acts on values that aren't responsive to her circumstance (or who fails to act on those that are) is irrational. Certainly, there are cases in which one's chosen end presupposes a prior end, so it is irrational to neglect the latter (i.e. when faced with starvation, it would be irrational to devote oneself to the arts.) One might also act *immorally* if she neglects important obligations to others for her chosen end (when one devotes oneself to meaningless hobbies to the extreme neglect of her children.) But there is yet another reason one's values should be responsive.

---

[231] This certainly overstates the matter because the character significance of each of her choices is extended over time and as an object she is the result of all of these in the aggregate.

Our capacity of self-determination incurs some onus to respond appropriately to the circumstances before us and not merely in the way demanded by minimum rationality or our basic moral obligations. It is this ability that endows us with the potential for character in the first instance, and the judgment we make of ourselves as objects concerns how we use it respond to our basic circumstance. There is a degree of *can implies ought* here, where one is a poorer object for failing to take up a cause when she is both capable of the task and aware of its value. This falls short of being a straightforward moral obligation because it doesn't appear to be owed to anyone in particular and isn't a case of clear irrationality because there's nothing irrational about simply be lazy or appetitive. It's rather that neglecting the task entails that one has failed (at least provisionally) to express her potential as an object. One wants to say this is a failure of *appreciation*, that one fails to appreciate her circumstance as an object capable of making the world (even in some small way).[232] Her failure to appreciate this is equally part of the world, of course. The world is just a bit paler for it.

The hues of character created by constrained choice in a life are typically heterogeneous and, in their totality, frequently transcend our moral language of good and evil. One is a devoted parent, but a reluctant spouse. Another is uncommonly kind to strangers, but unintentionally cruel to those closest to her. A third is a patient and supportive friend, but is so unreliable others know not to call. Character is messier than the paradigm cases of moral action because it's determined by a multitude of choices across different dimensions of a life. Our success in character-defining projects is

[232] This is overstating the matter because in the ordinary course of a life a person faces a multitude of such moments and her character will be defined by complicated amalgam of choices.

typically partial and placed alongside our failures in the ultimate assessment of our character. The ordinary result is that one is neither as bad as she fears nor as good as she hopes. It doesn't, however, follow that we're all the same because the manner and degree in which we succeed and fail will be distinctive and this is what differentiates a person's particular character.

One might accept the preceding and yet not feel this is sufficient reason to guide a life-structuring decision. One might be entirely unmoved to know she is in some way constitutive of the world or she might, more plausibly, believe her affective perspective is more palpable and therefore more important to her than seeing herself as a thing defined by her reasons. Her affective experience is, after all, irrepressible whereas it's easy to simply forget that she's an object in the world. This we must simply grant. Our affective experience is more pronounced because we have no equivalent mechanism that allows us to zoom out and see our relation to the world. But we should not confuse the volume of particular reasons with their ultimate significance. Our most appetitive drives are often the loudest, but part of a moral education involves distinguishing their volume from their importance. We have no direct perceptual experience of moral laws, but we don't discount them for that reason. The fact that our objectivity is less immediately apparent than our subjectivity is no more an indication of its importance than is the fact that our concern for others is less immediately apparent than our physical hunger.

The Coextension of Character and Morality

The language of *scarcity* and *responsibility* have moral resonance and this tempts us to wonder if questions of character generally should be enveloped by morality. The argument might run in the following manner: the impoverished mother's obligation to her children requires that she forgo her own interests and our existing moral vocabulary is more than sufficient to characterize this. If she neglects her children to spend days engaged in folly, she is a *poorer* person for it. We needn't invoke a discussion of character or develop a separate vocabulary to understand this failing. The *character-defining* choice above is really nothing more than a renamed *moral imperative*. A parent has *obligations* toward her children that she violates if she doesn't provide for them. Other character properties will also likely match established moral predicates, and thus character is just morality by another name.

While morality and character certainly have *some* co-extension, the degree will be determined by how expansive our notion of morality is. Failing one's moral obligations will *always* be defining of character because character concerns one's response to her existential circumstance. Since our circumstance is one of a moral life that imposes certain obligations on a person, moral failures will constitute some species of character deficiency. Since even the most restricted moral conceptions presuppose some obligations to one another, failing or alternatively honoring these obligations will almost necessarily be relevant to character. However, if morality concerns just our obligations

to one another, then the domain of character is larger.[233] For character also concerns the category of permissible, but not required actions, and these lie beyond the focus of many moral conceptions.

Imagine a parent who devotes herself to her child so that this child can pursue an athletic or artistic endeavor to an unusual degree. This parent accepts lesser employment and poorer living conditions to move closer to the optimal training facilities for her child. We might well believe that such sacrifice is *not* required of this parent, that she would honor any obligation to her child by enabling her to train at a local facility. Just as it demeans such a person to hold that she must have acted on a hedonic motive (e.g. that she must have been motivated by the pleasure she took in helping her child), we also risk mischaracterizing her choice to hold that it was necessarily moral in nature. The profundity of her action may lie in it *not* being morally required.

---

[233] There certainly exist ethical systems that are not limited to other-regarding obligations. In *Justice for Hedgehogs*, for example, Ronald Dworkin distinguishes between morality and ethics where the former deals with the obligations we owe to others and the latter concerns how to live well (pg. 25). Such a conception would envelop at least some of what I'm calling character and place it within the realm of ethics. Dworkin's view differs from mine in a number of respects, including his view that we have an obligation *to ourselves* to live well (pg. 196). By contrast, I'm suggesting that we have the prerogative to be guided by self-making reasons and that being so guided (or failing to) can be the source of extra-moral judgments from ourselves and others. I'm indebted to Calvin Normore and Gavin Lawrence for reminding of me the need to distinguish such conceptions from my own. See Dworkin, Ronald. (2011). *Justice for Hedgehogs*. Cambridge, Massachusetts; London, England: Harvard University

Other expansive conceptions of ethical life could similarly place part of what I call character in the realm of morality. Barbara Herman suggests that we can see morality as a way of getting around in the world, where moral literacy requires a grasp of basic moral concepts, a practical awareness of the difference between persons, and the support of a variety of social institutions (pg. 23). One's character, on this view, is determined by the scope and content of her deliberative field where this is thoroughly latticed into her moral competence, or lack thereof (pg. 81, 97). See Herman, Barbara (2007). *Moral Literacy*. Harvard University Press.

For a more robust discussion of how my notion of character relates to morality, see chapter seven. For the way in which this intersects with Herman's notion of moral literacy see 214n 252, 222n257, 233n286, 236-238, n287-290.

There is, of course, moral language that one could use to try to capture what's normatively going on in this case, but it seems a stretch. Certain fully-directed moral views hold, for example, that a person is always in the grip of some moral obligation or other, such that this choice is morally required because the child's endeavor outweighs the parent's other commitments. But we might think this doesn't represent the extent to which this action was voluntary for her, especially if, for example, we believed positing boundless moral constraint over-determines one's action. Alternatively, one might believe that this parent's action supererogatory, since it was permissible but not required.[234] We think of the supererogatory as the realm of saints and heroes, something that would make someone not merely a morally good person, but a great one.[235] But while a parent sacrificing certain personal and professional comforts for her child's development is certainly noble, it might not rise to the level of being heroic. Moreover, this needn't be a matter of splitting hairs because it might be valuable to have distinct normative language for a noble person as opposed to a heroic one.

A great majority of our morality is concerned with what we shouldn't or must do, and thus the class of actions which we might, but needn't, perform can fall into a vast, undifferentiated category. Even moral conceptions that reserve a 'free space' for personal projects might be insufficiently nuanced to characterize this choice.[236] They might, for instance, lack a way to differentiate this parent's sacrifice from a less

---

[234] See Urmson, J. O.. "Saints and Heroes" in A. I. Melden (ed.), *Essays in Moral Philosophy*. University of Washington Press, 1958 (pp. 198-216). For a discussion of Urmson and the moral status of the supererogatory, see chapter seven here, "Supererogation and Character"

[235] Moreover, ethical theories that are limited to what we owe to one another often don't include a space for the supererogatory.

[236] Scheffler Samuel (2003) *The Rejection of Consequentialism*. Oxford University Press, Oxford

demanding personal project (i.e. they might see this parent's sacrifice as a personal prerogative, equivalent to taking up a new hobby). While we don't think this parent's choice was morally required, neither do we believe it was driven by either whim or hedonic motives. Moral conceptions that group agent-centered prerogatives into a space protected from moral demands might blur important distinctions between different kinds of projects. One way to avoid this would be to supplement an established moral vocabulary with language of character to provide a more nuanced characterization of our behavior. In this way, the question of whether a character-defining choice is enveloped by morality depends on how expansive one's moral conception is.

The intersection of character and morality provokes a further question about whether one's interest in her own character is intrinsically moral in nature.[237] It's a self-regarding interest and therefore distinct from the other-regarding concerns that we typically think of as moral. We hope that we act well *because* morality demands it of us, but we also have a self-regarding interest in being the kind of creatures that yield to such reasons. One might wonder if acting on the latter motive strips an action of its moral worth. The Kantian moral agent acts well *because* she recognizes that moral law requires as

[237] There is a question as to whether a duty to self could serve a moral role analogous to that served by self-making reasons. I set aside this question here, though I recognize that Kant might be interpreted as including some of these notions in the self-respect required in his imperfect duty for self-perfection. On Kant's imperfect duties, see chapter three.

much.[238] But an agent can be aware that she is an evolving creature with a character largely defined by her past and future choices. The knowledge that she has acted badly will diminish her self-respect and this gives her a self-regarding reason to act well.[239, 240] In acknowledging that she ought to do 'x', she acknowledges that she's diminished if she doesn't. Her failure, according to a measure she accepts, in a project she affirms, would diminish her character. What isn't entirely clear is whether acting to avoid this diminishment and maintain one's self-respect by itself entails that one is not treating moral law as the ground of her action, thus rendering her action morally neutral. If, alternatively, one believed that this concern to maintain her self-respect was simply a

---

[238] See "On the Value of Acting from the Motive of Duty" in *The Practice of Moral Judgement*, Barbara Herman. "One need not be indifferent to the possible satisfactions that a dutiful action may produce. It is just that the presence of such possibilities should not be the ground of the agent's commitment to acting morally. Overdetermined actions can have moral worth so long as the moral motive is the determining ground of action - the motive on which the agent acts." (pg. 21) One is tempted to probe the notions of moral ground and overdetermination here. One might act morally because moral law demands it, but one might herald the call of moral law because one is concerned with one's own character and wants to be something good. One wonders if the distinction between an appropriate and inappropriate ground of moral commitment persists as one peels back the layers on moral motivation, or if yielding to moral law because it is moral law should exhaust the analysis of one's motive in acting.

[239] Debate exists over whether Kant's duty of self-respect could plausibly serve as such a moral motive for this sort of project in an agent's life. In "Servility and Self-Respect" (*Monist* 57, 87-104), Thomas Hill notes that our duties to ourselves are a precondition of our duties to others (pg. 91) and thus a failure to honor one's duties to oneself may interfere with one's ability to honor one's duties to others. If that's true, duties to oneself may be fundamental to acting morally.

Kant famously notes that a person "...should pursue his end...always with consciousness of his sublime moral predisposition (which is already contained in the concept of virtue). And this *self-esteem* is a duty of the human being to himself" 6:435 (editor Mary Gregor's italics). There is, thus, ground to argue that a person's ability to see herself as an end and bestower of value is necessary for her self-respect and, thus, a moral duty to herself. One might connect this notion of self-respect with our imperfect duty for self-perfection, so that seeing oneself as a moral agent is *itself* something one owes oneself. Some, however, are critical of such a view of Kant's self-respect. (See Cynthia Stark, "The Rationality of Valuing Oneself" in *The Journal of the History of Philosophy* 35, pp. 65-82). While I believe there might be room in Kant's imperfect duties to locate such a notion, I don't take such a view to be a central part of the orthodox interpretation of Kant's ethics.

[240] Self-duplicity and rationalization can shield a person from recognizing that she's acted badly. In such a case, a person might enjoy the fruits of immoral action without the diminished self-respect I refer to above. The possibility that the desire to maintain self-respect is sufficient to motivate irrationalities such as self-duplicity underscores how powerful our desire (or self-making reason) is to see ourselves as good.

different way of articulating the supremacy of moral law, then a larger share of character motives would be intrinsically moral. The notion is that failing to act morally might only diminish her self-respect *because* she already recognizes moral law as overriding. All of this is just to say that a self-regarding concern for character might intersect with moral motivations in a multitude of ways and thus the co-extension of character and morality isn't a trivial matter.

But there's a sense in which wondering whether character is a subset of morality misses the point. The operative question is whether character is a measure that reveals an important dimension of our lived experience. If it reveals something of moral interest, we might expand our moral conception to include it. If it reveals something interesting, but morally irrelevant, we should treat it as a separate subject. If it doesn't reveal anything interesting at all, it needn't concern us. Whether character reveals 'morally-adjacent properties' or whether they're all properly subsumed under morality will have more to do with the extent of morality than the importance of character.

What's interesting about the case of the parent moving for her child is the extent to which it cuts between the ordinary opposition of other-regarding, moral obligation and self-regarding desire. She might have neither especially wanted to endure the poorer employment and living conditions, nor felt obligated to do so. She might have rather chosen to do this because she wanted to devote herself to a larger goal and protect her child through an arduous process. One of the benefits of developing a concept of character is that it gives us a lens through which we can meaningfully chart the realm of

the merely permissible; a neglected space in our moral discourse where we nevertheless spend most of our lives.

Endorsement

I should, at this point, further clarify the *value* to which I'm referring. One might challenge the intelligibility of the notion that one should develop her talents in accordance with that which she *values*, since, in some sense, all of her behavior is directed by that which she values. Actions are driven by intentions, which are fueled by reasons and, thus, the fact that someone *chose* to develop any capacity by itself indicates that she valued it in some sense. Otherwise, she'd have chosen to develop another capacity, or perhaps none at all. The thought is that since actions are intentional, they require preferences, which, in turn, express relative value. Actions *not* driven by preferences that reflect one's values wouldn't be actions at all. Thus, the claim that a person is motivated to cultivate a particular capacity because she *values* it can sound fairly trivial.

But intentional, preference-driven behavior might indicate one's *values* in at least two different senses. The aspiring poet described has, *ex hypothesi*, weighed various capacities she might develop in light of the kind of life they would allow and the kind of person she would become. She not only prefers poetry, she *endorses* that preference as a worthwhile use of her life. One might think of this as a *deliberative value*; an end

that a person *endorses* as being worthy of time required for its pursuit.[241] One need not enjoy or be predisposed to an end to endorse it and, in fact, this end might well be something to which a person is *prima facie* averse. All that's required for an end to be endorsed or deliberative is that a person avow it as being worthy for its own sake of the resources needed for its pursuit.[242]

We can distinguish these from *non-deliberative motives*, which motivate actions but are *not* consciously endorsed by the agent who acts on them. One might, for example, not explicitly consider what she values or ought to value. Let us assume our envisaged banker doesn't actually value banking itself and doesn't especially value the quantitative capacities that she has spent years developing. We're tempted to believe that she must value the security they allow and believe they are instrumentally valuable to her. She must, we reason, deeply value this security since a large portion of her life is devoted to providing it. But this might *not* be so. She might have chosen a career without

[241] Mention of 'endorsed desires' immediately brings to mind Harry Frankfurt's seminal piece on second-order desires and volitions, "Freedom of the Will and the Concept of a Person," *Journal of Philosophy*, Jan. 14, 1971 (pgs. 5-20). While my views, and surely any views in this area, are certainly informed by his work, they are not fully compatible, nor are they directed at the same subject. Frankfurt's principle interest is in identifying a concept of a person and a free will. In his view, one must have second-order volitions to be a person and one's will is free only when it is brought into conformity with her second-order volitions (pgs. 10, 15). Creatures who have second-order desires but not second-order volitions are not persons, for Frankfurt, and thus incapable of having a free will at all (pgs. 10-11, 14-15).

I am not here engaging the question of what makes a will free, nor am I supposing that having second-order volitions is a precondition of personhood. Moreover, the distinction between second-order desires (i.e. endorsements of one's first-order desires) and second-order volitions (i.e. the desire that a particular first-order desire becomes a person's will) is pivotal for Frankfurt and something I don't engage. My interest is not in locating necessary conditions for personhood, but in identifying factors that determine a person's character. Unlike Frankfurt, I'm disposed to believe that the possession of second-order desires is sufficient for personhood (a point for which I provide no argument here) and that one's character is determined by *both* one's second-order volitions and her second order desires (this point I do engage above). Nevertheless, my views in this area are undoubtedly informed by Frankfurt's distinction between first and second-order desires.

[242] Often, the scarcest of these resources is time, where the pursuit of a particular end (e.g. partnership with a particular person or a certain career) will be to the exclusion of other ends that might also be choice-worthy. This again reflects the scarcity required for choices to be determinative of character.

considering the effect it would have on her life and character. She might *not* have explicitly decided she'd structure her life around economic security because it was most important, or even considered the matter. She might simply have possessed strong quantitative faculties and proceeded along a path without inquiring into its significance in her life's trajectory, without ever having an all-things-considered, evaluative moment. It seems improper to ascribe to her the second-order volitions that *her interest in banking dictate her choice* or that *she exercise her quantitative capacities*. It rather feels more appropriate to say that she didn't have higher order desires and that she ended up taking the path of least resistance, even though she never actually had a desire to *take the path of least resistance*. This matters because it's a somewhat common lapse into a life, a failure to resist the drift from the main flow of a river into increasingly narrow channels.

When a person makes a life-structuring choice, it is determinative of character and this is equally true when person does so acting on a non-deliberative motive.[243] Rather than assuming the unreflective banker is a materialist (i.e. one who values pecuniary gain and endorses this value), we might think of her as having a *deficiency or lack of*

[243] This is a point of divergence from Harry Frankfurt's view. He sees the unreflective agent above as something akin to a *wanton*, one who is indifferent to the desires that move her to action. As a wanton, such a creature is not a person and, thus, not capable of responsibility. I'm skeptical that Frankfurt's distinction of first and second order desires and second order volitions sufficiently characterizes our practical deliberation about which talents to develop because I think the truth is messier than these distinctions allow. Gary Watson seems correct to hold that there is something arbitrary about Frankfurt's associating willfulness (and thus personhood) with second order volitions, as opposed to associating it with second order desires (or even first order desires). But it also doesn't seem obvious that our desires fall cleanly into three categories: desires for x, desires about our desires for x, desires that our desires for x become our will. Relatedly, I'm disposed to see the unreflective agent as both a person and capable of responsibility. On wantons being persons and the deficiencies of Frankfurt's distinction, see Gary Watson's "Free Agency," *Journal of Philosophy*, Vol. 2, No. 8, April 24, 1975, pp 205-220. On wantons being subject to responsibility and even especially so, see Marina Oshana's "Wanton Responsibility," *The Journal of Ethics*, Volume 2, Issue 3, September 1998 pp 261-276.

character. Such a person's development is not structured by her choice, but by the notable absence of an endorsed decision. In the extreme case, a person who never critically assessed any of her values or preferences and simply treated her dispositions as reasons would fail to have much character at all. More commonly, such a condition will admit of degrees, where different thresholds of character will require different degrees of introspection and deliberation. In this way, we might distinguish between a person who lacked a bit of character and a wanton who had almost no character whatsoever.

However, it's not the case that those who do not have a deficiency of character thereby successfully act on endorsed reasons after deliberating. For even among those who deliberate about the values on which they act, other hurdles must be overcome before they successfully endorse a value and then act on it. If the ideal picture of deliberation consists of reasonable desire-belief pairs that are variously weighed, this process can be stalled by interfering factors. A person might have desires, of the first and n-th order, but also be influenced by fears, neuroses, and unclarities which stymie her deliberation. The aspiring poet might have valued poetry more than security, but failed to endorse this value because she was *timid*. She might have doubted she had sufficient talent or even that she had the right to make such an audacious choice for her life. Alternatively, she might have endorsed poetry as her highest value, but still failed to manifest her will in action. This might have been because she was suffocated by *depression*, unable to focus because of *anxiety*, or simply *lazy*. But in each case, a psychological hinderance prevented her from acting on an endorsed value. We might

think of those who are, in this way, unable to (1) endorse values or (2) act on those which they do endorse as having an *impeded* character.

Even an impeded character will be subject to assessment, one's own and that of others. As is often the case in our normative evaluations, these assessments will be influenced by whether the impediments are seen as things that could have been reasonably overcome. If we believe the timid person could have easily pushed through her doubts, we might see her as *cowardly*. If we believe the person who failed to bring herself to action could have overcome her depression, anxiety, or lassitude, we might see her as *slothful* or *undisciplined*. On the other hand, if we believe that any of these impediments were beyond her control (i.e. because they were the result of early trauma, a lack of social support, or biologically imposed), we might see her as *tragic*. In all of these variants, the actor has a character, just one that was impeded from endorsing a value or acting on it if endorsed.

If a person deliberates and succeeds in acting on an endorsed value, she has a what we might think of as a fully functioning character. In such a case, her endorsement of the values on which she acts will follow from either a hedonic or a self-making motive. She might choose to become a poet or a banker because she thinks it will maker her happiest, where she sees happiness as the value most worthy of endorsement. Or she might choose a path because she thinks that it's what the world needs or because she believes it's beautiful, interesting, or noble. In such cases, we can read her endorsement off of her action and this makes assessment of her character clearest.

Of course, if character requires reasons, a threat of regress lingers. After all, if having character requires that one is not merely disposed toward the end that motivates her, but has a *reason* she endorses it, then should we not require that she have a further reason for this reason? One wonders if, as we peel back the layers of our action-guiding motives, we do not all run out at some point? Moreover, in celebrating those who have reasons for their reasons and reasons for those, are we not simply preferring a more contemplative and perhaps obsessive psychology?[244]

But we needn't always prefer more contemplation to less. A clever person's longer chain of reasons needn't imply that she has a finer character because not all deliberative foundations are equal. We might believe that someone who acts for pleasure should be expected to inquire deeper into the value pleasure has in a life, even when we are satisfied that one who acts to advance the ends of her loved ones has reasoned enough. It's perfectly conceivable that a shorter chain of reasons might even be preferable to a longer one, as is the case of *one thought too many.*[245] The very issue of how deep a person's reasons go is itself relevant to her character. At the extremes we find caricatures. On one hand, without reason a person lacks character entirely and on the other, excessively long chains of reason would prevent any action at all. But

---

[244] This worry has hints of a sorites paradox. We agree there's a difference between acting on reason and acting on unreflective inclination. We can disagree about how deep these reasons and endorsements must run without threatening our shared intuition that some reasoned endorsement matters. As with many alleged sorites cases, our concepts may be intelligible in the center even if they fray at the edges.

[245] In Bernard William's case of a man who deliberates more than he should before choosing to save his drowning wife over another drowning victim. Williams believes the case demonstrates that a moral system shouldn't require an agent to weigh all lives equally. I mention his example just because in William's hypothetical it's enough that he save one of the people drowning *because* she's his wife and no further reason is necessary. See "Persons, Character, and Morality," in B. Williams, *Moral Luck*, Cambridge: Cambridge University Press. (1981)

between the impetuous at one end and those paralyzed by deliberation at the other lie almost all of us, and this is where the gradations of character will be most revealing. It's within this range that we'll be able to differentiate the *timid* from the *brave* and the *skeptical* from the *assured*.

In summary, I've explored a species of life-structuring decision, the choice of which capacities to develop in oneself, in order to show how the reasons on which a person acts determine her character. As I employ the term, character is a function of a person's response to her existential circumstance. Because the capacities one chooses to develop recasts this existential question and because these capacities become the basis for the adoption of further associated values, I believe this choice is profoundly determinative of character.

I've argued that one is not, and should not, be guided in this choice by which capacities she believes will make her happiest or which are rarest or strongest in her because none is more likely to make her happy than another and the rarity or strength of a capacity is far less important than its content. I rather think one should be guided in such decisions by her interest in the object she will become and, accordingly, that she should develop capacities that she believes are valuable and appropriately responsive to her circumstance. I've tried to argue that, in the conditions of moderate scarcity required for the development of character, acting on what she takes to be valuable and responsive to her circumstance enables a person to become an instance of that which she values and, in doing so, writes that value upon the world.

I've claimed since character requires acting on valued ends, such action entails a certain kind of non-trivial endorsement. I believe that when one fails to do this because she acts on a non-deliberative motive, her character is *deficient*. When one deliberates but fails to endorse a value or endorses one but fails to act on it, her character is *impeded*. However, when one acts on an endorsed value she has a functioning character and because her choices are an expression of that which she values, we are able to infer what she *is* from what she *does*.

Lastly, I've tried to suggest that developing a richer concept of character might generate more nuanced language that we can use to chart a vast and largely undifferentiated category in our moral discourse: *the merely permissible*.[246] The different modalities of character described above are an instance of this, where the choices made are neither wicked nor noble, but the reasons that motivate them constitute telling hues of character. In what follows, I explore how character can help us understand other normatively interesting behavior that would otherwise fall into the often unsorted group of morally permissible action.

[246] By 'merely permissible' I'm referring to actions that are permissible but not required.

Chapter Seven
Supererogation and Character

In what follows, I try to clarify two related issues that remain after the preceding chapters. The first of these concerns the relation between character and morality. I've argued that a person has a self-regarding interest in the object she is and that a central part of this object, what I call her *character*, is constituted by a network of capacities and dispositions. Because this interest is self-regarding, we needn't appeal to traditional morality to find a motive for it. However, I've not only suggested that a person *has* an interest in her character, but that she *should* have such an interest and that some characters are finer than others. Such remarks can reasonably lead one to wonder whether questions of character are moral in nature. To this point I've mostly engaged this question obliquely (i.e. that one's interest in her own character need not derive from her concern for traditional other-regarding morality but that aspects of a person's character will be relevant to her moral fitness). Here, I try to clarify my view about the relationship between character and moral theory. At a general level, I think their relation has more to do with what we take the proper scope of moral theory to be than with the importance of character itself. More pointedly, I try to show that if we believe morality should be concerned with that which we can properly *demand* of one another, then many questions of character would be extra-moral but if we believe the scope of moral inquiry should be more expansive, we can find a place for character within this realm.

The second issue I address concerns the *range* of self-making reasons a person might have. I've argued that self-making reasons, or reasons that issue from one's concern for one's self as an object, motivate many of our life-structuring decisions. I also argued that self-making reasons can be formed by an appropriate application of one's capacities to the circumstances in which one finds oneself. Because the self-making reasons I focused on were those that structure large-scale decisions *and* because I emphasized the importance of being responsive to one's circumstance, one might have been left with the sense that self-making reasons must consist of *things one would do to improve the world* or *causes one might take on*.

While I think self-making reasons can be directed toward such ends, I certainly do not think these are the only form such reasons can take. I think this unclarity arises from having focused so centrally on a *particular* end toward which such reasons are directed, namely life-structuring decisions. To remedy this, there's a temptation to simply list other sorts of reasons that issue from one's character, in order to show that they needn't all be directed at problems in the world.

However, I think a more comprehensive solution is to enumerate some of the many *character properties* a person might seek other than the desire to be good or to improve the world in which one finds oneself. These latter properties generate a single sort of reason issuing from one's character, but different character properties would generate others. The underlying issue is not merely that there are other sorts of self-making reasons but, more profoundly, that *aspects of a person's character* that have not yet been discussed would tend to generate these other sorts of self-making reasons. Hence, instead of simply giving a list of reasons that are less world-affecting,

I think it will be more illustrative to explore some of the character properties that give rise to these reasons. I should also note that while, to this point, I've focused primarily on *praiseworthy* self-making reasons, there are, of course, equally many such reasons that are less laudable. I try to rectify this as well by considering a variety of character properties of which we might be legitimately critical.

I've chosen to discuss these issues together because they are related. Clarifying the relation of character and morality requires that we (1) understand the sort of properties that constitute one's character and (2) specify limits of traditional morality in order to see if character falls within that space. I come at this issue somewhat obliquely by considering first the limits of traditional morality as discussed by J.O. Urmson in his celebrated essay, "Saints and Heroes."[247] There, Urmson discusses a class of behavior, so-called *supererogatory actions*, that he believes don't find a place within traditional moral theory. However, I don't think his most profound contribution lies in identifying such actions, but in encouraging us to look beyond the limits of morality as traditionally conceived. When we do this, I think we find that his class of supererogatory actions are just one example of normatively interesting behavior that doesn't find a place within our traditional moral framework. I then provide examples of others, namely, the *extraerogatory*, *suberogatory*, *undererogatory*, and the *nonerogatory*, and suggest these are conceptually adjacent to Urmson's notion of the supererogatory.[248] I try to show that the common thread among these classes is that they do not follow from that which a person *owes* another, but rather from the sort of

---

[247] Urmson, J. O.. "Saints and Heroes" in A. I. Melden (ed.), *Essays in Moral Philosophy*. University of Washington Press, 1958 (pp. 198-216)

[248] En route to doing this, I consider and reject Julia Driver's notion of suberogation because I think it fails to capture the negative analogue of Urmson's class.

object she is. As such, I believe this behavior issues from one's character and is motivated by her self-making reasons (or a lack thereof). Examining such behavior serves two purposes. First, it enables us to see the aspects of character that give rise to self-making reasons that are *less* world-affecting than those I've discussed thus far. Second, considering these other classes of behavior alongside supererogatory actions allows us to understand more clearly what's at stake, if anything, in the relation of character and morality.

In his famous 1958 essay "Saints and Heroes" J. O. Urmson objected to what he saw as the inadequate 'trichotomy' of moral facts that pervades much of moral philosophy.[249] Almost all predominant moral theories, he believed, categorize morally relevant actions as either (1) duties or obligations, (2) permissible but not required behavior, or (3) prohibited or forbidden actions. Traditional morality, on his view, thus consists of *things we must do*, *things we might but needn't do*, and *things we must not do*. Urmson reasonably saw that such a taxonomy cannot capture the richness of our moral experience and, in particular, is insufficient to explain two related types of action, saintly and heroic behavior. He further held that any moral theory that couldn't find a place for such action (beyond the traditional threefold classification) was, by consequence, inadequate.[250]

It might strike one familiar with modern moral philosophy that Urmson's trichotomy doesn't exhaust *all* the possible modalities of duty. Kant's imperfect duties, for

[249] Urmson, 1958, pp. 198-216

[250] Urmson, 1958, pg. 199

example, are neither demanded without exception *nor* merely permissible but not required. Rather, as discussed at length in chapter three, Kant treats imperfect duties as defeasible obligations that an agent must honor when she has sufficient bandwidth, but which can be overridden by her perfect obligations. In addition, some interpret Kant as suggesting that a person has some discretion to determine the time and manner in which she satisfies these duties.[251] This might lead to a certain unclarity about the scope of imperfect duties and to potential disagreement about when they have been sufficiently discharged. Thus, it's reasonable to contend that Urmson's trichotomy doesn't find a natural place for some interpretations of Kant's imperfect duties.[252]

While this is true, and while one can find other examples in Aristotelian ethics and other moral theories, Urmson has nevertheless identified a general theme in our moral traditions. Even defeasible obligations, such as imperfect duties, still present themselves as *constraints* on a person. Behind the trichotomy Urmson describes lies a broader claim he is making, that traditional morality presents itself as *dicta* constraining what a person may do, and this is still present with imperfect duties. Though one might argue Urmson's distinction (i.e. that traditional morality consists of *things one must do, things one may but needn't do*, and *things one cannot do*) is not sufficiently

---

[251] See Thomas Hill's "Kant on Imperfect Duty and Supererogation," from *Dignity and Practical Reason*. Hill writes, "…one may avoid [acting on an imperfect duty] at *any* time (though not at all times) that one feels inclined" (pg. 159)

[252] After noting the scholarly tendency to interpret Kant's imperfect duties (obligatory ends) as allowing an agent discretion in the time and manner of execution, Barbara Herman proposes the more sensible interpretation that such duties are limited *only* by other duties. I follow Herman's interpretation of Kant's imperfect duties here, which is why I believe they *are* nevertheless subject to Urmson's critique. See Barbara Herman's "The Scope of Moral Requirement" in *Moral Literacy* (pp. 203-229, especially 213-214)

fine-grained, defeasible duties simply add a variation to this theme (i.e. there are also *things one must sometimes do, unless they're overridden by more firm obligations*.) Even if such duties allow some discretion in their execution, they are still external restrictions on action. Urmson's deeper thought is that traditional morality presents principles that seek to constrain a person's behavior, which, if violated, render the person immoral, and this is also true of imperfect duties. What Urmson is seeking, ultimately, is a moral category beyond external principles that constrain action, one that might allow us to evolve our moral judgment beyond good and evil.

Aside from fringe cases such as imperfect duties, it's easy to think that this traditional trichotomy covers the general categories of moral action. It's even tempting, especially for moral theorists, to see all of our behavior as falling somewhere on a continuum from that which we must do, to that which we can, but needn't do, to that which we cannot do. But if we abstract away, we can see that the trichotomy reduces to a dichotomy of actions. First, there is a category of compulsory behavior, whether the obligation is to perform a certain sort of action or refrain from performing another sort. Second, there is large, undifferentiated category of 'other' actions, which are neither prohibited nor required but are *merely permissible*. Failing to comply with the first category is immoral and behavior in the second category is morally neutral. This contributes to the tendency in moral philosophy to see morally relevant actions as either moral or immoral, good or evil, and to leave the rest of our behavior

unexplored.[253] What Urmson helps us see is that this entire continuum forms just a single dimension in a larger space of possible moral behavior.

Saintly actions, for Urmson, are those which most people's "inclination, desire, or self-interest" would prevent them for performing, but which some persons are able to perform either effortlessly or as a result of "exercising abnormal self-control" [254] Heroic actions are those which most people would not perform out of "terror, fear, or a drive to self-preservation," but which some persons, again, are able to perform either quite easily or out of an abundance of self-control. Both require an individual who either lacks the dispositions that restrict ordinary human behavior or has the ability to override them. They only differ in that performing *saintly* actions requires overcoming desire and self-interest, whereas heroic actions require overriding fear and the desire for self-preservation.[255] The unmarried daughter who stays to care for her ailing father

---

[253] This, admittedly, ignores a current in contemporary philosophy that sees morality not as merely a body of requirements, but a particular way of seeing the world and one's place in it. On such view, moral agency and action require a proficiency with (and concern for) basic moral concepts that presupposes a complex etiology of education and socialization. On such view, *moral literacy* is an achievement not just for an individual but also for a society because it has such deep historical and cultural requirements.

This view is *not* the target of Urmson's critique in part because it has resources to accommodate supererogation and other concepts that the traditional trichotomy cannot. According to this view, morally literate agents have some capacity to deal with new moral facts, such that a person's and culture's moral lexicons can grow in response to new circumstances and challenges. For this reason, I believe this view would be largely consistent with many of the character properties and extra-moral predicates discussed in this chapter.

On the requirements and features of moral literacy see Barbara Herman's "Responsibility and Moral Competence" in *Moral Literacy*, (pp. 79-105, especially 91-94 and 97-105) and her "Responsibility and Moral Competence" in the same volume, (pp 81-105, especially 97-99)

On an agent's ability to deal with new moral facts see Barbara Herman's "Can Virtue be Taught" in *Moral Literacy*, (pp. 106-129, especially 113-116) and her "Responsibility and Moral Competence." (pp. 104-105)

[254] Urmson, 1958, pg. 200

[255] Urmson, 1958, pg. 200

is saintly, for Urmson, and the terrified doctor who nevertheless remains to care for his patients in a plague-ridden city is heroic.[256]

Such behavior, for Urmson, falls outside the traditional moral classification because they're good actions, even great actions, but they aren't required of those who perform them. Because their performance isn't *required* of a person they can't properly be seen as duties, but because they're so obviously praiseworthy they seem to be *more* than merely permissible actions of no moral import. Urmson, thus, dubs such actions 'supererogatory.'[257] He admits those who perform such supererogatory actions can see themselves as obligated to perform them, but he holds that we could never demand such actions of a person, whether she be ordinary or extraordinary.[258] His stated reason we could not demand this of ordinary persons is familiar and uncontentious. But the reason he provides for not demanding such action of usually capable persons reveals an aspect of Urmson's distinctive moral perspective that's worth exploring because it informs the discussion of character given to this point.

Saintly or heroic actions, for Urmson, should not be demanded of ordinary persons simply because such actions might well exceed the capacities of such people. There is a strong element of *ought implies can* in this notion, where demanding more of a person than she's capable of would be unfair. But Urmson also believes that positing duties that lie beyond the reach of ordinary people would denigrate the concept of duty

---

[256] Urmson, 1958, pg. 200. Urmson's example of the doctor who stays to care for his patients in the plague-ridden city strikingly resembles Dr. Rieux in Albert Camus' groundbreaking existentialist novel from the same period, *The Plague* (1947).

[257] Although Urmson is sometimes credited for introducing this frequently used term into contemporary moral discourse, he only uses it once in this essay (Urmson 1958, pp. 214).

[258] Urmson, pg. 203, 210, 211, especially 214

by preventing duties from being *rigidly* obligatory. “The basic moral code,” he writes, “must not be in part too far beyond the capacity of the ordinary men on ordinary occasions, or a general breakdown of compliance with the moral code would be an inevitable consequence…”[259] It’s clear, for Urmson, that duties must remain firm obligations in order to retain their essential character, but he also believes that morality must be in some sense tailored for those whose behavior it’s intended to regulate. Thus, “…a moral code that would work only for angels would be a far from ideal moral code for human beings.”[260] Because he holds that a moral code should help contribute to human well-being, Urmson further believes it must be “…formulable in rules of manageable complexity,” and sensitive to the limits of ordinary persons.[261] If we want duties to be firm and we also want our morality to be sensitive to our limits, it makes sense that duties should not be so overbearing that ordinary persons could not satisfy them and, by consequence, we should not demand heroic or saintly behavior of them.

But this doesn’t explain why those who possess greater capacities couldn’t, in virtue of these capacities, bear greater obligations than ordinary persons.[262] While it’s important that the concept of duty remains firm and that it doesn’t demand more than we can undertake, if duties are supposed to be tailored to our abilities it would seem that a system of minimal duties might demand too *little* of the most capable among us. After all, since more capable persons could be reasonably expected to perform more

[259] Urmson 1958, pg. 212
[260] Urmson 1958, pg. 211
[261] Urmson 1958, pg. 210-211
[262] Urmson 1958, pg. 212

onerous tasks, requiring more of such persons might not overburden them. One might, along these lines, think that a restricted system of duties might be contrary to Urmson's aim of tailoring morality to *all* those it is supposed to serve.

By way of explanation, Urmson claims that a system of duties must not be excessively complex and that "…it would be absurd to try to formulate complicated rules to determine in just what circumstances such [a heroic] action is a duty."[263] But the absurdity doesn't seem obvious. If we believed, for example, that one had a provisional duty to assist others in dire need when such help wouldn't be burdensome or dangerous, it doesn't seem especially 'complicated' to specify that such duties should scale proportionately to the actor's abilities. We might, thus, require a person of ordinary ability to save a child drowning in three feet of water, but not expect her to save someone drowning a few hundred feet off-shore. However, we might, nevertheless, believe an extraordinarily good swimmer would have some obligation to save a person drowning off-shore in virtue of her capacity to swim very well.[264] In this way, scaling duties proportionally to a person's abilities might not be overly complicated and, it seems, we have at least *some* intuition that it could even be appropriate. Urmson is nevertheless clear that people of greater abilities should *not* thereby have greater obligations. He claims it would be "horrifying" to pressure a person to perform an act of heroism and "…a moral outrage to apply pressure on [a person] to do such a deed as sacrificing his life for others."[265] This notion seems

[263] Urmson 1958, pg. 213

[264] Here I am assuming that the person provisionally obligated to save the drowning party was a sufficiently good swimmer that it would neither cause her substantial risk or inconvenience to save the victim, but not that she was a lifeguard or had any special relation that would obligate her to save the drowning party.

[265] Urmson 1958, pg. 214

contrary to the intuition that we can reasonably expect more of the most capable among us, but his reason for this both assuages this intuition and reveals a distinctive way of thinking about duty.

Urmson takes himself to be following Mill in holding that if duties are like debts that can be demanded of people then they should be restricted to the "…minimum requirement of living together."[266] The scope of duty then is doubly restricted, limited first by ordinary people's abilities, and further to the *minimum* required for peaceful coexistence. This second limitation allows us to see why Urmson would not require more of the most capable persons. If duties did scale to fit people's capabilities, they might constitute a great deal more than the minimum required for peaceful co-existence. Restricting duty to this minimal standard constrains the moral demand on a person to potentially much less than she's capable of providing. Taken together, these two limitations allow us to see that duties might demand the most of those with fewest capacities, moderately burden ordinary persons, and demand relatively little of what the most capable among us have to offer. Duties, in this way, become the moral equivalent of a modest regressive tax.

This concept of duty, which Urmson takes to follow from *what we owe to one another*, leaves a great deal of our behavior neither required nor proscribed. An action is properly immoral, in Urmson's sense, only when it doesn't meet minimum requirements for peaceful cohabitation. Conversely, an action's being moral doesn't amount to any more than satisfying the basic requirements of social interaction. But there's a sense in which this is the *beginning* of the moral story, not the end. We are,

[266] Urmson 1958, pg. 209

after all, more likely to achieve consensus about the minimum requirements for peaceful cohabitation and therefore need less moral direction at that end. Where we look to morality for guidance is in the large class of *permissible but not required* actions that traditional morality leaves largely undifferentiated.

There is a question as to whether this body of behavior should be considered morally neutral or something more. This set might be smaller for less capable persons (i.e. those who struggle to even honor the minimum duties required of us) but it would seem to be quite large for very capable people, leaving them with a significant freedom from moral requirement. It's unclear, at least initially, what we expect people should do with this latitude, but it would seem that we'd think *better* of capable persons who used it for productive ends. Whereas the traditional trichotomy doesn't have a place for this, Urmson's contribution lies in suggesting a new ethical modality. He writes,

> A line must be drawn between what we can expect and demand from others and what we can merely hope for and receive with gratitude when we get it; duty falls on one side of this line, and other acts with *moral value* on the other…[267]

Urmson isn't calling for a freedom from morality to pursue personal projects of the sort that Samuel Scheffler and others have more recently argued for, but rather an entirely different moral space. Duties, by their nature, apply external pressure and require moral agents to work from a sense of constraint. While Urmson concedes that such pressure is necessary for the basic behavior that allows for peaceful interaction, he believes that "…free choice of the better course of action is always preferable to action under pressure…" and that, to the extent possible, praiseworthy behavior should be encouraged and not demanded.

---

[267] Urmson 1958, pg. 213, my italics for emphasis

Urmson, thus, wants an area beyond duty, one in which we can praise people for acting a certain way, even if we do not see them as immoral for failing to do so. Whereas on the traditional view a capable person who ceaselessly uses her talents in the service of others would be acting *permissibly*, Urmson's recognition is that we intuitively believe such actions are clearly more than morally neutral. Even characterizing them as praiseworthy or morally upstanding seems insufficient, as such dedication and sacrifice seems admirable and inspiring. But in order to even develop a language to characterize this behavior, we need to look beyond morality as merely a system of obligations we owe to others.

Urmson provocatively claims that saintly and heroic behavior are just two examples of "… a whole realm of actions…" that fall outside the realm of traditional moral theories.[268] He notes that, "…being a little more generous, forbearing, helpful, or forgiving than fair dealing demands…" would constitute such behavior.[269] These comments would seem to provide fodder for a broader discussion of supererogation. However, in the thirty years following the publication of "Saints and Heroes" Urmson grew so frustrated with the discourse on supererogatory actions that he asked to be "disassociated" from the notion and from those who believe there is a singular class of such actions.[270] His frustration was two fold. First, he believed that many of those who discussed supererogatory actions, R.M. Hare in particular, simply deemed them a

---

[268] Urmson 1958 pg. 205

[269] Urmson 1958 pg. 205

[270] Urmson J.O., "Hare on Intuitive Moral Thinking" in *Hare and Critics: Essay on Moral Thinking*, Oxford University Press, 1998, pp. 161-169, (especially pg. 168)

higher level of duty that could be demanded of more capable persons.[271] While this differentiates them from the minimal duties required for social cooperation, these actions would remain obligatory for very capable persons and therefore would close the space Urmson was trying to open. In such a case, morality would still consist predominately of obligations to principle, even if these were sensitive to people's different capacities, and we would still just have two classes of persons: moral and immoral. Such a response simply assimilates saintly and heroic behavior into the traditional framework, when Urmson was citing these as instances of a new category of moral behavior.[272]

Second, he felt that much of the commentary on supererogatory actions treated them as a single type, when he felt strongly that the class contained significant heterogeneity. Urmson writes, "…[W]e now have, instead of the old over-simple trichotomy of acts into right, neutral, and wrong, a new over-simple tetrachotomy…"[273] The class of supererogatory actions, on his view, became a way of classifying saintly and heroic behavior when he believed these were just two examples of an entire spectrum of morally significant behavior that was in some sense praiseworthy but not obligatory. Years after introducing the notion, Urmson provided more examples of supererogation including behavior that was "…kind, considerate, chivalrous, charitable, neighborly, sporting, decent, or acts of self-denial or self-abnegation…"[274] and noted that such actions are distinct from one another. Lending a bag of flour to a neighbor

---

[271] Hare, R.M. "Prudence, Morality, and Supererogation" in *Moral Thinking*, Clarendon Press, Oxford, 1981, especially, pp. 188-205, (especially 192-205)

[272] Urmson pg. 168

[273] Ibid

[274] Ibid

isn't chivalrous, Urmson points out, and helping an elderly woman across the street isn't sporting. "If we are to do justice to the rich complexity of moral life," he writes, "… we need a much more rich and varied set of concepts…"[275]

I think Urmson is right that if we are to adequately characterize our experience, our moral vocabulary must extend beyond obligations and prohibitions and this prompts one to wonder whether such inquiry would be moral or extra-moral. On one hand it might be reasonable to restrict 'properly' moral language to the traditional trichotomy of actions, in which case Urmson's class of behavior should be thought of as extra-moral or morally-adjacent. Alternatively, if we should see morality as much more than a system of duties and constraints on action, Urmson's class of behavior might constitute a promising area of moral inquiry.[276] But I think, in the end, this question becomes largely semantic. What's important is that there seems to be a class of behavior that is determinative of the people we are and the lives we lead that isn't sufficiently characterized by our existing moral vocabulary. If this behavior is relevant to our estimation of people and lives, as indeed it seems to be, it is of normative interest. Whether we call this a moral category or an extra-moral category might have

---

[275] Ibid

[276] Those who take moral education and literacy seriously will tend to see the development of new moral concepts as a perfectly ordinary part of moral theory. My thought is not that character related concepts would be new moral facts by themselves, but rather that they could evolve into moral categories that would allow us to access new moral facts.

Along similar lines, Barbara Herman notes that there are two kinds of new moral facts; "…facts that were there but were conceptually inaccessible… and things whose moral significance is of a new kind" (pp. 108). She mentions that the former sort are often 'new' because our prior moral concepts prevented us from seeing the moral importance of certain states of affairs. Herman argues that we might need to see gender as a morally relevant category, for example, to understand the wrongness of certain forms of pornography. (pp.112) I believe that fleshing out the moral significance of character properties, both positive and negative, would similarly put us in a position to see new moral facts that would otherwise remain veiled. See Barbara Herman, "Can Virtue be Taught? The Problem of New Moral Facts" in *Moral Literacy* (pp. 106-129, quoted on pp. 108)

less to do with whether it's worthy of further inquiry than to do with whether we believe morality should be limited to specifying obligations between persons.

What's more significant, I think, is the potential landscape revealed when we move beyond the traditional trichotomy. Urmson has identified a number of *non-obligatory, praiseworthy behaviors* that deserve our attention, but these constitute just one class of action in this space, when others exist. One of these is what we might think of as a negative analogue of Urmson's class, namely *behavior which is permissible, but nevertheless legitimately subject to critical assessments*. A well-known, but I think unsuccessful, attempt to identify this class is found in Julia Driver's "The Suberogatory." There, Driver argues that if Urmson is correct in thinking that there's a class of *supererogatory* actions, then there should also be a class of actions that fall short of being forbidden but are nevertheless blameworthy. Such actions, for Driver, are 'permissible,' but 'worse than the situation calls for,' things that we 'ought not do,' and are 'morally bad.'[277]

She provides three examples of such actions. In the first, a traveler, when boarding a train, sees a couple also boarding and instead of allowing them to take the last two seats together, takes one of those for himself, thus preventing the couple from sitting together. On Driver's view, the couple has no claim to the two seats and thus the traveler's action is not forbidden, but he has done something blameworthy and morally bad.[278] In the second, Bob is suffering from kidney failure and needs a kidney from the only compatible donor, his brother Roger. Again, for Driver, Bob has no right or

---

[277] Driver, pp, 290, 291

[278] Driver, pp, 286-287, 291

entitlement to Roger's kidney so if Roger refuses to donate the kidney, he is not violating any duty he has to Bob and has done nothing wrong. Nevertheless, she believes Roger *ought* to give his brother the kidney and would be deserving of blame or reproach if he did not.[279] In the third, Albert has done Bill many favors in the past but when Albert asks Bill to return some library books, Bill declines, even though it would cause little or no inconvenience for him. Driver concedes that Bill *owes* Albert a favor, but unless Bill explicitly agrees to take the books back, he's under no obligation to Albert, and thus has done nothing wrong. Like the prior two cases, she believes Bill would be appropriately subject to negative judgment because he's done something morally bad, though not morally wrong in refusing Albert the favor.

Driver's cases have frustrated some because it appears they could possibly be handled by existing moral concepts.[280] One might believe, for instance, that Bill has implicitly entered into a relationship of *reciprocal obligation* with Albert by having accepted favors from him in the past, and that Roger has a *defeasible obligation* to give his kidney to Bob, since they have a familial bond. One is also tempted to wonder if the seat-grabbing traveller might have violated a precept of *etiquette* and not morality *per se*,[281] or that he too has a defeasible obligation (i.e. always let groups sit together, unless it would cause undue hardship) that could be nullified by a reason that he needed that seat (e.g. he would be traveling a long distance and that seat was more

---

[279] Driver thinks of this as a 'morally charged situation' because Roger is forced to choose between two alternatives, one of which would be heroic while the other would be blameworthy (pp. 287-288)

[280] See Hallie Liberato, "Denying the Suberogatory" *Philosophia* (2012) 40: 395-402. Liberato suggests that Albert and Bob might have reciprocal obligations to one another as one reason that Driver's concept of suberogation is unnecessary.

[281] Foot, Phillippa "Morality as a System of Hypothetical Imperatives", *The Philosophical Review*, 81(3): 305–316, (pp. 308)

comfortable). If these cases can be sufficiently dealt with by existing moral concepts, Driver's concept of suberogation would be unnecessary.

But I think the deeper issue with Driver's version of suberogation is that it runs afoul of Urmson's initial worry, namely, it tries to incorporate the concept into traditional moral frameworks. Her intuition begins at a reasonable place, that even if an action must violate a right or obligation to be morally wrong, we believe other behavior can still be legitimately subject to criticism. However, because she treats suberogatory acts as 'morally bad' and 'things we ought not do', she's forced to make some awkward claims; that, for instance, what a person has done is *morally bad* but *not morally wrong*, and that a person *ought* to act in a different way but she's *not obligated* to. However, it's difficult to understand how an act could be morally bad and worthy of blame and yet *not* be morally wrong, or why one ought to do something which she has no obligation to do and isn't wrong not to do.[282] The attempt to push another kind of moral category into the our existing moral framework forces us to split concepts that are ordinarily coextensive (e.g. morally bad and morally wrong) and to run others together (e.g. that it's not wrong to take the seat, but one ought not take the seat).

Another effect is that to the extent suberogatory acts ought not to be done and are cause for reproach, they're still *constraints* on one's action. As such, to some extent, they become obligations by another name, which is what Urmson sought to avoid. For

---

[282] Driver claims that the distinction between 'one ought to do x' and 'one is obligated to do x' is *now familiar*, but it doesn't seem uncontroversial. There might be things one morally ought to do that aren't obligations, if obligations are only owed to individuals. One ought to be environmentally conscious, though it isn't clear that one owes any particular individual the obligation to do so. But we might also think of obligation more broadly, such that one has an obligation to all extant people collectively to be environmentally conscious, or an obligation to future generations, or to oneself. If one wanted to express oughts in terms of obligations, it does seem that we could expand our notion of obligation to do so.

Urmson, we aren't obligated to perform supererogatory actions but, crucially, there's also no suggestion in Urmson that we ought to, or that we're in any sense worse for not having done so. His is an effort to get us away from the shouldn't/ought, good/evil of traditional morality to something that prompts moral action through encouragement not constraint. Thus, I don't think Driver's notion of suberogation is useful in advancing our moral discussion.

But I think there is an important intuition underlying Urmson's concept of the supererogatory and hints of it even in Driver's notion of the suberogatory. It's that traditional moral language doesn't allow us to express the range of normative judgments we'd like to make. At the upper end of a spectrum between good and evil, Urmson sees actions that aren't just morally good, but morally great, and we have no way of expressing this in our traditional architecture. Morally neutral actions lie in the middle of that spectrum, but it does seem there might be things that fall short of being morally neutral but aren't *immoral* in the sense we normally intend, and we don't have good tools to characterize those either.

These include a negative analogue to the behavior that Urmson focuses on, namely behaviors that are permissible but which are nevertheless subject to legitimate criticism. Let us again imagine a very capable person whose talents could help others. If this person tirelessly used her talents to help others, I think Urmson is correct in noting that we'd think *better* of her. Along these lines, if she used just a bit of her talents to help others we might have a neutral assessment of her. Correspondingly, if

she didn’t help others at all and instead used her talents for her own benefit, we might think *less* of her.[283]

Let us assume this person has satisfied her basic duties and so we shouldn’t think of her as being immoral as a result, as that would assume a claim on her talents that would violate her autonomy. If we don’t see her as being immoral, our traditional architecture asks us to see her as morally neutral, but we’re not inclined to see her this way because we are in some sense disappointed in this behavior. Urmson gives us a class of praiseworthy assessments of permissible behavior that fall short of sainthood, but, again, we aren’t inclined to see her behavior as being praiseworthy. There seems to be a considerable distance between seeing a person as immoral on one hand and morally neutral or praiseworthy on another. We want some way of capturing the intuition that she hasn’t violated any obligation to others, so she’s *not* subject to moral reproach, but we think less of her for failing to put her capacities to greater use than we would otherwise.[284]

One might resist this impulse to make a negative assessment, perhaps because it could appear to be a covert way of assigning duties to more capable persons. The thinking is that such assessments place an expectation on more capable people, because if they don’t perform certain actions they’re subject to a species of criticism. Along these lines, one might think such assessments are just toned down charges of immorality, which is reminiscent of Urmson’s first criticism of the discourse on

---

[283] It’s notable that Driver explicitly denies that such a case would be suberogatory her in her view (see Driver, pp. 287, 288).

[284] Here I differ importantly from Driver’s view in that I don’t think such a person would be blameworthy, whereas Driver emphatically does. (See Driver, pp. 291)

supererogation and my criticism of Driver. But I think this response fails to distinguish between what we can *demand* and what we can *hope for*, and implicitly rejects the notion that our characterizations of human behavior can extend beyond good and evil.

As Urmson notes, an advantage of having a restricted set of duties is that it gives us a group of behaviors that can properly be *demanded* of others, where one who doesn't comply with such duties is properly seen as immoral. But beyond this set lies a considerable range of behavior in which we have a legitimate interest, even if we *cannot* demand it of others. That very capable persons direct their talents toward productive ends is just one of a variety of hopes we have of other's behavior. We also hope for a variety of different kindnesses that extend beyond that which we can reasonably expect. We hope that others will be patient with us and even go out of their way to help us pursue our ends, even though we cannot expect them to do so. When we move to a neighborhood all we can require is that those around us respect our privacy and basic entitlement to live peacefully. But we hope for kind and considerate neighbors who can offer occasional assistance and even the possibility of friendship. We're a bit disappointed when those around us are somewhat terse and extend nothing more than the minimum required. In the bustle of our daily lives, we hope others will hold the door when they see we're carrying groceries and that if we sit despairing and forlorn at our desks that others will offer a supportive smile and ask if we're alright. The fact that these things exceed what we can demand of others doesn't diminish our hope for them or remove the tinge of disappointment we feel when we're denied them.

It's telling how much our hope for and disappointment in others extends beyond what we can reasonably *demand* of them. We don't think of those who deny us this treatment as immoral, nor do we think of those who grant it as heroes. The cold neighbor and distant coworker aren't evil for being such, but we do have a lower estimation of them than of their analogues who are kind and supportive. To make space for this we need normative concepts and corresponding language that extend beyond obligation and prohibition. Such concepts could encourage behavior we'd hope for and discourage that which we'd like to avoid. This would open a space where we could make assessments of others, both laudatory and critical, that extend beyond the language of duty.

If we do allow a space for permissible actions which are nevertheless subject to critical assessments, the question becomes what these assessments would be, or more generally, how we would map the considerable ground between praiseworthy and immoral behavior. The assessments we make will depend on a person's reasons for acting. If the envisaged capable person both knew that she was able to help those in need and chose to serve herself instead, we might think of her as *selfish*. If she didn't realize her talents could be put to more productive use, we might think of her as *myopic*. If she knew she was capable of doing more and yet chose to do nothing, we might think if her as *lazy*.

The more we can understand about her reasons, the more nuanced our assessments can be. It might be that she knew her capacities could be productively used to help others *and* she was disposed to use them to this end, but found herself afraid of extending herself in an unfamiliar way. In such a case, we might see her as

timid, or, if we believed that her reticence was unjustified and that any reasonable person could have found the gumption to act in accordance with her convictions, we might see her as *cowardly*. While some of these evaluations are indeed critical (i.e. that a person is selfish, myopic, lazy or cowardly), others are less so (i.e. that a person is timid) and we should remember that such characterizations would lie alongside Urmson's class of laudatory judgments for praiseworthy actions that nevertheless fall short of being saintly (e.g. one who volunteered for a few hours a week in the service of an important charity might be seen as *generous*).

More profoundly, it's important to note that the landscape between sainthood and immorality is not exhausted by permissible behaviors that we'd want to *encourage* (e.g. Urmson's class including kind, chivalrous and sporting behavior) and those we'd want to *discourage* (i.e. the class including myopic, selfish, and lazy behavior). There exists a class of assessments we can make of permissible behavior which could be *either* critical or laudatory depending on further context. If, for example, a capable person didn't help others because she used all of her available resources to pursue a scholarly endeavor or mastery of a musical instrument, we might think of her as *devoted*. Such dedication might be deserving of either praise or criticism depending on other facts about her circumstance (e.g. whether she was mastering this instrument in the midst a a great famine). There are also assessments of permissible behavior that will be *neither* laudatory or critical, but rather illuminate behavior across another dimension. If, for instance, a person abstained from intercourse and libations for spiritual reasons we might think of her as as *pure*. Such purity would be a distinctive and important part of

understanding this person and her other behavior, but the purity itself needn't be either positive or negative.

If we extend use of Urmson's '___erogatory' convention (though given his own misgivings and the abuses of the term, it might be best to find new language), we have identified at least four classes of behavior that are normatively interesting and lie adjacent to Urmson's class of supererogatory; that which is *extraerogatory*, *suberogatory*, *undererogatory* and *nonerogatory*.[285] Put another way, we've considered permissible behavior (1) that which exceeds what we could demand but falls short of being saintly or heroic, (2) that which disappoints us, (3) behavior whose significance is unclear without further context, and (4) behavior that neither meets our hopes or disappoints us, but is nevertheless distinctive.[286]

This puts us in a position where I can explain why I believe such evaluations, both those that are clearly laudatory or critical and those that aren't, belong to the realm of character. Such assessments refer to qualities that exceed what we *owe* one another. Instead, they describe the kind of *things* we are, and this is the essence of character.

---

[285] The root of 'supererogatory' derives from the latin term 'rogāre', or that which is *asked for* or *requested*. In this sense, my use of of 'extraerogatory' and 'suberogatory' respectively refer to that which exceeds and falls short of what we'd asked for, 'undererogatory' refers to behavior whose significance isn't fixed without further context, and 'nonerogatory' refers to behavior that we wouldn't ask for but is nevertheless normatively significant.

[286] My use of 'extraerogatory' here refers to behavior that we hope for and cannot demand, but which falls short of saintly or heroic actions. My intention here is to preserve Urmson's original use of 'supererogatory' to refer to saintly or heroic actions and still make space for another class of actions that exceed what we could demand but fall short of being supererogatory. Urmson admits that this latter class, what I've termed extraerogatory, is normatively interesting but distinct from the supererogatory. (See Urmson, 1998, pp. 168)

My use of 'suberogatory' differs from Driver's more significantly, in that her class consists of actions that are morally bad and ought not to be done, whereas my use simply refers to actions that fall short of what we would generally ask of others.

As discussed in chapter six, a person's character is determined by the application of her capacities to her circumstance. Just as we have different expectations of a person born into a land of plenty than we have of one born into a land of scarcity, we will have different expectations of persons with vastly different capacities. We will tend to expect more of a person who is very capable, even though we cannot demand more of such a person. These expectations form the basis of the aforementioned assessments we make.

Character is, admittedly, also informed by a person's response to duty, as one's basic moral fitness is clearly relevant to the kind of thing she is. But two persons of equal moral fitness need not have equivalent character as a result. If we imagine two very capable persons who each adhered to all of their basic duties, one of whom did nothing more while the other was usually *good-natured* to those around her, we'd tend to have a higher regard for the latter. A productive way to express this is by noting that she has a finer character.

There's some temptation to believe that whereas the former is *minimally* moral, the the latter is somehow *more* morally fit, though still not a saint. But that still leaves a single hierarchy that doesn't allow us to make important distinctions. Imagine an equally capable third person who also adhered to all of her basic duties, but instead of being unusually kind to those around her, was so devoted to a charitable cause that others found her *inspiring*. We'd tend to think of her more favorably than the capable person who did nothing more, but it isn't clear that we'd think of her more favorably than the person who was unusually kind to others.

One might argue that there are grounds to rank these two. One might believe local kindness should count for more or, alternatively, that the person devoted to a noble cause had a greater overall impact.[287] While we can be sympathetic to such responses, we should see that they are attempts to force a comparison; to push two points in a matrix into a single line. They're driven by our impulse for moral hierarchy, which, in turn, derives from well-ingrained dichotomies (e.g good/evil, virtue/vice, moral/immoral). These oppositions form poles in our thinking and tempt us to sort behavior along a continuum in between. In doing so, they isolate aspects of complex phenomena, whether they be persons or actions, and compare them in reference to a single property. This sort of comparison can be important, even vital, when evaluating different potential actions. But it is a filter on our moral experience and, like any filter, it becomes a limitation if it becomes the only way we see the world. Allowing more properties, more moral predicates, pushes against this tendency and helps us see persons and actions in their complexity.

We might instead think that all three of these people are morally fit, but that the latter two have finer characters that first. This allows us to express our intuition that they are more deserving of our esteem than the capable person who did nothing with her talent, but it also doesn't force us to compare our esteem for the latter two or to rank them unnecessarily. Thinking in terms of character, as opposed to degrees of moral fitness, allows for a variety of assessments that needn't submit to a single

---

[287] Barbara Herman makes a case for the primacy of the local, where 'local' is not merely a function of distance but also of social connection in discharging obligations. Nevertheless, she also notes that, consistent with Kant's ethics, the failure of society to provide for its citizens could cause the burden to fall upon the shoulders of individuals as a 'secondary obligation' of beneficence to be discharged after her other obligations have been satisfied. See Herman's, "The Scope of Moral Requirement", (pp. 222-225)

hierarchy, thus enabling us to articulate our intuition that each of these persons is more deserving of esteem than the first *but in different ways*. We might think of the good-natured neighbor as being *kind* and the inspiring charity worker as being *selfless*. Both the extraerogatory and suberogatory behavior detailed here admit of significant heterogeneity. Permitting this heterogeneity and suspending our propensity for rigid moral hierarchy allows us to describe a more diverse array of human behavior and to make more nuanced assessments of others.

It's fair to ask why a more descriptive moral vocabulary is important. The answer, I think, turns on what we take the purpose of moral judgment, and moral theory generally, to be. If the purpose of moral theory is to provide norms for acceptable behavior in the service of social stability, moral judgment could reasonably restricted to distinguishing good from evil, right from wrong, and there might be little utility to a broader moral vocabulary. But if the function of moral theory is broader, we would want moral judgment to extend further. If, for instance, we believed moral theory should provide a richer understanding of human behavior, not just to promote peaceful coexistence, but so we can have greater insight into the kind of lives we're living and the kind of creatures we are, we would want moral judgment to characterize a broad range of behaviors, not merely distinguish between certain well-established poles.

It's tempting to think the value of this greater insight lies in enabling us to better understand others, but I think its principal value lies rather in enabling us to build a concept of ourselves. When we're considering the thing we want to become, the question of whether to be good or evil is hopefully answered straightaway. But after this question is answered, most of what a person will become is still undefined. The

trichotomy doesn't offer a great deal of direction as to the value of courage, humility, or sanguineness, and yet these qualities would be constitutive of a person who possessed them. Developing a vocabulary that describes character properties could give us a more robust conception of what we are and aspire to be.

It must be conceded that these concepts (e.g. timidity, lassitude, kindness, and courageousness) haven't been explicated here in anything but the most preliminary fashion. There's no doubt that the concepts themselves are underdeveloped relative to our established dichotomies, as is their relation to one another. It isn't clear, for example, whether *chivalry* is an instance of the patriarchal subjugation of women or an effort to redress existing gender imbalances, nor it is obvious whether it's a species of kindness or of etiquette. The concepts will likely also have complicated limiting relationships to one another. Excessive timidity, for example, could limit the extent to which a person could be kind or neighborly. But rather than being a reason not to explore such concepts, this indicates that there's interesting moral (or quasi-moral) work to be done.

Groups of these properties will tend to cluster with one another forming constellations in a person's character. Our general understanding of a person will be given by the clusters of qualities she possesses, but a person's unique character will be often be found in the manner in which she departs from the typical structure. Kindness, patience, and empathy, for example, will tend to rise and fall together, but the relationship needn't be fixed. A person might be kind and empathetic, but lack the degree of patience typical of others who have these other qualities. In some sense that dearth of patience, which will no doubt limit the extent to which she can be kind and

empathetic, will be a more distinctive aspect of her character than the standard configuration.[288]

Because the properties and clusters stand in a complicated relation to one another and turn on background information we often won't have when evaluating others, we cannot hope they will give a comprehensive insight into all those around us. But this manner of seeing character gives substance to what it means to *know* a person. Such understanding requires not just knowing a person's motive for a particular action, but seeing how this motive issues from clusters of moral properties in her, which in turn stand in complex relations to each other. In our acquaintances we might see just a few such relations, but the emotional intimacy of friendship or partnership lies in seeing a greater number of them. To *know* someone is not merely to know their reasons for actions, but to know how their reasons cluster with one another, what constraints they're operating under, and the way in which their eventual behavior is often the product of a complex deliberative interaction. It should be clear that such understanding of another's character must always be regarded as provisional, not just because we lack anything like comprehensive access to a person's reasons and moral properties, but also because these properties are, at least to some degree, in flux. This diminishes the hope of having a static concept of a person, but it is the inevitable consequence of a person being a thing-in-process.

---

[288] When discussing a person's need to 'self-shape' her character, Barbara Herman makes reference to similar sorts of character defects. She writes "[w]e each come to adult agency with a mix of tendencies: some beneficial, some inclining us to cause harm." Her point is that we have a responsibility to recognize these and become "trainers of ourselves." Along these lines, I think one's particular mix of beneficial and harmful tendencies as well as one's attempts to self-shape can differentiate a character from others. (See Herman's "Contingency in Obligation" in *Moral Literacy*, pp.307-308)

What's more central to our interest in building a character, however, is identifying these properties in ourselves. We certainly have greater access to our own reasons and the circumstances surrounding our own actions, but without deliberative effort we can easily fail to see how our reasons issue from the complicated clusters of properties that constitute our character. One of the unfortunate effects of a limited moral vocabulary is that it disposes us to see our own reasons for action as more atomic and isolated than they are.[289] Some of the value of adopting a larger moral vocabulary is found in reinterpreting our behavior as having complex clusters of motivations, when before we might have seen only discrete causes.[290] But in order to do this we must recognize that our actions are guided not only by reasons we endorse, but also by inherited dispositions and propensities of which we are often unaware. A desire to be good natured about a friendly soccer game can be thwarted by lingering insecurity about one's athleticism or age. Our self-making reasons are often, in this way, restrained by unreflective moorings. For our action to be expressive of the ends we endorse, we must try to identify our inherited proclivities and limitations and decide whether we will be guided by them. An intention to be generous by paying for a meal might only manifest in action if one recognizes her penchant for thriftiness and overcomes it in that instance. But to construct a character, one must not only overcome an instance of psychological resistance, but regularly examine one's

[289] In arguing for a two tier motivational structure grounded by a moral motive, Barbara Herman makes a similar claim about our motives not being atomic. She writes, "We need to resist our proclivity to think of motives in terms of (or built out of) single end-desire pairs- having a desire for drink, wanting to promote justice." See her "Responsibility and Moral Competence" (pp. 103).

[290] When explaining how Kantian moral motivation can accommodate a robust notion of character, Barbara Herman notes that instead of seeing our desires as being atomic, she sees that, "… what we take to be desires are already highly evolved intentional dispositions, drawing on various original sources for their force, and extensive learning for their content and array of possible objects." See her "Making Room for Character", in *Moral Literacy*, (pp. 24)

unchosen deliberative propensities and consciously choose whether to allow them to exert deliberative force. We might of think of this as periodically tilling one's psychological soil and this can be both difficult and unsettling.[291] Perhaps because of the unflinching introspection this practice requires, it is itself a character defining act and, conversely, those who cannot or will not engage in it are marked by that as well.

This allows us to see how the facets of a person's character generate self-making reasons that can guide her action. When a person actively chooses to be guided by a character property and this motivates a particular action over another, a deliberately adopted aspect of her character has generated a self-making reason for action. Because prior chapters focused on the motives that guide life-structuring decisions, such as which profession to enter and whom to marry, they tended to emphasize 'world-affecting' self-making reasons and the character properties that drive them. One's conscious commitment to improving conditions in the world around her likely generates self-making reasons that are 'world-affecting.' But not all self-making reasons will be as world-affecting, and indeed the vast majority of them will not. A commitment to become a patient, neighborly person will tend to generate self-making reasons that produce actions affecting the world in a far more local manner. It will be the reason one waits for others to pass through a door, that she takes a moment to

[291] Barbara Herman refers to a 'virtuous circle of character' in which an agent uses her existing moral knowledge and training to respond appropriately to new facts about her environment and herself. Herman's point is that by utilizing our moral training and knowledge we get better at the practice of morality and extend our moral knowledge. Much of her discussion is forward-looking, directed primarily at new developments in the world and in ourselves. The process I make reference to above might be seen as backward-looking analogue. In order to become better moral agents, or agents of character, we must examine our inherited motives and dispositions at something like a forensic level. Much of what drives us is so familiar and deeply-rooted it can easily escape notice. But through rigorous examination we can unearth implicit proclivities and dispositions and then decide whether we want to allow ourselves to continue being guided by them. See Herman's "Contingency and Obligation" (pp. 308).

chat on the street, or offers a warm greeting to a store employee. Our character, or lack thereof, will manifest itself in these sorts of seemingly minor behaviors far more frequently than in the few life-structuring decisions we make. But although they have not been the focus of discussion to this point, these small actions likely matter as much or more in a life than the more structurally impactful decisions we make. For while these are not the things that guide the trajectory of one's life and while they are not what distinguish good from evil, the pieces of our character expressed through our everyday dealings determine almost everything else about the objects we are.

Postscript

The choice between the traditional hedonic model and the one advanced here is, at some level, a choice between different deliberative pictures. The former suggests that our deliberation is more unified and, when done properly, likely to bring increased happiness. The latter portrays this process as less direct and doesn't promise a good life will be a happy one. If we were asked which we'd rather be true of our lives, we'd concede the traditional picture is more attractive. It would clearly be nicer if the things that were objectively good for us more often increased our happiness in a linear and sustained fashion. But as one imagines this picture more fully, one sees that, for all its comforts, it lacks a vital aspect of being human. It's missing the notion of a person as an evolving object, trying to refine itself, reaching for something greater than contentment. It doesn't express perhaps the most redeeming part of our humanity; that we're ultimately driven not by a desire to feel, but a desire to be.

I admit the case against this traditional hedonic view seems almost *too* strong for the intended effect. It appears to assail not just our ordinary conception of happiness, but the very notion of trusting our affective states in the first instance. I've argued that our hedonic cycle is a largely closed circuit and we're unable to predict our future feelings accurately or recall past feelings without distortion. I've also argued that we incorrectly retain every confidence that our affective states are straightforwardly caused by events around us and so we treat them as a reliable metric. These remarks are not principally unsettling because they ask us to relinquish our traditional notion of happiness, but because they seem to ask something impossible of us; that we

somehow ignore much of what we feel. Even if we wanted to, and surely we do not, this would be impossible for us.

But I think this might be inferring too much. First, while our hedonic states aren't always reliable metrics of how our lives are going, they can serve as provisional indications. There is an intelligible space between treating happiness as the sacrosanct end of all our action and ignoring happiness completely. There is good reason to prefer happiness to misery, even if we are ultimately striving after something else. Second, the hedonic distortions described above do not completely obfuscate our experience and they are not entirely unpredictable. We are not likely to recall periods of intense misery as blissful, or predict that horrible decisions will make us very happy. The distortions function in somewhat systematic ways, not to entirely invert our hedonic spectrum but to smooth out curves and return us to a baseline. A bit of emotional intelligence can help us recognize their cadence and even consciously discount for some of their effect. We cannot expect this to free us from our hedonic prism, but neither does it mean that we should turn away from it entirely (even if we could).

What is notable, though, is that the distortions attaching to our faculty of desire need *not* apply to our judgments. When we act to maximize pleasure, we're chasing a moving target, but when we act to make ourselves into a certain thing, our end is less evasive. Even when the particular aim of our self-creation changes, we can know, and later remember, that we were acting to make ourselves. Ironically, one of the advantages of the view advanced here is that it begins with a less idealized conception of persons-in-time and thus promises less. It accepts that our anticipation of the future

will be blurred by hope or dread, that our memory is an interpretation fueled by a need to reduce cognitive dissonance, and that even in the present, a variety of pre-conscious mechanisms filter our perception of the world. It's little surprise, then, that such a view doesn't predict that a life lived well will always, or even mostly, result in happiness.

It rather offers the suggestion that this is not the appropriate measure of a successful life. While it accedes that it would be better if we remembered our days with fondness than regret, this is principally because regret can be unproductive and painful, not because it necessarily indicates a life lived badly. On this view, it would be better if we generally respected the choices we've made. But, again, this is not solely because such affirmation indicates that these decisions were good, but also because, at the extreme, a deep dissatisfaction with prior decisions can erode one's confidence in her deliberative capacity and lead to paralysis. The reason to avoid regret, then, is not to do with what it indicates about that past, but with what it can entail for the future.

Nevertheless, if this picture asks us to relinquish future happiness as the measure that our lives have gone well, one can reasonably ask what would replace it. There is an answer, even if it is a bit somber. It's that we are tentatively charting a course, which we will correct and re-correct many times, the success of which is *not* to be measured by its effect on our affective states. Rather, the fact that one is authoring her own course is an end in itself, irrespective of the happiness or sadness it will cause. Much as we want to think that self-determination should feel liberating, it might well feel uncomfortable and even frightening. According to the view advanced here, relative comfort is a peripheral consideration, for the sought end is not a feeling to be had, but a

fact written into the deliberative account of a person's life. We can find some comfort in the fact that it is not subject to the distortion that our happiness is; we can *know* that our motive in a character-defining choice was rooted in a concern with our own being, and even our success or failure in this project is a peripheral concern. For here it is the motive that holds value, not an envisaged end-state. Moreover, even if we were to learn that our happiness *was* more connected to that which is good for us than this hedonic data suggests, such findings needn't imperil this account. In such a case, even if happiness were to follow more reliably from the pursuit of objectively valuable goods in life, it would still be important to distinguish whether a person was acting for the sake of increased happiness or acting out of a concern for her character.

If the history of a person's life-structuring decisions includes an enduring interest in the construction her own being, then that person sought to create something that existed in the world alongside mountains, roads, and poetry. Whether it was, in fact, a beautiful thing is, for epistemic reasons, something a person might not ever know. But she can know that allowing her action to be guided by a concern for herself as an object counts, not inconsiderably, in her favor.

While there is no ultimate accounting of all things, no arrival to prove the value of her journey, she can look up from her life-in-progress and know that she's trying to become something. Even if her affective experience is bound by internal tethers and her memory is distorted, this effort, by itself, is significant. For just as the observance of morality indicates a recognition that others are relevantly similar and deserving of respect, her interest in herself as an object locates her as a part of the world, and one to

which she has a special relation. She can help other people, or improve the world around her, but all the while she *is* something. Irrespective of whether she ultimately succeeded or whether it felt pleasant, her concern to become a good thing is itself a piece of the world. In such a case, a part of the world, however small, realized it was *a part of a world*, and wanted to be a good part. In that, there is a measure of value and one which is not adequately expressed in the traditional language of happiness or interpersonal morality.

Trent Teti

References Cited:

Pascal, Blaise, 1623-1662. (1958). *Pascal's Pensées*. New York :E.P. Dutton.

Adams, Robert Merrihew (1999). *Finite and Infinite Goods: A Framework for Ethics*. Oxford University Press.

Aristotle., Irwin, Terence, & Fine, Gail. (1996). *Aristotle: Introductory readings*. Hackett Pub.

Appley, Mortimer & Symposium on Adaptation-Level Theory. (1971). *Adaptation-level theory: a Symposium*. New York : Academic Press.

Arneson, Richard (2006). Desire Formation and Human Good. *Royal Institute of Philosophy Supplement*, 59, 9-32.

Audi, Robert (ed.) (1995). *The Cambridge Dictionary of Philosophy*. Cambridge University Press.

Bacon, Francis, 1561-1626. (1680). The Essays or Counsels, Civil and Moral, of Sir Francis Bacon Lord Verulum Viscount St. Alban : London.

Brännmark, Johan (2006). Leading a Life of One's Own: On Well-Being and Narrative Autonomy. *Royal Institute of Philosophy Supplement*, 59, 65-82.

Brandt, Richard (1979). *A Theory of the Good and the Right*. Prometheus Books.

Bernoulli, Daniel (1954). Exposition of a New Theory on the Measurement of Risk. *Econometrica, 22*(1), 23-36.

Brickman, Philip, & Campbell, Donald (1971). Hedonic relativism and planning the good society. In Appley, Mortimer (Ed.), Adaptation-level theory (pp. 287-305). New York: Academic Press.

Brickman, Philip, Coates, Dan, & Janoff-Bulman, Ronnie (1978). Lottery winners and accident victims: Is happiness relative? *Journal of Personality and Social Psychology*, 36(8), 917–927.

Brink, David (1992). Mill's deliberative utilitarianism. *Philosophy and Public Affairs* 21 (1):67-103.

Buss, Sarah (2004). The Irrationality of Unhappiness and the Paradox of Despair. *The Journal of Philosophy*, 101(4), 167-196.

Camus, Albert, (1991). *The Plague*. New York: Vintage Books.

Coope, Christopher (1998). Was Mill a Utilitarian? *Utilitas* 10 (1):33.

Cooper, Wesley; Nielsen, Kai & Patten, Steven (1979). *New Essays on John Stuart Mill and Utilitarianism*. Canadian Association for Publishing in Philosophy.

Costa, Paul, Zonderman, Alan, McCrae, Robert, Cornoni-Huntley, Joan, Locke, Ben, & Barbano, Helen (1987). Longitudinal analyses of psychological well-being in a national sample: Stability of mean levels. *Journal of Gerontology*, 42(1), 50–55.

Deaton, Agnus, & Stone, Arthur (2014). Evaluative and hedonic wellbeing among those with and without children at home. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(4), 1328–1333.

Diener, Ed, Suh, Eunkook, Lucas, Richard, & Smith, Heidi (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin*, 125(2), 276–302.

Driver, Julia (1992). The suberogatory. *Australasian Journal of Philosophy*, 70 (3):286 – 295.

Dryer, D.P. (1979) Justice, Liberty, and the Principle of Utility in Mill, *Canadian Journal of Philosophy*, 9:sup1, 63-73

Dworkin, Ronald (2000). *Sovereign virtue: the theory and practice of equality*. Cambridge, Mass: Harvard University Press.

Dworkin, Ronald. (2011). *Justice for Hedgehogs*. Cambridge, Massachusetts; London, England: Harvard University.

Easterlin, Richard (1974) Does Economic Growth Improve the Human Lot? Some Empirical Evidence In: David, R. and Reder, R., Eds., *Nations and Households in Economic Growth: Essays in Honor of Moses Abramovitz*, Academic Press, New York.

Easterlin Richard (2003). Explaining happiness. *Proceedings of the National Academy of Sciences of the United States of America*. 100(19),11176-11183.

Elster, Jon (2016). *Sour Grapes: Studies in the Subversion of Rationality* (Cambridge Philosophy Classics). Cambridge: Cambridge University Press.

Festinger, Leon (1957). *A Theory of Cognitive Dissonance*. Stanford University Press.

Festinger, Leon & Carlsmith, James (1959). Cognitive Consequences of Forced Compliance. *The Journal of Abnormal and Social Psychology*, 58(2), 203–210.

Foot, Phillippa (1972). Morality as a System of Hypothetical Imperatives. *The Philosophical Review, 81*(3), 305-316.

Frankfurt, Harry (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy*, 68(1), 5-20.

Frederick, Shane & Loewenstein, George (1999). Hedonic adaptation. In Daniel Kahneman, Ed Diener & Norbert Schwarz (Eds.), *Well-being: The Foundations of Hedonic Psychology* (pp. 302-329). New York: Russell Sage Foundation.

Gilbert, Daniel, Pinel, Elizabeth, Wilson, Timothy, Blumberg, Stephen, & Wheatley, Thalia (1998). Immune neglect: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology*, 75(3), 617–638.

Gilbert, Daniel, & Ebert, J. E. J. (2002). Decisions and revisions: The affective forecasting of changeable outcomes. *Journal of Personality and Social Psychology*, 82(4), 503–514.

Gilbert, Daniel, Lieberman, Matthew, Morewedge, Carey, & Wilson, Timothy (2004). The Peculiar Longevity of Things Not So Bad. *Psychological Science*, 15(1), 14–19.

Gilbert, Daniel. (2006). *Stumbling on happiness*. Alfred A. Knopf.

Hare, Richard (1981). Prudence, Morality and Supererogation. In *Moral Thinking: Its Levels, Method, and Point*. (188-205) : Oxford University Press.

Headey, Bruce & Wearing, Alexander. (1989). Personality, life events, and subjective well-being: Toward a dynamic equilibrium model. *Journal of Personality and Social Psychology*, 57(4), 731–739.

Herman, Barbara (1981). On the Value of Acting from the Motive of Duty. *The Philosophical Review, 90*(3), 359-382.

Herman, Barbara (2006). Reasoning to obligation. *Inquiry: An Interdisciplinary Journal of Philosophy* 49 (1):44 – 61.

Herman, Barbara (2007). Moral Literacy. Harvard University Press.

Herman, Barbara. (2007) Contingency in Obligation in *Moral Literacy*. Harvard University Press (pp. 300-331)

Herman, Barbara (2007) Can Virtue Be Taught? The Problem if New Moral Facts in *Moral Literacy.* Harvard University Press (pp. 106-129)

Herman, Barbara (2007) Making Room for Character in *Moral Literacy.* Harvard University Press (pp.1-28)

Herman, Barbara (2007) Responsibility and Moral Competence in *Moral Literacy.* Harvard University Press (pp.79-105)

Herman, Barbara. (2007) The Scope of Moral Requirement in *Moral Literacy*. Harvard University Press (pp. 203-229)

Hill, Thomas (1973). Servility and Self-Respect. *The Monist* 57 (1):87-104.

Hill, Thomas (1992). *Dignity and Practical Reason in Kant's Moral Theory*. Cornell University Press.

Kahneman, Daniel (1999). Objective happiness. In Daniel Kahneman, Ed Diener, & Norbert Schwarz (Eds.), *Well-being: The Foundations of Hedonic Psychology* (p. 3–25). Russell Sage Foundation.

Kahneman, Daniel., Diener, Ed, & Schwarz, Norbert. (Eds.). (1999). *Well-Being: Foundations of Hedonic Psychology*. Russell Sage Foundation.

Kahneman, Daniel & Krueger, Alan (2006). Developments in the Measurement of Subjective Well-Being. *The Journal of Economic Perspectives,* 20 (1), 3-24.

Kant, Immanuel, & Wood, Allen (1996). In Mary Gregor (Ed.), *Practical Philosophy (The Cambridge Edition of the Works of Immanuel Kant, pp. V-Vi)*. Cambridge: Cambridge University Press.

Liberto, Hallie (2012). Denying the Suberogatory. *Philosophia* 40 (2):395-402.

Lucas, Richard, Clark, Andrew, Georgellis, Yannis & Diener, Ed (2003). Reexamining adaptation and the set point model of happiness: Reactions to changes in marital status. *Journal of Personality and Social Psychology*, 84(3), 527–539.

Lucas Richard, Clark Andrew, Georgellis Yannis, & Diener Ed Unemployment alters the set point for life satisfaction. *Psychological Science*. 2004 (1) 8-13.

Lykken, David & Tellegen, Auke (1996). Happiness is a stochastic phenomenon. *Psychological Science*, 7(3) 186–189.

Jevons, William Stanley “A Brief Account of a General Mathematical Theory of Political Economy” *Journal of the Royal Statistical Society*, London, XXIX June 1866, pp. 282-87

Mill, John Stuart (2014). *Utilitarianism* (Cambridge Library Collection - Philosophy). Cambridge: Cambridge University Press.

Moore, George Edward (1903). *Principia Ethica*. Cambridge: Cambridge University Press.

Nozick, Robert (1974). *Anarchy, State, and Utopia*. Basic Books.

Olsaretti, Serena (2006). Preferences and well-being. edited by Serena Olsaretti ; contributors, Richard Arneson ... [et al.]. *Royal Institute of Philosophy Supplements,* 5. Cambridge; New York: Cambridge University Press.

Oshana, Marina (1998). Wanton Responsibility. *The Journal of Ethics, 2*(3), 261-276.

Plamenatz, John (1958). *The English Utilitarians*. Oxford: Blackwell.

Parfit, Derek (1984). *Reasons and Persons*. Oxford University Press.

Ondaatje, Michael (1992). *The English Patient: A Novel.* New York: Knopf.

Qizilbash, Mozaffar (2006). Well-Being, Adaptation and Human Limitations*. *Royal Institute of Philosophy Supplement*, 59, 83-110.

Qizilbash, Mozaffar (2006). Capability, Happiness and Adaptation in Sen and J. S. Mill. *Utilitas*, 18(1), 20-32.

Rachels, James & Rachels, Stuart (2007). *The Elements of Moral Philosophy*. Boston: McGraw-Hill.

Railton, Peter (1986). Facts and Values. *Philosophical Topics*, 14(2), 5-31.

Rawls, John (1971). *A Theory of Justice*, Belknap Press/Harvard University Press.

Rice, Christopher (2013) Defending the Objective List Theory of Well-Being. *Ratio*, 26: 196-211.

Rosati, Connie (1995). Persons, perspectives, and full information accounts of the good. *Ethics* 105 (2):296-325.

Rosati, Connie (2006). Preference-Formation and Personal Good. *Royal Institute of Philosophy Supplement* 59:33-64.

Sayre-McCord, Geoffrey (2001). Mill's “Proof” of the Principle of Utility: A More than Half-Hearted Defense. *Social Philosophy and Policy, 18*(2), 330-360.

Schneewind, Jerome (1977). *Sidgwick's Ethics and Victorian Moral Philosophy*. Oxford University Press.

Scheffler, Samuel. (1984). The Rejection of Consequentialism: A Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions. *Journal of Philosophy* 81(4): 220-226.

Shaver, Robert (1997). Sidgwick's False Friends. *Ethics* 107 (2):314-320.

Silverstein, Matthew (2000). In Defense of Happiness: A Response to the Experience Machine. *Social Theory and Practice, 26*(2), 279-300.

Sidgwick, Henry (1874). *The Methods of Ethics*. Thoemmes Press.

Sobel, David (1994). Full information accounts of well-being. *Ethics* 104 (4):784-810.

Stark, Cynthia (1997). The Rationality of Valuing Oneself: A Critique of Kant on Self-Respect. *Journal of the History of Philosophy 35* (1):65-82.

Sumner, Leonard (1979). The Good and the Right. *Canadian Journal of Philosophy* 9 (sup1): 99-114.

Sunstein, Cass & Thaler, Richard (2006). “Preferences, Paternalism, and Liberty”. *Royal Institute of Philosophy Supplement*, 59, 233-264.

Urmson, James (1958). Saints and heroes. In Abraham Melden (ed.), *Essays in Moral Philosophy*. University of Washington Press.

Urmson, James (1988). “Hare on Intuitive Moral Thinking”, in *Hare and Critics*, D. Seanor and N. Fotion (eds.), Oxford: Clarendon Press.

Watson, Gary (1975). Free Agency. *The Journal of Philosophy, 72*(8), 205-220

Weintraub, Elliot (2002). Neoclassical economics. The Concise Encyclopedia of Economics, https://www.econlib.org/library/Enc1/NeoclassicalEconomics.html

Williams, Bernard. (1981). *Moral Luck: Philosophical Papers 1973–1980*. Cambridge: Cambridge University Press.

Wilson, Timothy, Meyers, Jay, & Gilbert, Daniel (2003). "How Happy was I, Anyway?" A Retrospective Impact Bias. *Social Cognition*, 21(6), 421–446.

Wilson, Warner (1967). Correlates of Avowed Happiness. *Psychological Bulletin*, 67(4), 294–306.