UC Santa Barbara
UC Santa Barbara Electronic Theses and Dissertations

Title
Adam Smith on the Nature and Authority of Conscience

Permalink
https://escholarship.org/uc/item/3r02j3jk

Author
Shin, Albert

Publication Date
2014

Peer reviewed|Thesis/dissertation

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

UNIVERSITY OF CALIFORNIA

Santa Barbara

Adam Smith on the Nature and Authority of Conscience

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Philosophy

by

Albert Shin

Committee in charge:

Professor Aaron Zimmerman, Chair

Professor Thomas Holden

Professor Kevin Falvey

Professor Anthony Brueckner, *In Memoriam*

September 2014

The dissertation of Albert Shin is approved.

Thomas Holden

Kevin Falvey

Aaron Zimmerman, Committee Chair

Anthony Brueckner
In Memoriam

September 2014

Adam Smith on the Nature and Authority of Conscience

Copyright © 2014

by

Albert Shin

ACKNOWLEDGEMENTS

I would like to express my very great appreciation to my committee chair, Dr. Aaron Zimmerman, for his support and encouragement, as well as his extensive and valuable feedback. I am also grateful to Dr. Thomas Holden, Dr. Kevin Falvey, Dr. Anthony Brueckner, and the entire University of California, Santa Barbara faculty, who have shaped my philosophical thinking.

I wish to acknowledge the deep insight and friendship provided by my fellow graduate students, Rob McIntyre, Tim Butzer, Jenessa Strickland, and Austin Somers. To my undergraduate students – you have consistently challenged me to improve as a thinker, writer, and instructor. I would also like to offer my thanks to the UCSB staff and administrators who have supported me through their time, effort, and funding.

Special thanks to my father, Jonathan, who sparked my curiosity in philosophy and theology, to my mother, Esther, who always believed in me, and my brother, Joseph, without whom I would not have been able to pursue this degree.

Finally, to my loving wife, Sonya: my deepest heartfelt thanks. You supported me through the most difficult phase of my dissertation process, teaching me to have grace on myself and providing perspective when I was in doubt. I could not have completed this dissertation without you. I dedicate this work to you.

VITA OF ALBERT SHIN
September 2014

EDUCATION

Bachelor of Arts in Psychology and Philosophy, University of California, Los Angeles, June 2005 (summa cum laude)
Master of Theological Studies in Ethics, Philosophy and Theology, Boston University, May 2008 (magna cum laude)
Doctor of Philosophy in Philosophy, University of California, Santa Barbara, September 2014 (expected)

PROFESSIONAL EMPLOYMENT

2009-2012 Teaching Assistant and Associate, Department of Philosophy, University of California, Santa Barbara
2012-2013 Research Associate, Department of Philosophy, University of California, Santa Barbara
2013-2014 Teaching Assistant and Associate, Department of Philosophy, University of California, Santa Barbara
2014-2015 Visiting Assistant Professor, Department of Ethics, Villanova University

AWARDS

Certificate in College and University Teaching, University of California, Santa Barbara, 2014
Dean's Prize Teaching Fellowship, University of California, Santa Barbara, 2014
Graduate Humanities Research Fellowship, University of California, Santa Barbara, 2012-2013
Summer Teaching Institute for Associates Certificate, University of California, Santa Barbara, 2012
RHA/ORL Outstanding TA Recognition, University of California, Santa Barbara, 2012
Graduate Division Fee Fellowship, University of California, Santa Barbara, 2009-2012
Ralph W. Church Fellowship, University of California, Santa Barbara, 2008-2009
Academic Scholarship, Boston University, 2006-2008
Phi Beta Kappa Honor Society, University of California, Los Angeles, 2005

FIELDS OF STUDY

Major Field: Ethics, 17th and 18th Century Philosophy, Moral Psychology

ABSTRACT

Adam Smith on the Nature and Authority of Conscience

by

Albert Shin

Conscience plays a central role in our moral lives. We have all felt the pangs of conscience when we fail to obey it; we are terrified by those who lack it; and we look to it as a moral guide, often using it to justify our actions. Yet there is little understanding of what it is, how it operates, and what justifications we have for obeying it.

The aim is to provide one plausible account of the nature and authority of conscience, that of Adam Smith. According to Smith, conscience, or what he calls 'the supposed impartial spectator', is the moral agent judging herself from the situation of an impartial spectator, just as she would judge others. Under this account, conscience is not an ideal or prototypical judge, but rather the agent judging as she would judge others. As a result, conscience is liable to all the same errors that any spectator is, especially partiality. Thus, there is a need to cultivate our conscience. We do so, I will argue, primarily through encountering diversity, which leads to disagreements, which prompt us to reevaluate how we judge others. Furthermore, Smith also claims that conscience has authority in virtue of the respect we give it. Our respect for conscience is rooted in our recognition that conscience is a better moral judge, demonstrated by our appeal to conscience to correct for our immediate, unreflective moral judgments. However, this account fails to capture a key feature of our phenomenological experience of conscience: conscience presents itself as *the* moral

authority, not merely a helpful moral guide. I argue that Smith’s theory, though not Smith himself, provides an alternative account of the authority of conscience: because conscience is our judging ourselves using our own faculties, we are committed to the accuracy of conscience’s judgments the same way we are committed to the accuracy of all of our judgments.

TABLE OF CONTENTS

I. Historical Development of 'Conscience' ………………………………………………….. 1

A. The Medieval Period: Bonaventure and Aquinas ……..……………………….. 4

B. Thomas Hobbes and the Security of the State …………………………………. 12

C. Bishop Butler and the Voice of God …………………………………………… 19

D. Immanuel Kant and Self-Awareness …………………………………………... 26

E. John Stuart Mill and the Sense of Duty ………………………………………... 32

F. Sigmund Freud and Psychoanalytics …………………………………………… 38

G. C. D. Broad and Contemporary Psychology ……………………………………44

H. Core Features of Conscience ……………………………………………………47

II. From Judging Others to Judging Ourselves: Conscience as Reflective Agent ………… 51

A. The Sympathetic Mechanism ……………………………………..……..……...53

B. How We Judge Others: Propriety/Impropriety and Merit/Demerit ……………. 58

C. How We Judge Ourselves: The 'Supposed Impartial Specator' ………………..66

D. Alternative Interpretation: Conscience as Prototype? …………………………. 69

E. Interpretive Issue #1: 'Supposed' or 'Imagined' Spectators …………………… 76

F. Interpretive Issue #2: 'Impartial' Spectators ……………………………………. 81

G. Interpretive Issue #3: 'Ideal' …………………………………………………… 98

H. Assessing Smith on Conscience ………………………………………………...108

III. Cultivating Conscience: The Amiable Virtues ………………………………………... 112

A. How Does Conscience Err? ……………………………………….………… 114

B. Correcting an Erroneous Conscience: The Explicit Account ….………………..124

C. Limits of General Rules ……………………………………………………... 127

D. Correcting Conscience by Correcting Our Moral Judgments of Others ……….. 134

E. How to Cultivate the Amiable Virtues ……………………………………………. 144

F. Concluding Remarks …………………………………………………………… 159

IV. The Authority of Conscience …………………………………….……………………161

A. Butler and Smith on the Authority of Conscience ……………….……………. 163

B. Distancing Smith from Butler: Smith on Conscience and 'Moral Faculties' ….. 167

C. Distancing Smith from Butler: Smith on 'Authority' ………………………….. 170

D. Smith on the Authority of Conscience …………………………………………. 181

E. Why Should I Obey My Conscience? …………………………………………….. 186

F. Problem for Smith: Conscience as *The* Moral Authority ………………………. 189

G. An Alternative: The Authority of Conscience as Commitment to Self ………... 193

H. Huckleberry Finn and Our Commitment to Self ………………………………. 202

References ……………………………………………………………………………….204

1

Historical Development of ‘Conscience’

In no sense do I advocate evading or defying the law, as would the rabid segregationist. That would lead to anarchy. One who breaks an unjust law must do so openly, lovingly, and with a willingness to accept the penalty. I submit that an individual who breaks a law that conscience tells him is unjust, and who willingly accepts the penalty of imprisonment in order to arouse the conscience of the community over its injustice, is in reality expressing the highest respect for law.

Martin Luther King, Jr. from “Letter from Birmingham Jail”[1]

Conscience plays a key role in our moral lives. We view it as a feature of a morally developed human being.[2] We have all felt the pangs of conscience when we fail to act in accordance with its dictates. We are terrified (and maybe even fascinated) by those who lack it (e.g. psychopaths).[3] And like Martin Luther King, Jr., we look to conscience for moral guidance, and appeal to it in justifying our actions. Yet there is great confusion as to what it is. We are given images of Jiminy Cricket or little angels on our shoulders, whispering moral commands into our ears. Often, these figures are thought to be representatives of God, society, or morality. Others are more skeptical, reducing conscience to mere opinion or feeling. But all these are merely hinting at a particular phenomenological experience we

[1] Martin Luther King, Jr. “Letter from Birmingham Jail” in *Why We Can’t Wait* (1963, 68).
[2] For instance, Kohlberg claims conscience is a feature of the later stages of moral development (Kohlberg 1981, 1984).
[3] Psychopaths are often characterized as lacking a conscience. For example, Robert D. Hare, a central figure in psychological research on psychopaths, cites a lack of conscience as one of the identifying characteristics of psychopaths (Hare 1999, 2004).

have: the seemingly immediate access we have to what is right or wrong to do in a given situation, presented to us as authoritative. Given the role that that conscience plays in our everyday moral lives, it is important that we gain a better grasp of it nature, and consequently its authority over us.

Our views on conscience have a long history. The term 'conscience' comes from the Latin term 'conscientia', which in turn comes from the Greek term 'syneidesis'. These terms carried a wide range of meaning: "holding of knowledge in common, fact of being privy to a crime, complicity, private knowledge, consciousness, moral sense, consciousness of right and wrong" (OED "Conscience"). C. S. Lewis divides into two branches: conscience as a state of awareness or knowledge (what Lewis calls the 'weakened' branch), and conscience as a state of shared knowledge (or the 'together' branch).[4] In the medieval period, 'conscience' retained its status as knowledge, either of general rules or of the application of those rules. The knowledge of general moral rules was thought to be given by the great Author of Nature, but its application was vulnerable to mistakes. From the 17th to the 19th Century, we find a much more diverse landscape. Some, like Bishop Butler, continued to appeal to conscience as an authoritative voice that provided moral knowledge, while others, like Thomas Hobbes, questioned its authority, reducing it to the status of mere opinion.[5] Alternatively, John Stuart Mill viewed conscience as a particular *feeling*, the sense of duty, which holds little connection to reasoning and knowledge. Today, our conceptions of conscience are influenced by the work of Sigmund Freud, who, similar to Butler, thought conscience was the function

[4] C. S. Lewis, *Studies in Words* (1960, 182-187). According to Lewis, conscience as shared knowledge often carried a negative connotation: the shared knowledge, more aptly described as shared secret, was between conspirators of some crime. But this was not always so, for instance, in the case of Hobbes, which I will demonstrate later in this chapter.

[5] I discuss their views in detail below. A helpful starting point for discussions on conscience in this period is *Contexts of Conscience in Early Modern Europe, 1500-1700* (Braun, Harald E. & Vallance, Edward, eds., 2004). Synderesis, which was closely tied to conscience and widely discussed in the medieval period, also remained in discussions in the Early Modern period, though not to the same extent (see Greene 1991).

of a distinct faculty, the super-ego. But generally, we often speak of conscience in all these various ways (as a source of knowledge, as a feeling, and as a distinct faculty), which may explain the appeal of C. D. Broad's definition of conscience as a multifaceted phenomenon that involves cognitive, affective, and conative components.

The general aim of this dissertation is to give what I take to be the most plausible and fruitful account of conscience, that of Adam Smith, and to show its significance in moral discourse. But to fully appreciate Smith's account, it will be helpful to compare his account to alternatives. And the historical backdrop, along with some contemporary views, will be especially helpful because in many ways Smith is addressing the same issues with his own account of conscience. Ultimately, our interest is the authority of conscience and the degree to which we ought to follow its guidance. But to address its authority, we need to first explain what it is, how it functions, and how reliable it is in providing moral knowledge. So I will present these differing views by how they addressed the following questions:

[A] What is conscience?
[B] Does it provide moral knowledge, and if so, how reliable is it as a source of moral knowledge?
[C] Does conscience have authority over us, i.e. should we follow the dictates of conscience?

Though the presentation here is a more general overview, I will focus on discussions of conscience in the 17th and 18th century for two reasons: 1) the primary aim of this overview is to set the stage for a discussion of Adam Smith (18th-century); and 2) with the exception of Butler's account, discussions of conscience in this period have garnered little attention, even though it was during this period that ideas of conscience began to diverge in important ways.

I will try to highlight some of these developments, and in the next chapter, place Smith within that context.[6]

1.1 The Medieval Period: Bonaventure and Aquinas[7]

Conscience was not discussed widely until the medieval period, during which conscience was used to explain *akrasia*: how is it that we fail to act in accordance with what we know to be good? The thought was that we had knowledge of general principles or rules about what is good to do, but in any particular situation, we failed to apply those rules correctly. Conscience, it was generally thought, was involved in this process of practical reasoning, from the general rules to the particular application of that rule. The debate was on exactly *how* conscience was involved in this process, and in particular the relationship between conscience and synderesis.

In *Commentary on the Sentences*, Book II, distinction 39, Bonaventure describes conscience (*conscientia*) as a disposition of practical, not theoretical, reason. Conscience first provides us with innate knowledge of very general moral principles (e.g. obey God's commands, do not steal). Conscience then takes that knowledge, and coupled with some

---

[6] In summarizing these accounts in chronological order, I do not intend to present any positive thesis about how our understanding of conscience changed over time. I simply pick out these particular accounts because they represent the wide range of views that have been considered. And in my discussion of Adam Smith, these accounts will primarily be used as points of contrast. The exceptions will be the views of those with whom Smith was, or plausibly was, familiar, like Butler. But I will be explicit if I think Smith is drawing directly from any other writers.

For an interesting interpretation of the historical development of 'conscience', see Langston 2001. According to Douglas Langston, the lack of contemporary discourse on conscience results from the dominance in the 19th and 20th century of the view that conscience is a distinct faculty, and the subsequent dismissal of theories holding such a view, such as Freud's (Langston 2001). I hold some reservations about Langston's explanation because, as I will discuss, I do not think that Freud believed conscience is a faculty, and that this misunderstanding of Freud reflects a failure to appreciate Freud's subtle distinction between conscience and the superego.

[7] Much has already been written on accounts of conscience in the medieval period, so I shall keep the review here brief. For a more thorough overview of views of conscience in the Medieval period, see Timothy Potts' *Conscience in Medieval Philosophy* (1980), Langston's *Conscience and Other Virtues* (2001), and Karen Feldman's *Binding Words* (2006).

beliefs (e.g. God commands you to pray, taking a penny from this jar is stealing), applies that general moral principle to come to a less general claim (e.g. I should pray, I should not take a penny from this jar).[8] In this way, conscience provides knowledge, specifically practical knowledge, i.e. knowledge of what to desire or what to do:

> *[C]onscientia* is a disposition perfecting our thought to the extent that it is practical, or to the extent that it directs us towards deeds…. Such a disposition is, accordingly, not just called 'knowledge' (*scientia*), but '*conscientia*', so as to signify that this disposition does not in itself perfect the theoretical potentiality, but does so as joined in some way to desire and deed (Bonaventure 111).[9]

Bonventure is using the 'con' in '*conscientia*' to distinguish conscience from theoretical knowledge; unlike theoretical knowledge, conscience provides knowledge that directs us toward (i.e. joins us to) some action. But conscience provides this knowledge by first discovering very general moral principles, then applying them to form less general ones. But this practical knowledge is not itself motivating. Conscience actualized potential knowledge through practical reasoning. So it provided knowledge of what one should do in a particular situation. And though this knowledge directs us toward the right action, it alone does not *move* one to act accordingly; it does not make one *want* to do what conscience directs us to do. To explain our motivation to do good, Bonaventure appeals to synderesis, a disposition that drives us towards the good:

---

[8] Langston draws out this distinction (between conscience as that which gives us access to the innate, general moral principles and that which applies those general principles to situations) more clearly in *Conscience and Other Virtues* (2001). He labels the former "potential conscience" and the latter "applied conscience".

According to Potts, the relevant distinction for Bonaventure is not between general and particular deontic propositions, but rather between necessary and contingent deontic propositions. He also draws some connections to the a priori/a posteriori distinction (Potts 1980, 61-64). For simplicity's sake, I will set aside both the necessary/contingent and the a priori/a posteriori distinction. But I will note that the innate general moral principles are the first principles, i.e. principles from which all other deontic propositions are derived.

[9] All quotes by Bonaventure and Aquinas are Potts' translations in *Conscience in Medieval Philosophy* (Potts 1980). I will cite by indicating the original author (Bonaventure, Aquinas, or Potts) and the page number in Potts 1980.

> *[S]ynderesis* is a disposition with respect to good and bad in general, *conscientia* a disposition with respect to good or bad in particular, and natural law is related indifferently to either. But because, as was argued above, *conscientia* is an intellectual disposition, either it is necessary to suppose that there is something over and above *conscientia* and *synderesis* which directs us, or it is necessary to suppose that *synderesis* is in the desiring part [of the soul] (Bonaventure 116).

So conscience (in the reasoning faculty) tells us to what is good to do, and synderesis (in the desiring faculty) motivates us to do good generally. Synderesis never motivates us to any particular action (e.g. stealing this candy bar in front of me right now), but just to do good, whatever it may be.

Though conscience provides moral knowledge, it can be mistaken from time to time. Errors, however, do not result from the general moral principles, which are infallible. Errors occur in the applying of the general principles to particular cases:

> And thus, although *conscientia* is always right so long as it sticks to the general and is moved by simple inspection, it can become mistaken when it descends to particulars and brings things together, because the actualization of deliberative reason is mixed with it. This will be clear from the following. The *conscientia* of the Jews first told them itself by natural pronouncement that God is to be obeyed, and they assumed henceforth that God *now* directs circumcision and keeping [certain] foods separate. From this their *conscientia* is formed in the particular [matter], that they should circumcise themselves and abstain from [certain] foods. This mistake does not come from the first premiss, which was indeed true, but comes from adding the minor premiss, which was not from *consientia* as a natural tribunal, but rather from mistaken reason, which has regard to free choice (Bonaventure 120).

We are infallible in our knowledge of the general rules, like “God is to be obeyed”. But we believe, mistakenly, that God commands circumcision and keeping certain foods separate. So

when this belief is combined with the general rule, we come to a false conclusion that we ought to circumcise and keep certain foods separate.[10]

Since conscience gives us knowledge of the laws of God, we ought to follow its dictates. The authority of conscience, then, is derived from conscience's role as the representative of God. We are bound to the dictates of conscience the same way we are bound to the laws of God. But given that conscience can err and command us do that which is against the laws of God, should we *always* follow its commands? Bonaventure thinks we should, even if conscience commands us to act against the laws of God. His reasoning is that though in disobeying a mistaken conscience we are *acting* in accordance with God's commands, we are acting with the *wrong intention*. And acting with the wrong intention is sin (Bonaventure 1980, 115). Of course, Bonaventure is not claiming we are justified in acting against the laws of God. Rather, we are bound to change our conscience whenever it is mistaken in its dictates. But given we are unaware of the mistake we must still act in accordance with its dictates.

Thomas Aquinas gives a similar account of how we come to form practical moral knowledge: we have innate knowledge of general moral rules and we apply them to specific situations. However, Aquinas provides a different division between conscientia and synderesis.[11] Whereas Bonaventure associated conscientia with both general moral principles and their applications, and synderesis with the desire to do what is good, Aquinas associated

[10] Similarly, synderesis is never mistaken since it does not issue any commands of what we should do in a particular situation. Rather, it is the spark that motivates us; it "murmurs back against evil and prods us to good" (Bonaventure 1980, 120).
[11] Aquinas adopts his distinction between conscientia and synderesis from Philip the Chancellor. For those interested in the details of Philip's account, see Potts 1980 (12-31; 94-109).

conscientia strictly with the applications of the general principles, and synderesis with the general innate principles:

> Accordingly, just as there is a natural disposition of the human mind by which it apprehends the principles of theoretical disciplines, which we call the understanding of principles, so too it has a natural disposition concerned with the basic principles of behaviour, which are the general principles of natural law. This disposition relates to *synderesis*; it exists in no other potentiality but reason (Aquinas 124-5).
>
> For the noun '*conscientia*' signifies the application of knowledge to something, so that to be conscious of something (*conscire*) is, as it were, to know simultaneously (*simul scire*). But any knowledge can be applied to something, so '*conscientia*' cannot name some special disposition, or some potentiality, but names the actualisation which is the application of some disposition or other, or of something or other known, to a particular actualization (Aquinas 130-1).

Conscientia simply denotes the application of the general moral principles, grasped by synderesis, to narrower cases. What is interesting about Aquinas' account is that as we apply general principles through reasoning, we become *aware* or *conscious* of what is right to do. For Aquinas, consciousness from applying knowledge is itself a kind of knowledge.
For Aquinas, all knowledge requires first principles: the principles from which all other knowledge is derived:

> Nature, in all its works, aims at what is good and at the maintenance of whatever comes about through the working of nature. Hence, in all the works of nature, its first principles are always permanent and unchangeable and conserve right order, because first principles must endure, as Aristotle says (*Physics* 1.6, 189a19). For there could be no stability or certainty in what results from the first principles, unless the first principles were solidly established. Anything which is variable goes back, accordingly, to some first fixed thing (Aquinas 127).

We must have some starting point that grounds that knowledge because without any such starting point, all supposed knowledge is merely unfounded conjectures. Knowledge can only proceed from knowledge, and so for anyone to have knowledge, there must be some starting point, some piece of knowledge that is not derived from any other piece of knowledge. That starting point is the first principles. Moral knowledge is no different. If we are to be capable of doing any good, we need to have first principles that direct our actions toward good rather than evil. And these first principles are grasped by synderesis:

> Hence, in order that there can be some rightness in human deeds, there must be some enduring principle which has unchangeable rightness and by reference to which all deeds are tested, such that this enduring principle resists everything evil and gives assent to everything good. This is what *synderesis* is, whose job is to murmur back in reply to evil and to turn us towards what is good. Hence, it is to be admitted that it cannot do wrong (Aquinas 128).

We are born with certain general moral principles, first principles, which need not be derived from any other piece of knowledge. For Bonaventure, these first principles were contained in conscience; for Aquinas, they are contained in synderesis.

Of course, it is unclear exactly what these first moral principles are. As Potts rightly notes, though first principles are general deontic propositions, not all general deontic propositions are first principles. For instance, “Do not murder for profit” is a general deontic proposition, but Aquinas may not consider this proposition one of the first principles. Though it is not quite clear what Aquinas considers a first principle, Aquinas seems to treat the first principles as something more general from which “Do not murder for profit” can be derived. So Aquinas seems to think of first principles as something closer to Bonaventure’s first

principles, e.g. “Only do what is good and never what is bad”, “Obey God”, “Do not murder”, and “Do not steal”.[12]

As for what motivates us to do what is good, Aquinas, like Bonaventure, believes it is synderesis. However, Aquinas thinks that synderesis is part of the reasoning faculty that gives us access to the first principles. And among the first principles is “Obey God” and “Only do what is good and never what is bad”. These principles alone, part of the reasoning and not the desiring faculty, motivate us to do what is good. Hence, it is *reason*, not desire, that motivates one to do good, but it does so by means of general commands of what is right and wrong.

Like Bonaventure, Aquinas thinks we are infallible in our knowledge of the first moral principles:

> [N]o *conscientia* can mistakenly judge: ‘God is not to be loved by me’ or ‘Something bad ought to be done’. This is because in both kinds of judgment, the theoretical as much as the practical, the major premiss is known *per se*, as existing in the general judgment” (Aquinas 133).

As stated earlier, Aquinas thinks that all knowledge requires first principles. So if we have any moral knowledge at all, there must be first principles about which we are not mistaken. Supposing we have this knowledge of first principles, if we do make mistakes, it must then be in the applying of these first principles:

> Mistakes can occur in two ways in this application: first, because what is applied contains a mistake; second, because it is not applied properly. In the same way, mistakes in reasoning can occur in two ways: either because some false [premiss] is used, or because one does not reason correctly (Aquinas 132).

[12] I leave it open exactly what level of generality would count as first principles for Aquinas. However, it is clear that he has in mind propositions that do not specify any particular situation or particular end.

So if I come to a false judgment, like “I should exclude information on a loan application to increase my chances of getting the loan”, either I have a false belief, e.g. that excluding information is not lying, or I made an error in reasoning. What I was not, and cannot be, mistaken about was the first principle “do not lie”. And so with diligent care, i.e. coming to true beliefs and using proper reasoning, we can correct for errant moral judgments.

Like Bonaventure, Aquinas thinks that conscience is binding in virtue of God’s commandments. And like Bonaventure, Aquinas thinks that, in general, even mistaken conscience is binding because conscience is what we take to be God’s commands, and to act against one’s conscience is to act with the intent of disobeying God. Moreover, one cannot be bound to some command if one does not know that command. And so generally we are excused when we obey our conscience. The exception, however, is when we have a second-order duty to know God’s command, in which case, we are not excused because we failed to follow a command that we *should have* known even though we did not.[13] As Potts points out, since we cannot be mistaken about the first principles, any failure to fulfill this second-order duty must be a failure in deriving the correct general deontic propositions from the first principles (Potts 1980, 59). But Aquinas does not specify which commands are those that we have a second-order duty to know, so it is unclear how lenient Aquinas is in excusing mistaken conscience.

In summary, the general picture for both Bonaventure and Aquinas is that conscience is an aspect of practical reasoning that provides us with moral knowledge, specifically

[13] “Hence no one is bound by an injunction except by means of knowledge of that injunction and, therefore, anyone who is not capable of being informed, is not bound by the command; nor is someone who is ignorant of an injunction of God bound to carry out the injunction, except in so far as he is obliged to know the injunction” (Aquinas 135).

knowledge of what we ought to do in particular cases, derived from our knowledge of what is good and bad. Conscience can be mistaken because though we cannot be mistaken about the first moral principles, we can misapply those principles to come to the wrong conclusions. However, if we apply the general principles to particular cases correctly, then we have moral knowledge. The authority of conscience is derivative; we are bound to conscience's dictates because conscience gives us access to God's commands and we are bound to obey God's commands. Since conscience is representing God's commands, to act against one's conscience is wrong to do even if we are mistaken because to do so is to act with the intention of violating God's commands.[14] As we shall see, each of these points has come under attack through the years. Some question whether conscience is a reliable source of moral knowledge; others whether we must follow the dictates of our conscience. Moreover, whether our motivation for acting in accordance with conscience is located in the affective faculty (as Bonaventure claims) or in the rational one (as Aquinas claims) remains a point of contention today.

1.2 Thomas Hobbes and the Security of the State

Though conscience was not prominently discussed in the 17th and 18th century, it remained of concern for various reasons. Thomas Hobbes was especially concerned with a corrupted use of 'conscience' used as grounds for disobeying the sovereign, resulting in the

[14] In "Four Conceptions of Conscience" (1998), Thomas Hill characterizes one conception of conscience, which he labels the "popular religious conception", in terms of its connection to God as the source of an "instinctual access to moral truth" (Hill 1998, 17; for more on religious conception of conscience, see Kelly 1967 and Lyons 2009). Though many of the Medieval conceptions, including those of Bonaventure and Aquinas, can be characterized in this manner, doing so simply overlooks the most crucial aspect of these conceptions: conscience provides moral knowledge through reasoning from first principles. We can accept that aspect of these views without adopting a divine command theory or God as the source of our grasping of first principles. We can instead take "God's command" to mean commands of moral law. The same applies to Butler's account (§ 1.3). I will return to this point in my discussion of Adam Smith, whose account of conscience need not rely on any connection between God and morality, though he does speak of conscience as a representative of God.

dissolution of the commonwealth. I want to discuss Hobbes' account in more detail because his discussion highlights the various uses of 'conscience' in this period.[15]

In *Leviathan*, Hobbes distinguishes three uses of 'conscience', one proper (corresponding to shared knowledge), one metaphorical (corresponding to private knowledge), and one corrupted (corresponding to mere opinion). 'Conscience', properly defined, signifies shared knowledge of some fact:

> When two or more men know of one and the same fact, they are said to be CONSCIOUS of it one to another, which is as much as to know it together. And because such are fittest witnesses of the facts of one another, or of a third, it was and ever will be reputed a very evil act for any man to speak against his *conscience*, or to corrupt or force another so to do, insomuch that the plea of conscience has been always hearkened unto very diligently in all times (*L* vii.4).[16]

Hobbes is returning to one meaning of conscience: conscience as awareness of some moral fact. But he is understanding conscience as a *shared* awareness, not some private judgment. Later, Hobbes sometimes refers to this (proper) use of conscience as 'public conscience' to distinguish it from other uses of conscience. But unlike those before him, Hobbes thinks conscience is an awareness that is shared because it is nothing more than the civil law in a commonwealth. His thought is that when we enter into a commonwealth, we agree to bind ourselves to the commands of the sovereign. We essentially give the power to judge to a sovereign, who by civil law dictates what is good and bad. I will say more about this point in discussing the authority of conscience. But for now, I want to highlight the significance of

[15] For my purposes here, I will not discuss Calvin's influence on Hobbes regarding conscience, which admittedly is significant. Hobbes often cites Calvin in discussions on conscience. And much of the influence of Medieval thought on Hobbes' account of conscience may have come by way of Calvin. For more on Calvin's influence, see Domanique Weber's "Thomas Hobbes's Doctrine of Conscience and Theories of Synderesis in Renaissance England" (2010, 56-66).

[16] *Leviathan* (Indianapolis: Hackett, 1651/1994). Passages from *Leviathan* will be referred to by chapter and paragraph number, prefaced by *L*.

this development. Though prior to Hobbes, ‘conscience’ was used to refer to a shared awareness, it was primarily shared in the sense that everyone has access to the same piece of knowledge through their own reasoning. For Hobbes, private judgment is excluded altogether from conscience; instead, everyone has access to the same piece of knowledge because it is the decree of the sovereign.

Though one might worry that as civil law, public conscience may not always be true, Hobbes thinks that it has the status of knowledge. He treats ‘conscience’ as derivative of ‘science’, which is a kind of knowledge. And given Hobbes’ understanding of science, his account of conscience is quite similar to that of Aquinas:

> And therefore, when the discourse is put into speech, and begins with the definitions of words, and proceeds by connexion of the same into general affirmations, and of these again into syllogisms, and end or last sum is called the conclusion, and the thought of the mind by it signified is that conditional knowledge, or knowledge of the consequences of words, which is commonly called SCIENCE (*L* vii.4).

Aquinas thought we had infallible knowledge of first moral principles. These building blocks were necessary for knowledge to even be possible. From these first principles, through correct reasoning, we are able to come to more particular moral knowledge. Similarly, Hobbes thinks knowledge must begin with the basic building blocks: not first principles (as Aquinas maintained), but rather definitions of words. And through correct reasoning, we derive knowledge of other truths. Conscience in its proper use, however, is not merely knowledge, but knowledge that is shared, in that multiple individuals are aware of that piece of knowledge.

Hobbes also points out that some used ‘conscience’ metaphorically to refer to private knowledge:

> Afterwards, men made use of the same word metaphorically, for the knowledge of their own secret facts and secret thoughts; and therefore it is rhetorically said that the conscience is a thousand witnesses (*L* vii.4).

This metaphorical use treats conscience as self-awareness, in which we are witnesses to ourselves, the same way that others can be witnesses to us. Hence, in some sense we share knowledge, not with different individuals, but rather with ourselves. As we shall see, Hobbes is not primarily concerned with this use of 'conscience'. So for now, I will leave discussions of is aside. But we will return to this use in discussions of Butler, Kant, Freud and Smith, all of whom, as we shall see, use 'conscience' in a similar fashion.[17]

Of greater concern for Hobbes is the corrupted use of 'conscience' (or what he calls 'private conscience' elsewhere), in which one simply uses the name 'conscience' to elevate the status of one's private opinion to that of knowledge:

> And last of all, men vehemently in love with their own new opinions (though never so absurd), and obstinately bent to maintain them, gave those their opinions also that reverenced name of conscience, as if they would have it seem unlawful to change or speak against them; and so pretend to know they are true, when they know, at most, but that they think so (*L* vii.4).

The first two uses of 'conscience' (as shared knowledge and as a metaphor) both use conscience to refer to *knowledge.* Even the metaphorical use referred to knowledge, just shared with oneself, not with others. But this third use is corrupted because it takes what is

[17] In *Binding Words*, Karen Feldman claims "conscience may be seen as *the* most dangerous metaphor for both Hobbes' nominalism and his political philosophy as a whole, for it is precisely the metaphoric shift in our understanding of conscience that, in Hobbes's account, corrupts knowledge into opinion, making error and deception possible" (Feldman 2006, 20). Though I do not disagree with the overall claim about the threat of metaphors, I think the use of Hobbes' discussion of 'conscience' as a prime example might be an overstatement. Hobbes distinguishes between the metaphorical use of 'conscience' and the *corrupted* use, and he only attacks the corrupted use, never the metaphorical one, throughout *Leviathan*. In fact, apart from the initial distinction, he never again mentions the metaphorical use in *Leviathan*.

not knowledge but opinion and gives it the misleading status of knowledge.[18] Hobbes may be responding here to multiple figures who treated conscience as private to an individual, including Calvin and some of the Medieval tradition he draws on (see Weber 2010) and philosophers like Locke (see Andrews 1999). But it may also be that he finds this corrupted use to simply be prevalent in his time.

Not surprisingly, the reliability and authority of conscience depends on the particular use of 'conscience'. Public conscience is a kind of *moral knowledge*, since it is derived from definition. But then we might be a bit suspicious here because public conscience is derived from definitions that the sovereign sets. The sovereign decides what is good and bad, instead of leaving the decision up to individuals. And by defining what is good and bad, public conscience cannot be mistaken, at least in the definitions. And if we derive any propositions from these definitions through sound reasoning, then we are not mistaken about those propositions and have knowledge of them. Private conscience, on the other hand, is *mere opinion*. And though our private opinion may be correct, our private opinions are often mistaken. Moreover, private opinion is not derived from definitions the way that knowledge is. Consider Hobbes' definition of opinion, contrasted with his definition of science (given earlier):

> But if the first ground of such discourse be not definitions, or if the definitions be not rightly joined together into syllogisms, then the end or conclusion is again OPINION, namely of the truth of somewhat said, though sometimes in absurd and senseless words (*L* vii.4).

[18] In my presentation here, I am primarily drawing on Hobbes' discussion of conscience in *Leviathan*. However, in *Elements of Law* (*EL*), Hobbes simply presents the corrupted use of 'conscience' in which conscience is a mere opinion that one takes to be knowledge grounded on some evidence (*EL* vi.8). In *Leviathan*, Hobbes is much more careful in distinguishing the three different uses. I am taking *Leviathan* to contain Hobbes' more considered account of 'conscience'.

Though not always false, opinions are by their very nature different from knowledge because they are either not the products of definitions, or not the products of sound reasoning. So even if an opinion happens to be true, it cannot be granted the status of knowledge.

Public conscience has authority over us, demanding that we act in accordance with it. But its authority stems from our entering into a commonwealth. Public conscience is civil law, the laws passed by the sovereign, but civil law only exists in a commonwealth. Hence, Hobbes distinguishes between acting against our conscience in the state of nature and doing so in a commonwealth:

> In the second place, I observe the *diseases* of a commonwealth that proceed from the poison of seditious doctrines, whereof one is: *That every private man is judge of good and evil actions*…. Another doctrine repugnant to civil society is that *whatsoever a man does against his conscience is sin*; and it dependeth on the presumption of making himself judge of good and evil. For a man's conscience and his judgment is the same thing; and as the judgment, so also the conscience may be erroneous. Therefore, though he that is subject to no civil law sinneth in all he does against his conscience, because he has no other rule to follow but his own reason, yet it is not so with him that lives in a commonwealth, because the law is the public conscience, by which he hath already undertaken to be guided. Otherwise, in such diversity as there is of private consciences, which are but private opinions, the commonwealth must needs be distracted, and no man dare to obey the sovereign power farther than it shall seem good in his own eyes (*L* xxix.6-7).

Whether or not the doctrine "*whatsoever a man does against his conscience is sin*" is a "repugnant" one depends on whether or not one is acting on private conscience without a commonwealth or is acting on private conscience within a commonwealth. If without a commonwealth, then it is sin to act against private conscience, because a person would then

be acting against her own laws of reason and sin is the violation of law.[19] So it would be a mistake to say that private conscience has no authority over us. Rather, private conscience *without a commonwealth*, i.e. within the state of nature, is authoritative. However, if within a commonwealth, then it is not a sin to act against private conscience because in a commonwealth, a person leaves the power to judge what is good or bad to the sovereign. So, if a person violates private conscience, but is following public conscience, then she does not sin.[20] In other words, within a commonwealth, it is public conscience that has authority over us. Regardless of our private judgment of what we ought to do, we are required to do what public conscience dictates.[21]

It should be noted that Hobbes thinks the sovereign's authority is limited to actions, not thoughts:

> There is another error in their civil philosophy, which they never learned of Aristotle (nor Cicero, nor any other of the heathen): to extend the power of the law, which is the rule of actions only, to the very thoughts and consciences of men, by examination and inquisition of what they hold, notwithstanding the conformity of their speech and actions…. [T]o force him to accuse himself of opinions, when his actions are not by law forbidden, is against the law of nature (and especially in them, who teach that a man shall be damned to eternal and extreme torments if he die in a false opinion concerning an article of the Christian faith) (*L* xlvi.37).[22]

[19] "A *sin* is not only a transgression of a law, but also any contempt of the legislator. For such contempt is a breach of all his laws at once" (*L* xxvii.1). It may be that Hobbes thinks that violating the laws of reason without a commonwealth is to show contempt for the legislator of Nature, i.e. God.

[20] *EL* xxv.12

[21] I am admittedly setting aside Hobbes' account of what we ought to do when the sovereign's laws contradict the laws of nature and laws of God, if such is even possible. But even if there were a conflict, Hobbes appears to hold up the authority of public conscience, at least over action within a commonwealth: "For [violating the sovereign's laws] is always a violation of faith, and consequently against the law of nature, which is the eternal law of God" (*L* xlii.131). For more in-depth discussion on conscience within and without the commonwealth, see Edward Andrew's "Hobbes on Conscience within the Law and without" (1999). Andrew also provides a useful discussion of the views and criticisms of conscience around Hobbes' time in his book *Conscience and is Critics* (2001).

[22] Also see *EL* xxv.3. "Conscience" in this instance should be read as private conscience.

So even though we ought to act in accordance with public conscience, even when doing so contradicts our private conscience, civil law can never extend to anything beyond the action. Thoughts and intentions are only considered insofar as they lead to some action. But one cannot be said to violate public conscience by simply thinking or intending alone, since public conscience only extends to actions.

Hobbes in some ways retains the Medieval account of conscience as knowledge, in the form of shared awareness. But he understands it in the context of a commonwealth in which people have given their power to judge what is good and bad to the sovereign, who, through civil law, form a *public* conscience. So conscience is shared knowledge in that it is civil law, open and available to all within a commonwealth. At the same time, Hobbes recognizes that some have begun to corrupt the use of 'conscience' to refer to their private judgments. It is this corrupted use that has led to much suspicion of those who justify their actions by appealing to their conscience. If conscience is nothing more than one's private judgment, there is ground for questioning the reliability and authority of conscience, resulting in the reinforcement (or dislocation) of the original motive.

1.3 Bishop Butler and the Voice of God

Commentators have long been aware of Bishop Butler's influence on Smith's work, and especially with regards to conscience. According to Butler, conscience is a distinct part of human nature that has the role of judging other faculties and principles, and so by nature holds authority over them. Smith's discussion of our moral faculties is strikingly similar to that of Butler's. And so it is tempting to view Smith as simply adopting Butler's account of conscience. I think doing so is a mistake. But to see why, we first need to understand Butler's

account of conscience, which I present here. In Chapter 2, I will present Smith's account, and highlight the various points of agreement and disagreement between Butler and Smith.

Butler's fullest illustration of conscience comes in the preface and the first three sermons of *Fifteen Sermons Preached at the Rolls Chapel* (1729).[23] According to Butler, human nature consists of several parts: "[a]ppetites, passions, affections, and the principle of reflection" (*Sermons* Pr.14). The 'principle of reflection' is what Butler elsewhere calls 'conscience' (e.g. *Sermons* Pr.14, Pr.19, Pr.25, 1.8, 2.3, 2.8). However, using 'conscience' and 'principle of reflection' interchangeably is misleading. Appetites, passions, and affections are genuine principles, i.e. *motive for action*: "We are plainly constituted such sort of creatures as to reflect upon our own nature. The mind can take a view of what passes within itself, its propensions, aversions, passions, affections, as respecting such objects, and in such degrees; *and of the several actions consequent thereupon*" (*Sermons* 1.8, emphasis added; also see *Sermons* 2.8, 2.9, 2.14, and 2.15).[24] And in *Sermons* 1, Butler covers two key principles ("tendencies", "propensions", "instincts"): self-love and benevolence. But conscience is *not*, strictly speaking, a motive for action. Rather, it is a faculty, "a capacity of reflecting upon actions and characters, and making them an object to our thoughts" (*Analogy* Diss.II.1). This capacity is not one of the other principles mentioned earlier: "There can be no doubt but that several propensions or instincts, several principles in the heart of man, carry him to society, and to contribute to the happiness of it, in a sense and a manner in which no inward principle leads him to evil. These principles, propensions, or instincts which lead him

[23] *Fifteen Sermons Preached at the Rolls Chapel* in *The Works of Bishop Butler* (Rochester, NY: University of Rochester Press, 1729/2006). Passages from *Fifteen Sermons* will be referred to by sermon, or preface ('Pr'), and paragraph number, prefaced by *Sermons*. Passages from *The Analogy of Religion* (1736/2006) will be referred to by chapter, or dissertation ('Diss'), and paragraph number, prefaced by *Analogy*.

[24] "Ground of action; motive – e.g. 'As no *principle* of vanity led me first to write it, so much less does any such motive induce me now to publish it.' (Wake)" (from Johnson's *Dictionary*, 7th ed., 2:201). Samuel Johnson's *Dictionary* is available in electronic form through the Eighteenth Century Collections Online (ECCO) from Gale Digital Collections (gdc.gale.com).

to do good are approved by a certain *faculty* [i.e. conscience] within, *quite distinct from these propensions themselves*" (*Sermons* 2.2, emphases added). Conscience, then, is properly speaking a faculty, and not a principle.[25]

Of course in forming judgments, this faculty does motivate a particular course of action, but it does so by adding to an existing motive, i.e. an existing appetite, passion, or affection. In judging an action right (or wrong), conscience simply affirms (or condemns) a motive, resulting in a settling (or unsettling) the original motive:

> And that this faculty tends to restrain men from doing mischief to each other, and leads them to do good, is too manifest to need being insisted upon. Thus, a parent has the affection of love to his children: this leads him to take care of, to educate, to make due provision for them; the natural affection leads to this: but the reflection that it is his proper business, what belongs to him, that it is right and commendable so to do; this, added to the affection, becomes *a much more settled principle*, and carries him on through more labor and difficulties for the sake of his children, than he would undergo from that affection alone (*Sermons* 1.8).

The motive to care for one's child is originally one's natural affection. Conscience simply affirms this motive, to the degree or strength that it is, by judging it to be good. As a result, one is assured in moving forward with that original motive, with acting on that natural affection. The same is the case if conscience disapproves of an original motive. If one is

[25] In interpreting conscience as a faculty, and not a principle, I am agreeing with the interpretations of D. D. Raphael (1949) and Douglas Langston (2001), among others. It is unclear whether Terence Penelhum (1985) and Stephen Darwall (1995) interpret conscience as a faculty. Penelhum primarily uses 'principle', and Darwall explicitly describes conscience as one of the several motives of action, while at several other times referring to it as a faculty. But neither seems to consider the distinction between principle and faculty that I present here.

There were two primary uses of 'faculty' in his time, the former broad and the latter narrow: "[t]he power of doing anything; ability; whether corporal or intellectual" and "Powers of the mind, imagination, reason, memory" (from Samuel Johnson's *Dictionary*, 7th ed., 1:383). In presenting Butler as adopting conscience as an "unerring faculty", Douglas Langston may be understanding 'faculty' in the latter, narrower sense, what Langston calls 'substantial entity' (see Langston 2001). But there is evidence that Butler is using faculty in the broader sense. At one point, he compares conscience to the many other 'natural facult[ies] or power[s]', and does not limit his discussion to distinct 'Powers of the mind', including the faculty of speech (*Sermons* 4.7-8). It may be that he moves back and forth between the two uses. I remain neutral on the issue.

motivated by excessive self-love to inflict harm on another, conscience would disapprove. And in disapproving of this particular motive, one finds discomfort in moving forward with this excessive self-love. Thus, through its approving and disapproving, conscience motivates actions indirectly by motivating one to *change* the motive from which one acts.

Butler's primary aim in the *Sermons* is to show that we ought to follow our conscience as dictated by our natural constitution. Butler does not say much regarding the *accuracy* of conscience in forming judgments of our other faculties and principles. It is easy to think Butler simply assumes that conscience cannot err, given he thinks conscience always directs us towards what is good and should always be obeyed (to be discussed below). But as we saw with both Bonaventure and Aquinas, we can draw distinctions here between the general drive towards the good (which Bonaventure called 'synderesis'), general (first) moral principles, and particular moral judgments derived through reasoning from the more general moral principles. And we can hold that conscience provides a general drive towards the good and is infallible about the general moral principles without accepting that conscience is infallible in the particular moral judgments formed. Butler does not discuss conscience in these terms, but may have a picture like this in mind. He recognizes that people do come to different moral judgments. They approve and disapprove differently in the same cases. And unless Butler is a relativist about moral judgments (which he is not) then at least some of these judgments must be erroneous. But Butler attributes these errors not to conscience, but rather to the "peculiarities in their own temper" or "the effect of particular customs", both distinct from conscience itself (*Sermons* 2.1). He also points to inattentiveness as a source of differences (*Sermons* 2.1), which may be akin to errors in reasoning in Bonaventure's and Aquinas' accounts. But even so, it is unclear whether *conscience* is erring since, in all these

cases, the source of error is something distinct from conscience. In the least, Butler thinks the first moral principles, including the imperative that we ought always to obey our conscience (see *Sermons* Pr.26), given by nature, cannot err.[26]

Butler continually stresses that conscience by its nature has authority over the other faculties and principles.[27] This authority is not *mere power*, which is simply the strength of the motive, the way it is for the various appetites, passions, and affections. For instance, right now my self-love is much stronger than my benevolence, and so I act out of self-love rather than benevolence. But conscience does not necessarily overpower the other motives in this manner. Authority is not a matter of the strength; it is a matter of *superintendency*. Conscience has authority the same way that a government has authority over a country, *regardless of its power to exercise that authority*. Conscience is to be considered:

> as being superior, as from its very nature manifestly claiming superiority over all others: insomuch that you cannot form a notion of this faculty, conscience, without taking in judgment, direction, superintendency. This is a constituent part of the idea, that is, of the faculty itself: and, to preside and govern, from the very economy and constitution of man, belongs to it. Had it strength, as it has right; had it power, as it had manifest authority; it would absolutely govern the world (*Sermon* 2.14).

Each part of the constitution of humans has by its nature a particular role or function. And the role of conscience is to judge other faculties, principles, and actions. ('Judging' here should be understood broadly to include directing, superintending, presiding, and governing.

---

[26] The discussion here is limited to creatures with all the different faculties and principles common to "the bulk of mankind". Butler recognizes that there may be individuals who simply lack certain capacities or principles, like those "who are in great measure without the natural affections toward their fellow-creatures" and those "without the common natural affections to themselves" (*Sermons* 1.13). Likewise, a person may lack the capacity for reflection, i.e. conscience, altogether. But the concern here is explaining the apparent differences in judgments of those that seem to have all the same faculties and principles.

[27] Though Butler's approach in the *Sermons* is to appeal to human nature to show that we have an obligation to obey our conscience, he also thinks we can come to the same conclusion by considering the abstract relations of things, similar to Hobbes' approach (see *Sermons* Pr.12).

By approving and disapproving, conscience is *telling*, metaphorically, what the other principles and faculties to do.) In being placed in this role, conscience is placed in a position of authority, just as someone appointed to the position of a judge, in virtue of that appointment, has the authority of that office. Because conscience has this function, we cannot make sense of it without viewing it as authoritative.[28]

Of course, as Stephen Darwall rightly notes, that conscience has this role alone does not answer the more important question of whether or not we *should obey* conscience at all times (Darwall 1995, 267-269). Conscience simply has the function of approving and disapproving. But why should we change our motives and actions to align with conscience's approvals and disapprovals? Why are we obligated to obey conscience always? The answer is that we have an obligation to obey the law of our nature, and the law of our nature is that conscience guides and directs us for nature has given conscience that role. Thus, we have an obligation to make conscience our guide, i.e. to obey it.[29] But this answer needs some unpacking.

First of all, Butler distinguishes between three senses of 'natural', from broad to narrow:

[i] "some principle in man, without regard either to the kind or degree of it" (*Sermons* 2.5)

[ii] "those passions which are strongest, and most influence the actions" (*Sermons* 2.6)

[iii] "What that is in man by which he is *naturally a law to himself* is explained in the following words: *Which show the work of the law written in their hearts, their consciences also bearing witness, and their thoughts the meanwhile accusing or else excusing one another*" (*Sermons* 2.8)

[28] Admittedly, I am glossing over an important distinction between constitutional order and teleological/ functional order (see Darwall 1995, 244-270). It is unclear whether Butler holds this distinction, or if he does whether he feels justified in collapsing the two.

[29] Butler summarizes the argument in *Sermons* 3.9.

Under the first sense, anything that is a principle of humans would count as natural. So disobeying conscience by acting on a different principle of action would not be unnatural in this sense. Under the second sense, whichever principle of action is strongest, and hence the one from which we act, would count as natural. But once again, disobeying conscience would not be unnatural if the principle of action was simply strong enough. Instead, Butler is concerned with giving an account of how following conscience is 'natural' in the third sense, "in the strictest and most proper sense" (*Sermons* 2.10). The third sense is a claim about how what we *ought* to do, how we *ought* to function, as intended by the Author of Nature. It is in this sense that disobeying conscience is unnatural, for conscience was intended to guide, judge, and direct the other faculties and principles.

Next, Butler distinguishes between the nature (in this third sense) of the different parts of humans and the *whole* nature of humans. We can speak of each faculty or principle as intended for some purpose. For instance, self-love is intended for the preservation of self, and benevolence for the benefiting of others. Conscience is intended for judging other faculties and principles. But when Butler speaks of an obligation to obey the law of our nature, he is referring to our *whole* nature, which consists not only of the individual parts and their roles but also of their relations to one another. And given the parts and their roles, our whole nature includes conscience's exercising authority over the other faculties and principles, to adjust and correct for those principles. It is when conscience and the other faculties and principles interact in this manner that we are acting in accordance with our nature, as the Author of Nature intended, and hence virtuously (*Sermons* 2.4).

Finally, Butler assumes that we have an obligation to "*act agreeably to the constitution of our nature*" (*Sermons* 2.15). He is drawing on teleological accounts of ethics

of Ancient philosophers, most explicitly in *Sermons* 2.4: "And yet the ancients speak of deviating from nature as vice, and of following nature so much as a distinction, that according to them the perfection of virtue consists therein". What we ought to do is follow our nature, understood in the third sense of what nature has intended for human beings. He is also drawing on the writings of St. Paul, who asserts that our moral (or divine) obligations are embedded in our human nature (see *Sermons* 2.4, 2.8). So what Butler means by his repeated claims that humans are a law to themselves is that our moral obligations are written into our nature, in not only the roles given to each part and their relations to one another, but also the way that conscience judges. Like Aquinas and Bonaventure, moral principles are innate (though we may not always be aware of them). And for Butler, these principles are written in the way that conscience judges, in the fact that conscience approves of some motive and not others in various circumstances. Our obligation to our conscience, then, stems from our obligation to nature. And so Butler says in response to the question of *why* we should obey our conscience: "Your obligation to obey this law, is its being the law of your nature. That your conscience approves of and attests to such a course of action, is itself alone an obligation" (*Sermons* 3.5).[30]

1.4 Immanuel Kant and Self-Awareness

Though conscience was not a central feature of Kant's moral theory, he did give conscience an important role in our everyday moral life. Conscience is how we recognize whether or not we are living up to our duties. And in doing so, it serves as motivation for doing what is right. Moreover, a feature of a virtuous person is a highly developed

---

[30] I have drawn my interpretation of Butler on the authority of conscience primarily from Raphael (1949) and Darwall (1995), but my interpretation is consistent with most others, e.g. O'Brien (1991), Penelhum (1985), Sturgeon (1976).

conscience. His account is particularly of interest here because as we shall see it is remarkably similar to Smith's, maybe more so than Butler's is.[31] I will present Kant's account here, which will draw heavily from the most comprehensive presentation of Kant on conscience, "Four Conceptions of Conscience" (1998); I will highlight the points of divergence between Kant and Smith in the next chapter.[32]

In *Metaphysics of Morals*, Kant defines conscience as the "[c]onsciousness of an *internal court* in man ('before which his thoughts accuse or excuse one another')" (*MM* 6:438).[33] Kant is drawing on the familiar metaphorical use of conscience as self-awareness, in which we are witnesses to ourselves.[34] This court includes three figures: the prosecutor, the advocate (defense attorney), and the judge. So, it would be a mistake to identify conscience with any one of these figures.[35] To avoid contradiction, Kant thinks the agent, the one under trial in this court, is distinct from the court, and all the figures in the court.

---

[31] See Fleischacker 1991. Fleischacker also argues that Kant may have been familiar with Smith's work in ethics. There is clear evidence that Kant was at least familiar with *Wealth of Nations*, but the extent of his familiarity with Smith is unclear.

[32] A similar overview is provided by William Lyons in "Conscience – An Essay in Moral Psychology" (2009). He also provides his own model of conscience, one that is similar to Kant's account in many respects.

[33] *Metaphysics of Morals* (Cambridge: Cambridge University Press, 1797/1996). Passages from *Metaphysics of Morals* will be referred to by volume and page numbers in the Akademie edition, prefaced by *MM*; and passages from *Religion Within the Boundaries of Mere Reason* by *Religion*.

Kant's account of conscience is dispersed throughout several works: *Metaphysics of Morals* 6:233-5, 6:394, 6:400-1, and 6:437-41; *Religion Within the Boundaries of Mere Reason* (6.185-187); *Groundwork on the Metaphysics of Morals* (4:422 and 4:404), and *Second Critique* (5:98). I will draw primarily from *Metaphysics of Morals*, and refer to other texts when helpful.

[34] As mentioned earlier, Hobbes recognized this metaphorical use, but did not see it as the proper definition of 'conscience'. Thomas Hill, who provides the most comprehensive account of Kant on conscience, gives Kant's account of conscience the title of the "Kantian conception" of conscience. I think this way of labeling Kant's account is grossly misleading, since Kant is adopting a use of 'conscience' that was not all too uncommon in the 17th and 18th centuries. Kant's account is also remarkably similar to Smith's, as we shall see later, and there is evidence that Kant was familiar with at least some of Smith's writings. A better way of characterizing Kant's account is as one that was commonplace but adopted to fit into Kant's wider moral project.

[35] Hence, anytime Kant speaks of conscience as judging, as acquitting and condemning, one should interpret Kant as speaking of the judgments rendered by the *tribunal* after all figures play their role. Hill should be read similarly in his presentation of Kant, e.g. "[C]oncience as "an inner judge" that condemns (or acquits) one for inadequate (or adequate) effort to live according to one's best possible, though fallible, judgments about what (objectively) one ought to do" (Hill 1998, 16).

Kant also at times treats conscience as the tribunal instead of a consciousness of this tribunal, for instance, when he speaks of conscience as administering a verdict or when he states "the moral self-awareness

> Now, this original intellectual and (since it is the thought of duty) moral predisposition called *conscience* is peculiar in that, although its business is a business of a human being with himself, one constrained by his reason sees himself constrained to carry it on as at the bidding *of another person*…. But to think of a human being who is *accused* by his conscience as *one and the same person* as the judge is an absurd way of representing a court, since then the prosecutor would always lose. – For all duties a human being's conscience will, accordingly, have to think of *someone other* than himself (i.e., other than the human being as such) as the judge of his actions, if conscience is not to be in contradiction with itself. This other may be an actual person or a merely ideal person that reason creates for itself (*MM* 6:438-9).

The court will always conclude in favor of the accused if the accused is also the judge. The court, then, must be represented as independent of the accused. Hence, the judge must be a projection of an actual or ideal person. This projecting is what leads to our treating the judge as an internal representative of God.[36] So our awareness of the court is not awareness from our playing the role of the various figures in the court, but rather as the accused witnessing the court.

The internal court does not form any moral judgments. Instead, the court takes an agent's moral judgments, whatever it may be, lays them before the her, and determines whether or not she lived up to her own standards. As Hill puts it, "[C]onscience presupposes but is not the same as 'moral judgment' in the sense of 'drawing from the moral law a more determinate specification of our duties'" (Hill 1998, 35). So in being aware of this court, an agent is aware of her own moral judgments as well as the court's verdict on whether or not

---

of conscience" (MM 6:439). I think these are mere slips, but I am giving greater weight to his formal definition of 'conscience'. But given that Kant is not always consistent in his use of 'conscience', it is understandable why Hill does not distinguish between conscience as the court and conscience as *consciousness* (awareness) of this court.

[36] "Now since such a moral being [i.e. the projected judge] must also have all power (in heaven and on earth) in order to give effect to his laws (as is necessarily required for the office of judge), and since such an omnipotent moral being is called **God**, conscience must be thought of as the subjective principle of being accountable to God for all one's deeds" (*MM* 6:439).

she lived up to her own standards. And this awareness is involuntary: "this authority watching over the law in him is not something that he himself (voluntarily) *makes*, but something incorporated into his being" (*MM* 6:438). Hence, conscience should be disconnected from any moral deliberation, reasoning, or judgment (Hill 1998, 23).

Under this picture, one can say that conscience can provide us with moral knowledge in the sense that we become aware of our own moral judgments and whether or not we lived up to those judgments. But presenting conscience as such is not quite accurate. For one, conscience is not involved in making the moral judgment, as it was for Bonaventure and Aquinas. Conscience does not assent to nor reject any deontic proposition. The moral judgment is already made, and conscience simply makes us aware of it. Since conscience does not form any moral judgments, conscience cannot *err* in making moral judgments.[37] Any mistake would be in the forming of the moral judgment, in practical reason, distinct from conscience. But more importantly, conscience makes us aware of our moral judgment *within a process of assessing whether or not we lived up to that judgment*. Insofar as awareness is a form of knowing, we know our own moral judgments as a *byproduct* of coming to know whether or not we lived up to that those moral judgments. We may have already known what is right or wrong, assuming our moral judgment is correct, prior to our conscience determining whether or not we have lived up to those standards. So what conscience actually provides is not moral knowledge, i.e. knowledge of any deontic proposition, but rather knowledge of some empirical fact, of whether or not we acted in accordance with our own moral judgment.

[37] Hill notes that there is one way in which we might have an erroneous conscience: "Errors of conscience, if there were any, would have to be a matter of failing, even after we raised the question, to recognize either the fact that what we intentionally did was (or was not) against our best moral judgment or the fact that we had (or had not) exercised due care to determine whether our act was right" (Hill 1998, 35). But Kant does not seem to consider this possibility (see *MM* 6:401).

Whether or not we come to the correct moral judgment, conscience motivates us to do the right action by inflicting psychological discomfort when we fail to live up to our duties. At the same time, this motivation is a mere psychological fact; we are not obligated to obey our conscience. For one, conscience is not the proper motive for doing the right action. If we do what is right out of fear of our conscience, then we are doing the right action for fear of the psychological discomfort resulting from the tribunal's decision. However, such a motive would be no different from obeying the law for fear of criminal prosecution, which would be the wrong motivation for Kant. It is true that conscience makes us aware of our duties in assessing whether or not we lived up to those duties. So, conscience often, if not always, comes with awareness of one's duties. But there is a difference between doing what is right for the sake of those duties and doing what is right for the fear of psychological sanctions by conscience (Hill 2002, 352). If by "acting from conscience", one is acting from a desire to avoid the pangs of conscience, then one is not acting for the right reason. But if by "acting from conscience", one is acting out of respect for the moral law which one is aware of through conscience, then one is acting for the right reason. But then the motive is not conscience, but rather a respect for the moral law.

Also, if we had a duty to obey our conscience, then we would need some mechanism for determining whether or not we lived up to such a duty. But conscience is that mechanism. What we would need, then, is a *second* conscience to determine whether or not we lived up to the duty to obey our conscience, and yet a *third* conscience to determine whether or not we lived up to the duty to obey our *second* conscience, and so on (*MM* 6:401). Conscience simply does not operate in this manner. It speaks "involuntarily and unavoidably", and we

can choose whether or not to heed to its judgment by acting from duty. But there is no additional duty to obey conscience, only a duty to obey the moral law.

Though we have no duty to obey our conscience, Kant thinks we have a duty to *cultivate* our conscience, "to sharpen one's attentiveness to the voice of the inner judge and to use every means to obtain a hearing for it" (*MM* 6:401). Why might this be? According to Kant, we have a more general duty to cultivate virtue. Virtue is "the strength of man's maxims in fulfilling his duty", but done freely, "in accordance with a principle of inner freedom, and so through the mere representation of one's duty in accordance with its formal law" (*MM* 6:394). In other words, virtue is based on "the *capacity* to master one's inclinations when they rebel against the law" (*MM* 6:383). Conscience is crucial for cultivating this capacity for free self-constraint because it lays our duties before us and shows us when we have failed to live up to them so that we can better ourselves by freely correcting for our failings and aligning ourselves with our duties.[38] And thus, the first command of all duties to oneself is "*know* (scrutinize, fathom) *yourself*", your "heart" or the maxim from which you act (*MM* 6:441), because it is only when we know whether or not we are living up to our duties that we are able to improve ourselves and cultivate virtue. Kant, then, thinks that though conscience is not authoritative in that we have no obligation to obey it, we still have an obligation to cultivate our conscience so that we can fulfill our duty to cultivate virtue.

Kant, then, takes conscience to be the judging of oneself, restricted to whether or not one has acted in accordance with one's duties. And unlike Aquinas, Bonaventure and Butler,

[38] Kant makes a similar claim in *Religion* when he states the one who seeks comfort from "*moral* sufferings, the reproaches of their conscience" should have one's conscience "*stirred up* and *sharpened*, in order that whatever good yet to be done, or whatever consequences of past evil still left to be undone (repaired for), will not be neglected, in accordance with the warning" (*Religion* 6:78 footnote).

Kant does not think we have a duty to obey our conscience. However, we have a duty to cultivate our conscience, to sharpen it and make it more attentive, because doing so is necessary for cultivating virtue. As we shall see, this connection between conscience and virtue is an ever-present theme in Smith's moral theory.

1.5 John Stuart Mill and the Sense of Duty

In Chapter 3 of *Utilitarianism*, Mill provides a brief account of conscience and its authority. The aim of this chapter is to give a general account of the binding force of the moral duties (i.e. the principle of utility) laid out in the earlier chapters: "What is its sanction? what are the motives to obey it? or more specifically, what is the source of its obligation? whence does it derive its binding force?" (*U* 3.1).[39] There are external sanctions like the disapprobation of others and punishment under the law that motivate us to do our duty. But he spends most of his energy on the *internal* sanction of duty:

> The internal sanction of duty, whatever our standard of duty may be, is one and the same—a feeling in our own mind; a pain, more or less intense, attendant on violation of duty, which in properly-cultivated moral natures rises, in the more serious cases, into shrinking from it as an impossibility. This feeling, when disinterested, and connecting itself with the pure idea of duty, and not with some particular form of it, or with any of the merely accessory circumstances, is the essence of Conscience (*U* 3.4).

Once we understand our duties, we tend to feel pain whenever we violate those duties. And when this pain is from the simple recognition that we violated a duty (generally, not any particular duty), then that feeling is conscience. For instance, suppose that according to the principle of utility, I have a duty to give $100 to Oxfam, but I decide to buy a new pair of

[39] *Utilitarianism* (Oxford: Oxford University Press, 1863/1998). Passages from *Utilitarianism* will be referred to by chapter and paragraph number, prefaced by *U*.

shoes. But when I realize that I did not do my duty, I feel terrible. This feeling is connected to the thought that I violated a duty, not the duty to give to Oxfam in this particular situation but rather *any* duty; this feeling is conscience. Conscience, then, is nothing more than a "subjective feeling in our own mind" associated with the idea of duty. So Mill is equating conscience to the pangs of conscience, the feelings of guilt or remorse we have when we realize we did not do what we ought to do.[40]

For Mill, conscience is not necessarily the source of moral knowledge. Though conscience relies on an awareness, belief, or knowledge of one's duty, conscience itself is a particular kind of feeling in response to the recognition of violating one's duty. This affective response can bring to awareness various duties of which we were previously unaware. But conscience is not the awareness nor does it need to provide awareness. It is rather a particular kind of feeling in response to what we take to be our duties. In this way, his account is similar to Bonaventure's synderesis: that spark of conscience that moves us towards what is good. Moreover, like Bonaventure, who thought synderesis motivated in virtue of its being a dictate of practical reason (which instructs a person on what to do), Mill thinks conscience motivates because it is an affective response to the violation of one's duty. However, for Bonaventure, the dictate of practical reason itself is the motivating force, whereas for Mill, the *affective response* to such a dictate is the motivating force. Furthermore, whereas

---

[40] In a letter to W. G. Ward (1859), Mill draws a distinction between two similar feelings based on whether or not the feeling is caused by something internal to oneself:

> The pains of conscience are certainly very different from those of dread of disapprobation; yet it might well be, that the innumerable associations of pain with doing wrong which have been riveted by a long succession of pains undergone, or pains feared or imagined as the consequence of wrong things done, or of wrong things which we have been tempted to do (especially in early life), may produce a general and intense feeling of recoil from wrongdoing in which no conscious influence of other people's disapprobation may be perceptible (Mill from *The Collected Works of John Stuart Mill* Vol. 15 Letter 423).

The "pains of conscience" is distinct from the "dread of disapprobation" *because* the former is essentially an internal sanction, caused by one's recognition of one's own violation of a duty, whereas the latter is an external sanction, caused by the presence of external observers.

Bonaventure thought synderesis was innate and can never be extinguished, Mill leaves open the question of whether or not conscience is innate or acquired (see *U* 3.7-8), and believes that some may not have this feeling.

In the early accounts of conscience, we are bound to the dictates of conscience because it is or provides knowledge of what we ought to do, which are commandments of God or morality. In other words, the authority of conscience is derived from its providing moral knowledge. But for Mill, conscience does not necessarily have any connection to moral knowledge. Whether what we take to be duties align with our actual duties (in accordance with the principle of utility) depends on our training.[41] Furthermore, even if conscience provides awareness of what to do, its authority is not derived from providing awareness of our duties. Rather, its "binding force" is derived from its being a feeling:

> Its binding force, however, consists in the existence of a mass of feeling which must be broken through in order to do what violates our standard of right, and which, if we do nevertheless violate that standard, will probably have to be encountered afterwards in the form of remorse. Whatever theory we have of the nature or origin of conscience, this is what essentially constitutes it (*U* 3.4)

The binding force of conscience is no different than the external sanctions of punishment. We are bound to do our duty because doing otherwise is difficult: we are faced with (the prospect of) strong negative feelings. Conscience serves to prevent us from acting against our

[41] "Like the other acquired capacities above referred to, the moral faculty, if not a part of our nature, is a natural outgrowth from it; capable, like them, in a certain small degree, of springing up spontaneously; and susceptible of being bought by cultivation to a high degree of development. Unhappily it is also susceptible, by a sufficient use of the external sanctions and of the force of early impressions, of being cultivated in almost any direction: so that there is hardly anything so absurd or so mischievous that it may not, by means of these influences, be made to act on the human mind with all the authority of conscience" (*U* 3.8). So what we take to be our duties will be more or less accurate depending on our upbringing. But Mill also thinks that there is a "powerful natural sentiment" that directs our moral development toward the acceptance of the principle of utility: "the social feelings of man; the desire to be in unity with our fellow creatures, which is already a powerful principle in human nature, and happily one of those which tend to become stronger, even without express inculcation, from the influences of advancing civilization" (*U* 3.10).

duty through its intensely negative feeling that we would rather avoid. The stronger this feeling is, the more binding force we feel when considering acting against our duty. Of course, this force is only felt by those who actually have this negative emotion: "this sanction has no binding efficacy on those who do not possess the feelings it appeals to" (*U* 3.5). So if a person lacks a conscience, in Mill's sense, then she feels no remorse in violating her duty, and will not feel bound to do her duty in the way that someone with a conscience would.

Now, we might find this account problematic. Since there is no binding force on those who do not have a conscience, then there is nothing we can say to them when they ask why they should do their duty. We can point to external sanctions, like fines or imprisonment, but nothing beyond that, nothing internal to the agent. But for Mill, this is not a problem. For one, he seems to treat these cases as limited in scope. Most people do have this feeling. At one point, he refers to it as "the conscientious feelings of mankind", and then points out "the feelings exist, *a fact in human nature*, the reality of which, and the great power with which they are capable of acting on those in whom they have been duly cultivated, are proved by experience (*U* 3.5, emphasis added). He also thinks that this feeling is founded in a powerful, shared natural sentiment: "the social feelings of mankind; the desire to be in unity with our fellow creatures" (*U* 3.8). Mill is speaking specifically of the principle of utility, but he is also pointing out in this section that as a generalization, people share this social sentiment. This sentiment is the source of the feelings of duty, i.e. conscience. So even if there are instances in which people are devoid of conscience, he thinks these are the exceptions.

Furthermore, Mill thinks he is in no worse position than any other moral theory because for people without conscience, *no* internal sanction will motivate: "On them morality of any kind has no hold but through the external sanctions" (*U* 3.5). If upon recognizing her

duty *however understood*, an agent does not *feel* bound to do her duty, then there is no internal sanction at all *regardless of whatever moral theory she adopts*, utilitarianism or otherwise. Since conscience is a feeling connected to the thought of duty *generally*, if an agent lacks this feeling, then she lacks any motivational force to do her duty *generally*. Like Hume, Mill thinks that feelings are required for someone to be motivated to act; reason alone cannot motivate. If reason alone could motivate, then conscience, a feeling, would not be necessary; the recognition that some action is required by duty is sufficient for motivating action. But it is clear that Mill thinks *some* feeling is necessary to motivate action. And because conscience is a feeling associated with duty *generally*, apart from the particular moral theory one holds, failing to have a conscience means one just does not have that internal sanction, no sense of duty, no feeling 'binding one' to do one's duty.

There is also a slightly different worry that Mill recognizes. If we realize that moral obligations are not rooted in some "transcendental fact, an objective reality belonging to the province of 'Things in themselves'", but rather in a mere feeling, then we might not be motivated to do our duty (*U* 3.6):

> [I]f a person is able to say to himself, This which is restraining me, and which is called my conscience, is only a feeling in my own mind, he may possibly draw the conclusion that when the feeling ceases the obligation ceases, and that if he find the feeling inconvenient, he may disregard it, and endeavour to get rid of it (*U* 3.6).

So then for conscience to be effective in motivating us to do our duty, we cannot be aware that conscience is a mere feeling. By realizing that conscience is but a feeling, we might then come to question its authority and fail to be motivated to do our duty. But Mill is quite skeptical that realizing that conscience is a feeling somehow reduces its efficacy:

> Does the belief that moral obligation has its seat outside the mind make the feeling of it too strong to be got rid of? The fact is so far otherwise, that all moralists admit and lament the ease with which, in the generality of minds, conscience can be silenced or stifled. The question, Need I obey my conscience is quite as often put to themselves by persons who never heard of the principle of utility, as by its adherents. Those whose conscientious feelings are so weak as to allow of their asking this question, if they answer it affirmatively, will not do so because they believe in the transcendental theory, but because of the external sanctions (*U* 3.6).

Even if we recognize that our conscience is a feeling, we cannot simply turn it off. We still feel guilt and remorse when we realize we did not do our duty. And those feelings continue to motivate us to act in a particular way, even if we realize that they are but feelings. In fact, for Mill, the only ones who can even question their consciences are those with weak, or non-existent, consciences. Now this claim might be an overstatement; those with strong consciences can still doubt whether they should act in accordance with them *on the grounds that* they recognize that their consciences are but feelings. But Mill's point is that it is a mistake to think that conscience would somehow fail to be motivating if we did not also believe that moral obligation is a transcendental fact.

Mill, then, is focusing on conscience as a negative feeling, an internal punishment, that arises when we consider acting against our duty or when we recognize that we have not done our duty. In this way, conscience motivates us, just like punishments (or threat of punishments) by law or the disapprobation of others motivates us, to do our duty. There is nothing that binds us to our duties other than our feelings arising in response to these potential or actual sanctions. By treating conscience as a particular kind of feeling, Mill represents break from prior accounts that attempt to ground conscience strictly in practical reason. For Mill, though conscience relies on practical reason, insofar as practical reason

provides awareness or knowledge of our duties, it is part of the affective faculty that *responds* to practical reason.

1.6 Sigmund Freud and Psychoanalytics

In *Civilization and its Discontents*, Freud provides a detailed account of conscience. He does so in two stages: he first presents the origins of conscience, and then specifies some subtle distinctions between 'conscience', 'super-ego', and other related terms.[42] If one focuses on the first description of conscience, one will be led to think that conscience is the super-ego, a faculty of the mind. But in his second explanation, Freud recognizes that he has been far too loose with his remarks, and so highlights some subtle differences between these concepts, including conscience, which is a *function* of the superego, not the superego itself. In what follows, I will try to elaborate on the subtle, yet important, distinction that has so often been overlooked in the literature to bring a more accurate picture of what conscience is in Freud's theory.[43]

At the beginning of Chapter 7 of *Civilization and its Discontents*, Freud claims his aim in the chapter is to give "the history of the development of the individual" (*Civ* 83-84), but he is essentially giving an account of the development of the superego and conscience. The question he is trying address is: "What means does civilization employ in order to inhibit the aggressiveness which opposes it, to make it harmless, to get rid of it, perhaps?" (*Civ* 83). By 'aggression', Freud is referring to sadism, in which one desires the exercising of power, of control and mastery, over another. Such a desire, if fulfilled, would be harmful to the

---

[42] *Civilization and its Discontents* (New York, NY: W. W. Norton & Co., 1930/1989). Passages from *Civilization and its Discontent* will be referred to by page number, prefaced by *Civ*.

[43] More recently, Langston (2001) interprets Freud as stating conscience is a faculty. Langston then claims the decline in the discussion of conscience is a result of the failure to locate this faculty.

sustaining of a civilization, which requires that people live, to a large extent, in harmony. Now, initially, civilization uses external force to keep individuals from being aggressive. In the civilization's inhibiting aggression in individuals, those individuals begin to turn their aggression, which can no longer be directed toward others, in toward themselves. To do so, one sets aside a part of one's ego to become the 'super-ego', "in the form of conscience", which exercises aggression against the ego.[44] The aggression that was once directed at others to satisfy the ego is now directed at oneself so as to exercise control over oneself. Thus, the civilization need not exercise control for one now controls oneself such that one's aggression is not directed outward endangering others (*Civ* 84).

With the forming of a super-ego, we are now faced with internal tension whenever our ego conflicts with our super-ego. The super-ego demands one course of action, while the ego desires another. The result of such tension is the sense of guilt, which "expresses itself as a need for punishment" (*Civ* 84). Now all tensions between the super-ego and the ego will require that one did or intended to do something that one recognizes as 'bad' or 'sinful' (*Civ* 84). And so the sense of guilt "presupposes that one has already recognized that what is bad is reprehensible, is something that must not be carried out" (*Civ* 84). So the moral judgment is presupposed, and the sense of guilt is merely a product of the tension between the super-ego, which presents that judgment, and the ego, which fails to act or to intend to act in accordance with that judgment. Freud's account of the sense of guilt, then, resembles Kant's

[44] "His aggressiveness is introjected, internalized; it is, in point of fact, sent back to where it came from – that is, it is directed towards his own ego. There it is taken over by a portion of the ego, which sets itself over against the rest of the ego as super-ego, and which now, in the form of 'conscience', is ready to put into action against the ego the same harsh aggressiveness that the ego would have liked to satisfy upon other, extraneous individuals (*Civ* 84).

Given that the super-ego is a product of the development of an individual and not innate, one might think that the super-ego is not what I call a *F*aculty. Though it is true that historically, *F*aculty is often thought to be one of several innate Powers of the mind, it need not be. The Powers may be cultivated to some degree through experience. But I will leave this discussion to the side, and use 'faculty' to refer to any capacity.

account of conscience, in which conscience simply makes one aware of whether or not one lived up to one's own moral standards. The difference is that for Freud, this awareness only occurs when there is a tension or conflict between the ego and super-ego. Thus, the process only involves the negative feeling of guilt and never the corresponding positive feeling, praise, when one does live up to one's own moral standards.

At this stage, the super-ego simply controls the ego out of the fear of punishment from civilization. So, the super-ego is only concerned with controlling the ego insofar as one is able to avoid punishment. If one can avoid punishment, then the super-ego does not control and prevent the ego from exercising aggression. The internal tension that results in the sense of guilt only arises when the super-ego recognizes that the ego is driving one to do something that will result in external punishment. Moreover, because the super-ego is checking the aggression of the ego to avoid the punishment from civilization, what is 'bad' or 'sinful' that is presupposed by the sense of guilt is merely 'bad' or 'sinful' *according to civilization*.

But the super-ego does not stop at this stage:

> A great change takes place only when the authority is internalized through the establishment of a super-ego. The phenomena of conscience then reach a higher stage. *Actually, it is not until now that we should speak of conscience or a sense of guilt*. At this point, too, the fear of being found out comes to an end; the distinction, moreover, between doing something bad and wishing to do it disappears entirely, since nothing can be hidden from the super-ego, not even thoughts (*Civ* 86, emphasis added).

When the super-ego is initially formed, the super-ego controls the ego so as to avoid external punishment. Because the civilization only has direct access to what one *does*, the super-ego initially shows aggression based on what the ego *does*. But because the super-ego has access to all the intentions of the ego as well, eventually, the super-ego shows aggression toward the

ego for the intention alone. Thus, though the super-ego reflects the standards of civilization, it acts apart from civilization, enacting aggression when civilization would not. And it is only when the super-ego gains this independence from civilization, enacting aggression on the ego regardless of the possibility of external punishment, that we have conscience, and the sense of guilt, *proper*.

Thus, we conclude Freud's initial presentation of conscience and its formation. The discussion so far presents conscience as the super-ego, a faculty formed from a part of the ego to direct aggression inward, and one that eventually gains independence from the fear of external punishment. Freud, then, seems to hold an account quite similar to Butler's, in which conscience is the capacity to reflect on one's own intentions and actions and approve and disapprove of them. But unlike Butler, conscience is not an innate faculty, but rather obtained through experience with the rest of society. But as Freud himself recognizes, he has been, thus far, too loose with his language.[45] And so he tries to clarify his account of conscience and all the related concepts:

> The super-ego is an agency which has been inferred by us, and conscience is a function which we ascribe, among other functions, to that agency. This function consists in keeping a watch over the actions and intentions of the ego and judging them, in exercising a censorship. The sense of guilt, the harshness of the super-ego, is thus the same thing as the severity of the conscience. It is the perception which the ego has of being watched over in this way, the assessment of the tension between its own strivings and the demands of the super-ego…. We ought not to speak of a conscience until a super-ego is demonstrably present. As to the sense of guilt, we must admit that it is in existence before the super-ego, and therefore before conscience, too. At that time it is the immediate expression of fear of the external

[45] "Though it cannot be of great importance, it may not be superfluous to elucidate the meaning of a few words such as 'super-ego', 'conscience', 'sense of guilt', 'need for punishment' and 'remorse', which we have often, perhaps, used too loosely and interchangeably. They all relate to the same state of affairs, but denote different aspects of it" (*Civ* 100).

> authority, a recognition of the tension between the ego and that authority (*Civ* 100-101).[46]

This passage makes a few things clear. For one, conscience is *not* the same as the super-ego. Conscience is a *function* of the superego. And this function is to watch over the ego and to judge it. Thus, we should not speak of a conscience unless the super-ego is "demonstrably present", i.e. *functioning*. The sense of guilt, a feeling, is the perception of the tension between the ego and the (what once was external, but now) internal authority, i.e. super-ego. The greater the gap between the super-ego's demands and the ego's intent, the stronger the sense of guilt will be. But the sense of guilt is a product of conscience, not the conscience itself.

Freud's primary aim is not to give a moral theory, and so has nothing to say about the reliability of conscience in obtaining moral knowledge. But he does give an explanation for how we do form moral judgments. Unlike Bonaventure and Aquinas, Freud rejects the existence of any innate access to the first principles of morality, from which one can come to more particular moral judgments (*Civ* 84). Rather, the super-ego is influenced by something external to it that shapes its standard of judging what is good or bad to something different from the ego's. This influence is civilization. Hence, the super-ego is influenced by and often simply adopts the moral standards of civilization, whatever they are. But this account of the origin of conscience alone says nothing of the reliability of conscience as a moral guide, as a source of moral knowledge. So unless Freud is a cultural relativist, his account of conscience

---

[46] Though Freud is most explicit of these subtle distinctions here, he does make some attempt to draw them earlier in his work. For instance, he distinguishes conscience from the sense of guilt when he says "*conscience* – the readiness to feel guilty" (*Civ* 94). But Frued is right to note that he has been inaccurate in his usage of these terms throughout, e.g. "For it tells us that conscience (or more correctly, the anxiety which later becomes conscience)" (*Civ* 90).

does not determine whether or not conscience is a reliable way of obtaining moral knowledge.[47]

More interesting is Freud's description of the authority of conscience. Like Mill, Freud thinks there is nothing more to the authority of conscience than its motivational force. Though, like Mill, Freud takes the motivating force of conscience to be the sense of guilt, Freud thinks the sense of guilt is grounded in a more general feeling: the fear of the loss of love, specifically the love of those with the power to protect or to punish. Externally, this being would be other individuals with greater power, including the governing body and the rest of civilization. Internally, it would be the super-ego, which through conscience can punish with a strong sense of guilt.[48] So we are motivated to obey conscience because we are afraid of its punishment, i.e. guilt, but we fear this punishment because we want to retain the love of conscience. The punishment, guilt, is an indication that we have or will lose the love of conscience. Moreover, like Mill, the sense of guilt influences action only to the degree that it is felt. The greater the tension between the ego and super-ego, the greater the sense of guilt will be because the greater threat to the loss of love is. And the greater the sense of

[47] Freud's account is more or less what Hill calls the 'Extreme Cultural Relativist (ECR)' conception of conscience, except under ECR, conscience is identified as the feeling, i.e. the sense of guilt, and not the function of a faculty (Hill 1998, 21-23). In a footnote, Hill points to Ryle's account (Ryle 1940) as an example of ECR (Hill 1998, 44 fn. 16), but one can just as easily point to Freud. To be careful, though, it is unclear whether Freud is a cultural relativist, a moral nihilist, or neither. And though he sees the function of ethics as a way to maintain civilization, to protect it from the aggressiveness of individuals, he avoids making any value judgments about whether or not we should maintain civilization in this way, or whether or not some civilizations or the species as a whole is suffering from a kind of 'neurosis' (*Civ* 110-112).

[48] "Since a person's own feelings would not have led him along this path, he must have had a motive for submitting to this extraneous influence. Such a motive is easily discovered in his helplessness and his dependence on other people, and it can best be designated as fear of loss of love. If he loses the love of another person upon whom he is dependent, he also ceases to be protected from a variety of dangers. Above all, he is exposed to the danger that this stronger person will show his superiority in the form of punishment. At the beginning, therefore, what is bad is whatever causes one to be threatened with loss of love. For fear of that loss, one must avoid it. This, too, is the reason why it makes little difference whether one has already done the bad thing or only intends to do it. In either case the danger only sets in if and when the authority discovers it, and in either case the authority would behave in the same way" (*Civ* 85).

guilt, the more likely one is to act in accordance with the super-ego to avoid guilt, and hence avoid the threat of loss of love.

But as I said, Freud is not giving a moral theory, and so has nothing to say on whether or not we *should* follow our conscience and act to avoid the sense of guilt. He just claims that we do and to varying degrees depending on the strength of the super-ego. However, Freud does note that we identify the virtuous person with the one that has a more vigilant super-ego, which produces a stronger sense of guilt:

> For the more virtuous a man is, the more severe and distrustful is its behaviour, so that ultimately it is precisely those people who have carried saintliness furthest who reproach themselves with the worst sinfulness. This means that virtue forfeits some part of its promised reward; the docile and continent ego does not enjoy the trust of its mentor, and strives in vain, it would seem, to acquire it. The objection will at once be made that these difficulties are artificial ones, and it will be said that a stricter and more vigilant conscience is precisely the hallmark of a moral man (*Civ* 86-87).

The mark of (what people or society consider to be) a virtuous person is a "stricter and more vigilant conscience". This characterization of a virtuous person is the same as that of Kant (and as we shall see, of Smith as well).

1.7 C. D. Broad and Contemporary Psychology

Conscience is still discussed today in both psychology and philosophy. But there is still little consensus on exactly what conscience is. More recent approaches draw on C. D. Broad's view of conscience as a multifaceted phenomenon, which tries to capture the wide array of claims we make about conscience.[49] Broad thinks it is more helpful to focus on

[49] See Koops, Brugman, and Ferguson (2010) for an overview of the use of Broad's conception of conscience in contemporary psychology. Alternatively, Grusec applies Freud's account of conscience in contemporary psychology (Grusec 2006).

conscience as it is experienced by people. He therefore aims to describe the phenomenology of conscience:

> To say that a person "has a Conscience," when this phrase is used in its widest sense, is equivalent to asserting the following three closely connected propositions about him. (1) That he has and exercises the cognitive power of reflecting on his own past and future actions, and considering whether they are right or wrong; of reflecting on his own motives, intentions, emotions, dispositions, and character, and considering whether they are morally good or bad; and of reflecting on the relative moral value of various alternative ideals of character and conduct. (2) That he has and exercises the emotional disposition to feel certain peculiar emotions, such as remorse, feeling of guilt, moral approval, etc., towards himself and his own actions, dispositions, etc., in respect of the moral characteristics which he believes these to have. (3) That he has and exercises the conative disposition to seek what he believes to be good and to shun what he believes to be bad, as such, and to do what he believes to be right and avoid what he believes to be wrong, as such (Broad 1940, 118).

This experiencing of one's conscience involves various dispositions: cognitive, affective, and conative. These dispositions can work in conjunction with one another or in sequence by influencing one another. Conscience, then, is "a system of cognitive, emotional, and conative *dispositions*, and it is only when these dispositions are in operation that we have 'conscientious action' (Broad 1940, 125). Conscience, then, is not a distinct faculty (in Butler's sense), nor does it involve particular faculties at any given time. Rather, conscience is a complex system of dispositions, in which at different times these dispositions are active in different sequences and combinations to different degrees. Hence, conscience is experienced in different ways depending on the specific dispositions active at that particular time.[50]

[50] Langston adopts Broad's definition and uses it to explain why conscience is central to virtue (see Langston 2001).

In this chapter, I have presented different accounts of conscience and did nothing in the way of assessing their viability. And I do not wish to do so in any great detail here. But I do want to point out a general worry with these multifaceted approaches because it explains my motivation for abandoning contemporary approaches and tracing the history of conscience to find a more viable alternative. My worry is not novel. As Peter Fuss pointed out in his response to Broad, if we define 'conscience' in this broad sense, the term covers too much. Instead of denoting a feature of a human's moral psychology, it would seem to refer to the general capacity to be a moral being (Fuss 1964, 114). Conscience is responsible for moral judgments, moral feelings, and resulting moral actions. But discussion of conscience was not and is not aimed at capturing such a wide spectrum of experiences. Rather, the different approaches to characterizing conscience were attempts to capture a more particular phenomenon: our immediate awareness, knowledge, or belief of what is right to do in a particular situation, which also seems to strongly motivate us to act in accordance with that awareness, knowledge, or belief.

Of course, conscience is *associated* with a wide range of experiences. Most medieval thinkers thought that though conscience was a part of practical reason, it was also closely tied to the general desire to do what is good. But *conscience* did not include both reason and sentiments. Rather, conscience was a part of reason, and interacted with certain desires to give rise to right action. Freud thought that conscience was closely connected to feelings of guilt. But he distinguished between conscience and the feelings that *resulted* from conscience. What we need, then, is not a definition that simply includes cognitive, affective, and conative elements and leaves open the relationship among them as well as the degree to which each is active. We need a more systematic approach that explains exactly how each

element is connected to the other and where conscience is in the midst of these components, like the accounts provided by Aquinas, Bonaventure, Kant, and Freud. If the argument is that conscience involves all three components, then we need a better account of exactly how these components come together to form *conscience*, distinct from other components of a mature moral psychology.

1.8 Core Features of Conscience

As varied as these account are, we can notice a set of features that all accounts are attempting explain. Conscience was first and foremost thought to be something, whether a function or a special faculty, that provided moral *knowledge*. Conscience "tells" us what is right or wrong to do, what we ought or ought not to do. Even skeptics of conscience as a (reliable) source of moral knowledge, like Freud, attempt to explain why it at least *seems* to an agent that her conscience is providing moral knowledge.

Moreover, the moral knowledge that conscience gives is not of general principles, but rather of the *particular* case at hand. Though a few, like Bonaventure, claim conscience also provides moral knowledge of general moral principles, all are in agreement that conscience speaks to our present situation. For Aquinas, Bonaventure, and Hobbes, this process involved applying general moral principles to a particular situation. For Butler, conscience judged other faculties for how it functions in that specific instance. For Kant, conscience presents us with how we failed in some particular action to live up to our own moral standard. Even if we reduce conscience to some feeling, e.g. the feeling of guilt, that feeling is a response to a particular situation and a particular deed, e.g. I feel guilty about stealing money from my parents yesterday.

We can also see that among those who thought conscience provided moral knowledge, conscience was always believed to be *fallible* in its assessment of particular situations. All of the medieval writers thought conscience can err. Their primary struggle was in explaining why it is that we should or should not follow conscience in those cases in which conscience errs. Similarly, Hobbes thought that we can err in our moral judgments of particulars because we can err in reasoning from general principles to particular applications, from definitions to particular instances of those definitions. The notion that conscience must be inerrant simply does not fit with the historical development of conscience.

Though conscience is thought to be possessed by most people, each person only has direct, privileged access to her own conscience. Just as I cannot experience your feelings of pain directly, but must assume them based on your behavior or your situation, I cannot experience your conscience as you experience it. I can of course infer that you are experiencing conscience, say in your behavior or in my conversations with you. But I cannot experience your pangs of conscience; I cannot experience your awareness of what you ought to do the way you are experiencing it. Conscience is a kind of private *consciousness*; it is a phenomenological experience that only the one with that conscience can experience. It is this feature of conscience that gave rise to skepticism about conscience as a source of moral knowledge, e.g. for Hobbes and Freud.[51]

Conscience also makes us aware of what we have done with respect to some moral standard. Kant's account of conscience reduces conscience to making us aware of how we

[51] An example of an account of conscience that allows both an agent and a spectator to have the same access to the agent's conscience is Lynn Stout's definition of conscience as "unselfish prosocial *behavior*" (Stout 2010, 12, emphasis added).There is no privileged access that the agent has to her own conscience, for if her conscience is her behavior, then others have the same access to her behaviors (e.g. by seeing what she does) as she does to her own behaviors. But this account strays far too much from any common conception of conscience, in which conscience is a private experience. Behavior may be one way we can infer the presence and operation of another's conscience, but under no common conception of conscience do we have the same access to another's conscience the way we do have to our own.

have done or what we should do with respect to our own moral standard. But all other accounts present us with conscience as a *consciousness*, either in the form of knowledge or feeling, of what is right and wrong, and whether we are in fact doing what is right and wrong. For instance, according to Mill, our sense of duty makes us aware of what is right to do, and that if we do otherwise, we are violating that standard. For Freud, conscience makes us aware of our shortcomings in abiding by the moral code through the feeling of guilt.

Finally, conscience motivates behaviors. Often, this motivation comes through the feelings of guilt or shame, or feelings of tranquility and self-content. And the motivation can be of varying degrees, sometimes stronger and sometimes weaker. But if one has a conscience, it cannot simply be ignored because it directs us to a particular course of action over another; it motivates because it *prescribes*. Even under an account in which conscience simply provides knowledge, this knowledge is of what we *ought to do* or what we *should have done* in a given situation.

So, to summarize, the core features of conscience that any account of conscience must address are:

[1] Conscience at least *seems* to provide moral knowledge, i.e. knowledge of what is good, right, or proper.
[2] Conscience involves judgments of *particulars*, that some particular feeling, intention, or action is good, right, or proper in some particular instance.
[3] Conscience is *not* infallible; we are capable of doubting our own conscience, and admit that our conscience can be mistaken, at least with regard to particular moral judgments.
[4] Conscience is private in that one has direct access only to one's own conscience.
[5] Conscience is like a judge or witness in that it makes us aware of what we have done with respect to some moral standard.
[6] Conscience motivates certain behaviors, specifically through emotions like guilt and shame, as well as tranquility and self-content.

In the following chapters, I present Smith's account of conscience; one that I will argue plausibly includes all of these features. Smith's account is interesting for several reasons. First, like Butler, he places conscience at the center of his moral theory and at the center of a virtuous person. But Smith provides a more plausible account of the source of conscience's authority and why conscience is crucial for the virtuous person. Second, there were a variety of uses of 'conscience' during Smith's time, and Smith in many ways brings these divergent views together. So prior to C. D. Broad, Smith brought together the cognitive, affective, and conative aspects associated with conscience. But Smith did so in a way that allows for a useful working definition, one that is not overly-inclusive in the way Broad's definition is. Finally, Smith is in many ways returning to the original dual-meaning of 'conscience' as both a private and a shared consciousness. In one sense, conscience is private in that the agent only has direct, privileged access to her own conscience. But in another sense, conscience is shared, not only in that (most) all have a conscience, but also in that our consciences operate similarly, coming to similar moral judgments or arising in response to similar actions in similar situations. It is because of the shared nature of conscience that we are able to assess the quality of another's conscience, even though we do not have direct access to it.

2

From Judging Others to Judging Ourselves: Conscience as Reflective Agent

In everything do to others as you would have them do to you; for this is the law and the prophets.

Matthew 7:12 (NRSV)

Conscience is often characterized as an inner voice that speaks to us. And this speaker, it is claimed, is independent of the agent. It is an angel on our shoulder, speaking on behalf of God; it is Jiminy Cricket; it is the Freudian superego, an internalization of our parents or society. Some interpret Adam Smith along similar lines: conscience (or what Smith calls 'the supposed impartial spectator'[52]) is an internalized prototype of the ideal impartial spectator. This prototype is built up from our experiences of our judging others and

[52] In his writings, Smith does not explicitly state that 'conscience' refers to the supposed impartial spectator. However, he does use 'conscience' and 'supposed impartial spectator' interchangeably (e.g. *TMS* 3.1.3.1 and 3.1.2.32). Some commentators seem to imply some distinction between conscience and the supposed impartial spectator. For instance, at one point, Raphael equates conscience to "the approval and disapproval of oneself", whereas the 'supposed impartial spectator' is "a creation of my imagination" that makes the approval or disapproval (Raphael 2007, 34-35). Charles Griswold makes a similar claim: "Conscience is the *voice* of the impartial spectator addressing the agent and motivates us to do that which is praiseworthy and avoid that which is blameworthy" (Griswold 1999, 194, emphasis added). Also see Broadie (2006, 181) and Forman-Barzilai (2006, 93).

In general, I think these are merely unintended slips. Raphael seems to move between 'supposed impartial spectator' and 'conscience' throughout his presentation of Smith's moral theory (see Raphael 2007, chapter 5). Griswold explicitly equates conscience and the supposed impartial spectator: "Conscience, which is the internalized 'impartial and well-informed spectator'" (Griswold 1999, 133). Similarly, Broadie later equates the impartial spectator and conscience: "the impartial spectator, considered as our conscience" (Broadie 2006, 182). So these slips may be merely a reflection of the wide variance in the use of 'conscience'. But it is particularly important to be consistent in its use when discussing Smith's moral theory, which primarily uses 'conscience' in a narrower sense.

of others judging us from an impartial stance.[53] Under such an account, conscience's judgments may be our own in that they are products of our psychology, but they are *not* our own in that they are formed from principles that we do not necessarily endorse.

In this chapter, I argue that such an interpretation of Smith is mistaken. For Smith, an agent's conscience just is that agent when she is judging herself by placing herself into the situation of an impartial spectator. We judge ourselves using the same mechanism and principles we use to judge others. And so conscience gives us the flip side of the Golden Rule: judge *yourself* as you judge *others*. Our judgments of others and of ourselves are thusly intimately bound. The judgments of conscience are not the judgments of some independent prototype, but rather *our* judgments as the agents that author the judgment.

My interpretation is not wholly novel. Among others, D. D. Raphael (2007) and Jon Rick (2007) provide similar interpretations. But I want to highlight two elements of Smith's theory. For one, there is a close connection between us as conscience and us as spectators of others. This connection is lost in interpretations that treat conscience as a prototype that operates on principles distinct from those we employ as spectators generally. And second, conscience is an operation of the imagination, distinct from reason. This point is crucial for properly understanding how Smith distinguishes moral judgments from other kinds of judgments, and for recognizing the unique place of Smith's theory in the history of conscience.

[53] For example, see Broadie (2006), Campbell (1971), Darwall (1999), Evensky (1987), Griswold (1999), Hanley (2009), Heilbroner (1982), and Kennedy (2008).

2.1 The Sympathetic Mechanism

To understand how we judge others and ourselves, we first need the proper account of how we come to sympathize with another, for it is through sympathizing (or not) that we, as spectators, judge others. The process is as follows:

> By the imagination we [the spectator] place ourselves in [the agent's] situation, we conceive ourselves enduring all the same torments, we enter as it were into his body, and become in some measure the same person with him, and thence form some idea of his sensations, and even feel something which, though weaker in degree, is not altogether unlike them (*TMS* 1.1.1.2).

We observe a situation involving some agent that is to be judged, and by imagination, enter into the agent's situation to imagine what we would feel in the same situation. The result of this imaginative exercise is sympathy, a fellow-feeling or agreement with any sentiment between the spectator and the agent (*TMS* 1.1.1.5). I will call this mechanism by which we come to sympathize the 'sympathetic mechanism', which is not to be confused with the sympathy that results from the operation of this mechanism.

What is crucial to Smith's account of how we sympathize is that in entering into the agent's situation, in imagining what it is like 'to be the agent', so to speak, we are *not* taking on any features particular to the agent herself. Rather, we only consider the *situation* of the agent, and from that situation consider how *we* would think and feel, not how the *agent* would. Jon Rick (2007) provides a helpful distinction between two ways in which spectators "get into another person's shoes". One way in which the spectator enters into another's situation, what Rick calls 'spectator-partiality', is for the spectator to project *herself* into the situation of the agent. The spectator "loses no sense of herself and maintains all of her particular values, commitments, and affective dispositions". Moreover, the spectator

"acquires no sense of the agent's practical identity" (Rick 2007, 148). In other words, the spectator feels as the *spectator* generally would in that situation, given the spectator's dispositions. Alternatively, the spectator can enter into the situation of the agent and consider how the *agent* would feel, given the *agent's* dispositions, what Rick calls 'agent-partiality'. In this case, the spectator not only enters into the situation of the agent, but also takes on the agent's "practical identity and psychological economy" (Rick 2007, 148). Smith, as both Rick and I interpret him, endorses the 'spectator-partiality' account of how we sympathize with others. When we engage in this imaginative exercise, we hold steady the situational factors of the agent, and displace the agent with ourselves, the spectators. So what Smith means when stating "[we] become in some measure the same person with him" is that we imagine ourselves in the same *situation* as the agent. Hence, sympathy need not arise in response to the agent's actual sentiments, but rather her situation (*TMS* 1.1.1.10).[54]

To be clear, when we as spectators project ourselves into the situation of the agent, we are imagining ourselves *as if* we were in that situation, running a first-person simulation to come to some conclusion about what we would feel. We are *not* determining what we would feel by forming generalizations about how we feel in certain kinds of situation and deducing what we would feel. Also, we are *not* determining what we would feel by imagining how someone "like us", someone with our psychological profile, would feel in that situation; in other words, we are not simply viewing some image or representative of us as we do when passively viewing ourselves in a video. Rather, we imagine ourselves as *living through* that situation, as participants in that situation.

---

[54] I am using Rick's terms here, but his distinction between agent-partiality and spectator-partiality is not novel, particularly in discussions of mindreading or simulation theories of mind, which interestingly attribute their origin to Adam Smith's *TMS*. Robert Gordon, for instance, distinguishes between "just imagining being in X's situation" and "imagin[ing] being X in X's situation" (Gordon 1995, 741-742; also see Gordon 1986, Goldman 1989 and 2008, Stueber 2006 and 2009, and Oxley 2011).

The sympathetic mechanism gives us awareness of what *we*, as spectators, would feel if we were in the same situation as the agent. But Smith claims that it does more:

> As we have no immediate experience of what other men feel, we can form no idea of the manner in which they are affected, but by conceiving what we ourselves should feel in the like situation (*TMS* 1.1.1.1).

The sympathetic mechanism is supposed to give us some idea of what the *agent* feels as well. But how can this be, given that we are imaginatively entering into the situation of the agent to determine how *we*, and not the agent, would feel? One possibility is that the sympathetic mechanism does not, in fact, give us any understanding of what the agent feels. Instead, we infer what the agent feels by observing their facial, bodily, and verbal expressions. Smith seems to recognize such means as not only a way to know what an agent feels but also a way to sympathize, for by observing another's expressions of joy and pain, we as spectators come to feel something similar:

> The passions, upon some occasions, may seem to be transfused from one man to another, instantaneously, and antecedent to any knowledge of what excited them in the person principally concerned. Grief and joy, for example, strongly expressed in the look and gestures of any one, at once affect the spectator with some degree of a like painful or agreeable emotion. A smiling face is, to every body that sees it, a cheerful object; as a sorrowful countenance, on the other hand, is a melancholy one (*TMS* 1.1.1.6).[55]

So Smith recognizes that we might simply feel a corresponding emotion upon seeing someone visibly in distress or in joy, without any knowledge of the situation of the agent.

[55] Smith is referring to Hume's account of sympathy, first presented in *THN* 2.1.11.3: "When any affection is infus'd by sympathy, it is at first known only by its effects, and by those external signs in the countenance and conversation, which convey an idea of it. This idea is presently converted into an impression, and acquires such a degree of force and vivacity, as to become the very passion itself, and produce an equal emotion, as any original affection."

However, it is clear that Smith thinks that sympathizing by inferring from various expressions is an inadequate account of how we understand and judge others. That instantaneous fellow-feeling from observing the agent's expressions of pain is, for Smith, quite limited in scope. Only certain kinds of feelings seem to excite a fellow-feeling simply from observation, namely feelings whose effects are restricted to the agent and no one else, like grief and joy in response to the agent's bad or good fortune. But other kinds, like anger towards another person, we generally do not sympathize with unless we know *why* that person feels the way she does (*TMS* 1.1.1.7) because someone else is affected by that anger, i.e. the person to which the anger is directed (*TMS* 1.1.1.8). Also, even with feelings like grief and joy, we only have an "extremely imperfect" sympathy. Until we know more about the details of the situation, we can only feel "general lamentation", and not grief to the same degree over the particular misfortune (*TMS* 1.1.1.9). So Smith acknowledges that we can and do infer what another feels by observing various expressions. However, he thinks that such inference leaves us with an incomplete understanding of the agent's sentiments and an incomplete assessment of the propriety of those sentiments (*TMS* 1.1.1.10).[56]

Smith, instead, seems to rely heavily on the situation to accurately identify the agent's sentiments. For instance, we do not need to see any expression of fear to conclude that the agent feels fear. It is enough that we "see a stroke aimed and just ready to fall upon the leg or arm of another person" (*TMS* 1.1.1.3). We do not need to observe a person with "sores and ulcers" scratch themselves to know that they would feel itchy. It is enough that we see those

[56] For instance if we do sympathize by in this way, we only do so because the agent's emotions "suggest to us the general idea of some good or bad fortune that has befallen the person", and we are able to enter into that situation through the sympathetic mechanism (*TMS* 1.1.1.8). And even then, since we are not fully informed of the cause of that person's emotions, the feelings of the agent "create rather a curiosity to inquire into his *situation*, along with some disposition to sympathize with him, than any actual sympathy that is very sensible" (*TMS* 1.1.1.9, emphasis added).

sores and ulcers, to see what situation that person is in (*TMS* 1.1.1.3). This reliance on the situation to understand the agent's sentiments suggests the following: we assume that the agent shares some basic human faculties and principles, which give rise to certain sentiments to a certain degree in certain situations. Because you and I share these faculties, I can understand what you feel by simply considering what I would feel. However, this assumption is often how we make our initial assessment of the agent's sentiments. We can adjust that assessment based on observation of physical expressions, or some indication that the agent has lost some faculty (e.g. *TMS* 1.1.1.11 and 1.1.1.13). However, Smith seems to think that we often do not and need not rely on inferences from physical expressions to understand how an agent feels.

So far, I have stressed that for Smith, we sympathize by holding the situation of the agent fixed, and considering what *we* as spectators would feel if we were in that situation. We have yet to address what it is that we are holding fixed as part of the situation. Smith provides a couple of examples that might help clear up what he understands to be a part of the situation. One example is of a person who has lost the capacity to reason: "the poor wretch, who is in it, laughs and sings perhaps, and is altogether insensible of his own misery" (*TMS* 1.1.1.11). In sympathizing with this person, we place ourselves into the situation which does not include the complete insensibility that this person has of his own condition. In other words, just as we do not consider the agent's feelings as part of the situation, we do not include the agent's obliviousness to her situation.[57] Similarly, we do not consider a dead person's lack of any mental activity as part of the situation. In sympathizing with the dead,

---

[57] A similar conclusion can be drawn from Smith's example of an imprudent and rude person who is unaware of the impropriety of his behavior. In such a case, we do not include as part of the situation his ignorance, and we feel embarrassed for him because we know that we would feel embarrassed were we in his situation (*TMS* 1.1.1.10).

we hold as part of the situation being "deprived of the light of the sun; to be shut out from life and conversation; to be laid in the cold grave, a prey to corruption and the reptiles of the earth; to be no more thought of in this world, but to be obliterated, in a little time, from the affections, and almost from the memory of their dearest friends and relatives" (*TMS* 1.1.1.13). So the situation includes not only the physical state of the corpse, but also the lack of thoughts and feelings others have about the agent.[58] What the situation does not include is the agent's ability to feel and reason. If these capacities were part of the agent's situation, the spectator would not feel *any* feeling since the situations would then strip the spectator of the capacity to feel (also see the example in *TMS* 1.1.1.11).

2.2 How We Judge Others: Propriety/Impropriety and Merit/Demerit

The sympathetic mechanism is used to make two kinds of judgments: 1) judgments of propriety and impropriety, and 2) judgments of merit and demerit.[59] Though the two are different in kind, the latter is built up from the former. So it will be appropriate to begin with an account of judgments of propriety.

What exactly are judgments of propriety? It is the approving or disapproving of another's sentiments in a given situation by considering what we would feel in the same situation, and recognizing the correspondence of sentiments. To form this judgment, we need to use the sympathetic mechanism to determine what we would feel in the same situation as the agent. We engage in spectator-partiality, in which we enter into the agent's situation to determine how we, the spectator, would feel. We also need to determine what the agent in

[58] Several others have remarked on how 'situation', for Smith, includes more than just physical elements, extending to certain psychological features as well (e.g. Nanay 2010: 91-92; Darwall 1999: 141-142; Fleischacker 2006: 4).
[59] For simplicity, I will from time to time use 'propriety' to refer to both propriety and impropriety, and 'merit' to refer to both merit and demerit. The context should disambiguate the two usages.

fact feels in that situation. We might infer that the agent simply feels as we would in the same situation. But if that were all that we used to determine what the agent feels, then we would never be any disagreement with the agent. Instead, we also make inferences using some combination of cues, including facial expressions, verbal cues, and body language. The strength of the cries of a child indicates how much pain the child is in. The flare in the nostrils, the tenseness of the muscles, and verbal assaults indicate the height of the agent's anger. We might also engage in some agent-partiality, in which we as spectators enter into the agent's situation *as the agent*, embodying the agent's particular dispositions as we understand them given our experiences with that agent.

The spectator then compares the sentiments of the agent with the spectator's own upon entering into the agent's situation. If the two sentiments are similar, then the spectator approves of the agent to the degree that the two are similar; if the two are dissimilar, then the spectator disapproves to the degree that they differ. Judgment of propriety is *not* the correspondence of sentiments itself, but the *observing* of this correspondence (*TMS* 1.1.3.1), i.e. a "perception of this coincidence" (*TMS* 1.1.3.3). And so if you feel what I would feel in the same situation, but I fail to recognize this agreement, I do not make a judgment of propriety. The key here is that we do not need to actually feel the corresponding feeling. We only need to observe that we *would* generally feel the same, though at the moment. Hence, Smith says the judgment is dependent not on the observation of an actual agreement in sentiment, but rather the "consciousness of this *conditional* sympathy" (*TMS* 1.1.3.4, emphasis added).

Though the judgment of whether or not the agent has felt and acted properly depends on this correspondence in sentiments, the standard of propriety is always determined by the

spectator. The spectator does not judge by deferring to others, but always takes her reference point to be her own faculties and principles, never those of others:

> Every faculty in one man is the measure by which he judges of the like faculty in another. I judge of your sight by my sight, and of your ear by my ear, of your reason by my reason, of your resentment by my resentment, of your love by my love. I neither have, nor can have, any other way of judging about them (*TMS* 1.1.3.10).

In other words, when I judge an agent, I do not consider the situation nor do I consider the judgment of any other spectators to make *my* judgment. Thus, my judgment of propriety is not formed by considering what the society approves or what spectators generally feel. My judgments can be *informed* by such considerations. For instance, if a person faces disagreement with another, he might attempt to reassess his own judgment by entering more fully into the situation of the man with whom he disagrees, "to bring home to himself every little circumstance of distress which can possibly occur to the sufferer" (*TMS* 1.1.4.6), so that he might account for situational factors that he initially did not sufficiently account for.[60] Also, I may come to recognize that I am not my usual self, that my faculties and principles are not operating the way they normally do: "We may often approve of a jest, and think the laughter of the company quite just and proper, though we ourselves do not laugh, because, perhaps, we are in a grave humour, or happen to have our attention engaged with other objects" (*TMS* 1.1.3.3). So I may adjust my judgment accordingly, based on my knowledge that I would judge differently were I in a more normal situation or one more conducive to reacting as others normally react in the situation facing the agent under review. But in all such cases, I am not judging what is proper based on how others judge, but rather, I take their judgments to give me grounds for reconsidering how *I* would judge were my reactions not

[60] Broadie (2006, 175-176).

affected by factors peculiar to my mood or circumstance. I revise my understanding of the situation, or take into account peculiarities in the operations of my own faculties and principles, and form a new judgment that now takes into account these considerations. Judgments of what is proper are always determined by *my* faculties and principles.

It is important to highlight that judgments of propriety involve the effects of an agent's actions only in a very narrow way. The judgment of propriety is concerned primarily with an agent's *response* to some object, whether it is human or not. In entering into the situation of the agent, we are only considering how we would *respond* in the same situation. Hence, the effects *on* the agent may be considered, as part of the situation, e.g. the pain inflicted on the agent. Any consideration of the effects *of* the agent's motives or actions on others is not involved in judgments of propriety. Consider a situation in which my friend hits a stranger, breaking the stranger's nose. When I judge the propriety of my friend's motive and action, I am not concerned with the fact that the stranger was affected by my friend's actions. Rather, I am concerned with whether or not I would have felt and done the same as my friend in the same situation. The same holds if it turns out that my friend hit *me* and not a stranger. I am clearly affected by my friend's actions. But these factors are not part of the judgment of propriety. All that is considered in judgments of propriety is whether or not I would have felt the same. Because I am affected by it, I might be more motivated to come to an agreement in sentiment. But in such a case, the effects influence my motivation for coming to an agreement, for trying harder to enter into my friend's situation and taking into account all the details of his situation. What the effects do *not* do is factor into the sympathetic process through which I come to judge my friend's motives proper.[61] As we

[61] That Smith limits the role of effects of the motives and actions of the agent to the *motivation* of those affected is further evidenced by his discussion of the two kinds of situations in which we judge propriety. In the first, the

shall see, that judgments of propriety do not involve consideration of effects of the agent's motives on others is important for distinguishing judgments of propriety from judgments of merit.

In Part II, Smith presents a different kind of judgment: merit and demerit. And to give an account of this kind of judgment, he leaves aside the dyadic relationship between the spectator and the one judged in Part I, and instead shifts to a triadic relationship in which the spectator approves or disapproves of the person judged (agent) given the *effects* on another (recipient).[62] In this triadic relationship, the spectator imagines being in the situation of *both* the agent and the recipient. But the spectator does so in a particular order. The spectator *first* enters into the situation of the agent and determines whether the agent's motives are proper. The spectator then enters into the situation of the recipient to figure out how the spectator would feel in the same situation. If the spectator judges the agent's motives proper and feels gratitude upon entering into the recipient's situation, then the spectator judges the agent meritorious, i.e. deserving of reward. If the spectator judges the agents' motives improper and feels resentment upon entering into the recipient's situation, then the spectator judges the agent demeritorious, i.e. deserving of punishment. And if the spectator does not enter into the situation of either the agent or the recipient, the spectator does not make a judgment of merit. Merely entering into the situation of the recipient and feeling gratitude or resentment is not enough to come to a judgment of merit or demerit, respectively. In fact, if the spectator does

---

agent's sentiments do not affect the agent or us. And in such cases, we are not very motivated to come to any sort of agreement in sentiment (*TMS* 1.1.4.1). But in the second, the agent's sentiments do affect either the agent or us, and so we are more motivated to enter more fully into the situation of the agent to come to more of an agreement in sentiment (*TMS* 1.1.4.1). In neither case are the effects of the agent's motives and actions part of the judgment of propriety.

[62] I am adopting Broadie's terminology, 'agent' and 'recipient' (Broadie 2007, 178).

not first approve of the agent's motives, the spectator cannot deem the recipient's gratitude proper, and if the spectator does not disapprove of the agent's motives, then the spectator cannot deem the recipient's resentment proper (see *TMS* 2.1.3.2-3). Interestingly, in this imaginative exercise, the spectator need not assess the propriety of the recipient, though the spectator can do so. What is important is not the recipient's actual feeling, but rather what the spectator feels upon placing himself in the situation of the recipient.[63]

The key difference between judgments of merit and those of propriety is that judgments of merit do not focus on the response of the agent, but rather on the response of the *recipient*, since we are assessing the effects of the agent's motives on other people, i.e. the recipient. Judgments of propriety simply assess whether the agent's response is of the appropriate kind and degree given "the cause or object which excites it" (*TMS* 2.1.1.2). The judgment is strictly in response to the agent's sentiments or actions. And depending on whether we, as spectators, sympathize with those sentiments or actions, we either deem the agent deserving of approval or disapproval. Judgments of merit, however, are not grounded in considerations of the agent's response to the situation. Rather, they are grounded in our considerations of the *effects* of the agent's motives on other people. In making judgments of merit, what we are really doing is finding the motive of the agent improper and as a consequence attempting to change the effects resulting from those improper motives through the use of rewards and punishments. Hence, judgments of merit and demerit are judgments of the deservedness of *reward* or *punishment*, not approval or disapproval. So although

[63] *TMS* 2.1.5.11. This point is highlighted in Broadie (2007, 179). If the spectator is attempting to determine the propriety of the recipient, then the spectator would need to sympathize with the recipient to determine the recipient's response proper. But in such a case, it is not the agent that is judged, but rather the recipient.

judgments of merit make use of judgments of propriety, they are distinct in substantial ways from the judgments of propriety that we use to make them.[64]

Furthermore, judgments of merit and demerit, unlike those of propriety and impropriety, are grounded in *specific* sentiments, namely gratitude and resentment, respectively. Recall that to make judgments of propriety we simply need to employ our capacity for fellow-feeling once we have executed the imaginative exercise of entering into an agent's situation. Exactly what the agent felt does not play a determinative role in determining the propriety of the reaction under review. However, for judgments of merit and demerit, the spectator must enter into the situation of the recipient and determine that the spectator would feel gratitude and resentment, respectively.[65] Other sentiments will not give rise to judgments of merit and demerit because other sentiments do not prompt a response of *deservedness*, a response of *obligation* or *owing* another, notions which are central to reward and punishment. As Smith writes: "To reward, is to recompense, to remunerate, to return good for good received. To punish, too, is to recompense, to remunerate, though in a different manner; it is to return evil for evil that has been done" (*TMS* 2.1.1.4). Punishment and reward is only satisfied when the recipient gives the agent what the agent deserves. Though there are other sentiments that might want to prompt us to give to or receive from the agent, "there are none [besides gratitude and resentment] which so directly excite us to be the *instrument* of either" (*TMS* 2.1.1.5, emphasis added). Only gratitude and resentment satisfies

[64] Haakonssen makes the slip of collapsing the two kinds of judgments into one in his introduction to the Cambridge Press edition of *TMS* (2002, xvi): "The central concept is that of propriety. We first of all judge an agent, whether oneself or another, in terms of whether a motive or action is proper to the given situation of that agent; if it is, we say that the motive or action has moral merit, otherwise demerit".

[65] I am leaving open the possibility that the spectator makes a judgment of merit and demerit without feeling gratitude and resentment, but instead judges that she *would* feel gratitude and resentment under normal circumstances.

the proper exchange between the recipient and the agent, in which the recipient is the one that is giving what the agent deserves.[66]

Exactly which sentiments might we mistake to produce judgments of merit and demerit? Smith distinguishes gratitude from "[t]he love and esteem which grow upon acquaintance and habitual approbation" (TMS 2.1.1.5), and resentment from "[t]he hatred and dislike… which grow upon habitual disapprobation" (TMS 2.1.1.6). The distinction is meant to show that gratitude and resentment, unlike these other sentiments, are ones that not only prompt us to want benefit and harm, respectively, to come upon the agent, but also that we are the source of that benefit and harm. For instance, in TMS 2.1.1.5, Smith points out that like gratitude, the "love and esteem which grow upon acquaintance and habitual approbation, necessarily lead us to be pleased with the good fortune of the man who is the object of such agreeable emotions, and consequently, to be willing to lend a hand to promote it". But unlike gratitude which gives rise to a desire to thank or in some way repay the person to whom we are grateful, love and esteem do not demand that we be the *source* of the good fortune that falls on the object of our love and esteem. The same distinction holds true between resentment and hatred that arises from habit. Unlike resentment which gives rise to a desire to blame, scorn or in some way punish the person we resent, hatred does not demand that we be the *source* of the bad fortune that falls on the object of our hatred.[67]

[66] There is an issue here as to whether one can sympathize with another's resentment only within a triadic relationship, or whether one can do so within a dyadic relationship as well. There are passages in Part I that seem to indicate that one can directly sympathize with another's resentment without considering any third party. However, these passages can be read as a rhetorical tool, which Smith often makes use of, which is to begin with a more loose discussion and then to refine his concepts as he delves into the topic more deeply. Of course, Smith might simply think that one can sympathize with another's resentment in both the dyadic and triadic relationship, and that only the latter is a judgment of demerit. I do not think it makes much of a difference for my discussion. All that is needed is that judgments of demerit rely on a particular kind of resentment, one that must be assessed in the context of a triadic relationship.

[67] I have discussed two means by which we form moral sentiments: "First, we sympathize with the motives of the agent; secondly, we enter into the gratitude of those who receive the benefits of his actions" (*TMS* 7.3.3.16).

2.3 How We Judge Ourselves: The 'Supposed Impartial Spectator'

Smith begins Part III of *TMS* by summarizing what has done thus far, which is consider the "origin and foundation of our judgments concerning the sentiments and conducts of *others*" (TMS 3.1.1.1, emphasis added). He then turns to the main aim of Part III: consider the origin and foundation of our judgments concerning *our own* sentiments and conducts. Unfortunately, we are incapable of judging ourselves directly, for judging requires some comparing between the spectator's sentiments and the agent's. Hence, we need to imaginatively remove ourselves from our own situation (as agent), and view ourselves from the situation of a spectator:

> We can never survey our own sentiments and motives, we can never form any judgment concerning them; unless we remove ourselves, as it were, from our own natural station, and endeavour to view them as at a certain distance from us. But we can do this in no other way than by endeavouring to view them with the eyes of other people, or as other people are likely to view them. Whatever judgment we can form concerning them, accordingly, must always bear some secret reference, either to what are, or to what, upon a certain condition, would be, or to what, we imagine, ought to be the judgment of others. We endeavour to examine our own conduct as we imagine any other fair and impartial spectator would examine it. If, upon placing ourselves in his situation, we thoroughly enter into all the passions and motives which influenced it, we approve of it, by sympathy with the approbation of this supposed equitable judge. If otherwise, we enter into his disapprobation, and condemn it. (TMS 3.1.1.2)

There are two more principles that give rise to moral sentiments: "thirdly, we observe that his conduct has been agreeable to the general rules by which those two sympathies generally act; and, last of all, when we consider such actions as making a part of a system of behavior which tends to promote the happiness either of the individual or of the society, they appear to derive a beauty from this utility, not unlike that which we ascribe to any well-contrived machine" (*TMS* 7.3.3.16). The third principle is that with experience, we begin to form general rules that correspond to how spectators, ourselves included, judge the motives and actions of others. Once we form these general rules, we begin to look to these rules with reverence, "what is properly called a sense of duty" (*TMS* 3.1.5.1). I will discuss this principle in greater detail in the next chapter. The fourth principle, which Smith discusses in detail in Part IV, is that acting from this sense of duty is approved if a spectator does or would feel the same sentiment in the same situation. In a like manner, though we find some action useful, when we approve of such an action, we do not in the first instance approve of it because a spectator appreciates its utility, but rather because a spectator sees the action as beautiful, or pleasurable. Though in some sense both the third and fourth principle have their origin in the sympathetic mechanism, neither requires the use of the sympathetic mechanism at the moment at which we execute a moral judgment.

We act as a spectator to ourselves by imagining how we would judge ourselves were we an impartial person observing our actions in full knowledge of our motives and the situation to which we are reacting in formulating those motives. If I imaginatively enter into this different situation or at least think of things from a more impartial perspective, I will find myself with two sentiments, my original sentiments as agent, and the less violent sentiments I will experience after thinking of my motives and actions as a spectator would.[68] When these two sentiments coincide, then I approve of myself, and when these two sentiments do not, I disapprove of myself. So, as Alexander Broadie puts it, though conscience is "the product of an act of imagination… [i]n so far as the agent is imagining how he would be judged by an impartial spectator, the judgments he is imagining are his own" (Broadie 2006, 181).

We need to be careful about how we understand this process. The passage above states that in entering into the situation of a spectator, we are "endeavouring to view them with the eyes of other people, or as other people are likely to view them," "as we imagine any other fair and impartial spectator would examine it." It seems as if Smith is claiming that we are imagining how *other* people would judge us. However, this is not the case. Though we change situations in thought or imagination and view ourselves from the situation of an impartial spectator, we continue to exercise our own faculties to judge as *we* would judge, not exercise our faculties to judge as *others* would judge:

> When I endeavour to examine my own conduct, when I endeavour to pass sentence upon it, and either to approve or condemn it, it is evident that, in all such cases, I divide myself, as it were, into two persons; and that I, the examiner and judge, represent a different character from that other I, the person whose conduct is examined into and judged of. The first is the spectator, whose sentiments with regard to my own conduct I endeavour to enter into, by placing myself in his situation, and

[68] The latter sentiment differs not only in degree but also in kind from the original sentiment since the latter is a sentiment arising from the imagination.

> by considering how it would appear to me, when seen from that particular point of view. The second is the agent, the person whom I properly call myself, and of whose conduct, under the character of a spectator, I was endeavouring to form some opinion. (TMS 3.1.1.6)

Even though in judging ourselves we enter a different situation, that of an impartial spectator, *I* am still the judge entering into the situation of the agent and exercising *my* faculties to judge the agent: "by considering how it would appear to *me*." What Smith means, then, in referring to viewing ourselves "with the eyes of other people, or as other people are likely to view them" is that because we have no "immediate experience" of what another feels, we are limited to imagining how *we* would feel from the same situation: "We begin, upon this account, to examine our own passions and conduct, and to consider how these must appear to them, by considering how they would appear to us if in their situation" (*TMS* 3.1.1.5).[69]

Finally, conscience is necessarily a judge of oneself.[70] We might be able to imaginatively enter into the situation of an impartial spectator to form moral judgments of others, but such a spectator would *not* be conscience. Smith is not explicit on this point, but he only discusses conscience in the context of self-approbation and self-disapprobation and the view accords with common sense or ordinary usage of the term. 'Conscience' is initially introduced to explain how we judge ourselves (*TMS* 3.1.1). Even afterwards, conscience is only discussed as a means for correcting the sentiments of and cultivating virtues in the bearer of that conscience, not anyone else (*TMS* 3.1.2-3). Smith provides other means for correcting our judgments of others: we enter more fully into the agent's situation by taking into account factors we did not as a result of our partiality or we take into account the

---

[69] Cf. *TMS* 1.1.1.1.

[70] The exception is Carrasco, who claims that any moral sentiment, whether of oneself and of others, must appeal to the supposed impartial spectator (Carrasco 2004).

perspective of other spectators to identify our own errors as spectators (*TMS* 1.1.4.6). So even though it is possible that we judge others by imaginatively entering into the situation of an impartial spectator, Smith would not consider this imagined judge to be conscience.

2.4 Alternative Interpretation: Conscience as Prototype?

There is a common alternative interpretation of Smith's account of conscience: conscience is a prototype of an idealized impartial spectator. The general thought is that through experience, we build up an idea of how impartial people tend to evaluate different situations, or an idea of how impartial people would judge the actions and motives of another were they privy to all the relevant information about the situation of the person under review. Once we have formed this prototype in our mind, we use it as the standard by which we judge ourselves. Conscience, then, is our judging ourselves as we imagine how an ideal judge, who embodies all impartial judges, would think of us, based on what we know of how impartial people judge those they judge.

The key to this interpretation is that conscience, "the supposed impartial spectator", is *supposed* (or imagined) and *impartial* in a very particular way. Conscience is not an actual (living, breathing) spectator, the way that Jiminy Cricket, for instance, is an actual spectator. Yet conscience is *distinct from* us as agent. It forms moral judgments on principles different from ours, for it represents a different spectator, an ideal spectator. Conscience is 'imagined' in that we are thus *creating* a spectator using our imagination, which I refer to as imagined in the [S-Other] sense:

> [S-Other] Conscience is 'imagined' in that the agent considers what another spectator would judge, using the spectator's own faculties and principles.

Conscience, then, is not imagined in that it is a reflection of the agent's faculties and principles. Rather, what we are imagining is a spectator that forms moral judgments in a manner different than our own.

Furthermore, what we are imagining is not just any spectator, but rather an *impartial* one. Impartiality is an essential feature of conscience. If we are imagining how a partial spectator, like our parents or friends, would judge, then we are not appealing to conscience at all, because conscience *must be* an impartial spectator. And because conscience is an ideal, and part of this ideal is complete impartiality, some have gone further to claim that conscience must be a *fully* impartial spectator.

Several prominent commentators adopt this interpretation. For instance, T. D. Campbell claimed that conscience was an internalization of outside spectators that we encounter:

> [Conscience] does not therefore embody the reactions of any particular person, but is an empirical ideal type who may be said to represent all those aspects of human nature which are brought into play in the development of moral judgments. The spectator is "ideal" in the sense that he excludes all those features of actual spectators which relate to their special interests as particular individuals involved in the actual situation which they are observing; he is "empirical" in the sense that, once this abstraction is made, the responses of the spectator are identified with the consensus towards which any actual group of persons can be observed to approximate in their attitudes to the behaviour of their fellows (Campbell 1971, 127).

Conscience, then, does not reflect any particular spectator or impartial spectator. Rather, it is a being that represents an ideal in the sense that it reflects generalizations we believe to hold for impartial spectators, devoid of any particular details of particular spectators. The agent herself, of course, is *not* such a being. In other words, in imagining conscience we are

imagining a being wholly different from us as agent. Conscience is impartial to the degree that we often are not. It also judges as impartial spectators generally do, which we do not always coincide with how we judge. We are, in essence, imagining a being distinct from us, one that operates differently using different faculties and principles, and ultimately coming to moral judgments different from our own.

Similarly, Stephen Darwall claims: "Moral judgments involve an impartial projection into the agent's or patient's standpoint. We imaginatively project, not *as ourselves*, but impartially, *as any one of us*" (Darwall 1999, 142). For Darwall, all moral judgments are made by projecting an impartial self, and that the appeal to impartiality is a feature embedded into the sympathetic mechanism. Whether we are judging others or ourselves, we are not projecting ourselves, but rather some prototype, an impartial spectator that embodies how "*any one of us*" would judge. In some sense, we are this prototype, for we are doing the projecting and imagining from the agent's situation. But strictly speaking we are *not* this prototype, for this prototype does not reflect our own faculties and principles, but rather those of any impartial spectator.

Charles Griswold also presents conscience as a prototype that embodies generalizations a person embraces about how spectators judge when judging impartially:

> Moral self-consciousness requires that I "divide myself, as it were, in two persons" (*TMS* 3.1.1.6). The *internalized* or *idealized* judge is still a spectator. The imagination preserves the privileged position of this spectator – the stand-in for "the public." The theatrical relation is thus internalized; *we become our own public* (Griswold 1999, 108).

When we judge ourselves, we are appealing to a judge that reflects the community as a whole, i.e. "mankind". We are still judging ourselves, in that we are dividing ourselves and

entering into the situation of a spectator. But in entering into the situation of this spectator, we are judging as the *public* would. Conscience, then, does not apply *our* principles, which may not be ideal in any way, but rather apply the principles of the general community of impartial spectators, "the stand-in for 'the public'."

There are several reasons for interpreting conscience as a prototype. For one, Smith often speaks of how we are attempting to judge ourselves as *any other* impartial spectator would judge. So in judging ourselves, we make reference to how *others* judge, not to how *we* judge:

> We must, here as in all other cases, view ourselves not so much according to that light in which we may naturally appear to *others*…. When he views himself in the light in which he is conscious that *others* will view him, he sees that to them he is but one of the multitude in no respect better than any other in it. If he would act so as that the impartial spectator may enter into the principles of his conduct, which is what of all things he has the greatest desire to do, he must, upon this, as upon all other occasions, humble the arrogance of his self-love, and bring it down to something which *other* men can go along with (*TMS* 2.2.2.1, emphases added).

In judging ourselves, we are considering how we appear to *other* spectators, and whether *they* would approve (or disapprove) of us in accordance with *those* judges' sentiments, not our own. He repeats this notion that we are judging ourselves as *others* would judge us elsewhere in Part III:

> We can never survey our own sentiments and motive, we can never form any judgment concerning them; unless we remove ourselves, as it were, from our own natural station, and endeavour to view them as at a certain distance from us. But we can do this in no other way than by endeavouring to view them *with the eyes of other people*, or as *other people are likely to view them*. Whatever judgments we can form concerning them, accordingly, must always bear some *secret reference*, either to what are, or to what, upon a certain condition, would be, or to what, we imagine, ought to be the *judgment of others* (*TMS* 3.1.1.2, emphases added; also see *TMS* 3.1.1.5).

According to Fonna Forman-Barzilai, Smith in this passage is claiming that we are not simply entering into the situation of others to "[step] outside ourselves and [cool] ourselves off" (Forman-Barzilai 2010, 90). Rather, Smith's claim about the "secret reference" to the judgments of others suggest that we are entering into the situation of another to judge ourselves as *they* would judge us, using *their* principles of action. As Darwall puts it: "We imaginatively project, not *as ourselves*, but impartially, *as any one of us*" (Darwall 1999, 142). The implication is that though we are, as conscience, judging ourselves in some sense, we are doing so as *other* (impartial) spectators would judge, not as *we* would judge. Another reason for the prototype interpretation is that Smith refers to conscience as an 'abstract' being, a 'representative' of impartial spectators:

> The man within the breast, the *abstract* and ideal spectator of our sentiments and conduct, requires often to be awakened and put in mind of his duty, by the presence of the real spectator… (*TMS* 3.1.3.38, emphasis added).

[T]he prudent man is always both supported and rewarded by the entire approbation of the impartial spectator, and of *the representative of the impartial spectator*, the man within the breast. (*TMS* 4.1.1.11, emphases added).[71]

Conscience, thus, is not a matter of taking the perspective of any particular impartial spectator, including the agent herself; rather, it is adopting an abstract perspective, distinct from the perspective of any real spectator. The same is true of Smith's reference to 'the

[71] Smith makes several more references to both the 'abstact' spectator and the 'representative' in prior editions of *TMS* 3.1.2.32-37, e.g."This inmate of the breast, this abstract man, the representative of mankind, and substitute of the Deity, whom nature has constituted the supreme judge of all their actions, is seldom appealed to by them" (*TMS* 3.1.2.37).

representative of the impartial spectator.' 'The impartial spectator' (or other times, 'the indifferent spectator') is not some particular impartial spectator. Smith uses this term when providing tendencies, generalizations, of impartial spectators.[72] So in referring to the 'representative' of the impartial spectator, Smith is appealing to some abstract spectator who embodies all the generalizations the agent endorses about impartial spectators that he or she has built up from her experiences of being judged by impartial people and of judging others impartially.

So then my conscience is a reflection of what I as agent believe impartial spectators *generally* to be, a prototypical impartial spectator built up from my experiences with impartial spectators. This prototype is the 'ideal' judge within the breast (e.g. *TMS* 3.1.3.27-29, 3.1.3.38, 3.1.4.4). In part, it is ideal in that it is an impartial spectator, for as Smith writes, when we align ourselves with this ideal judge, we become "the *impartial* spectator of his own conduct" (*TMS* 3.1.3.28, emphasis added). Moreover, some suggest that conscience is also ideal in that it informed of all the details of the situation and is perfectly impartial so as to render the correct moral judgment. This ideal spectator is in stark contrast to how we

---

[72] For instance, in *TMS* 2.2.2.1, he writes: "To disturb his happiness merely because it stands in the way of our own, to take from him what is of real use to him merely because it may be of equal or of more use to us, or to indulge, in this manner, at the expence of other people, the natural preference which every man has for his own happiness above that of other people, is what no *impartial spectator* can go along with." Often, however, Smith presents generalizations without any explicit reference to 'mankind', 'the spectator', or 'the impartial spectator'. For instance, in *TMS* 1.1.3.1, Smith summarizes how we approve or disapprove of one another: by the *actual* correspondence (or lack thereof) of sentiments. But he then goes on to qualify this account by recognizing alternative means for approving, which, though grounded in the correspondence of sentiments, need not require actual correspondence (*TMS* 1.1.3.3). In other words, his initial account was a generalization of how we tend to approve.

Many have used the term 'the impartial spectator' to refer to conscience. So as D. D. Raphael suggests, we need to distinguish between two senses of 'the impartial spectator': the "more mundane sense of an actual spectator of the conduct of other persons" (Raphael 2007, 42 fn. 12), and the imagined spectator, "the man within", that judges oneself (Raphael 2007, 36). The latter, as Raphael puts it, is "a creation of my imagination. He is indeed myself, though in the character of an imagined spectator, not in the character of an agent" (Raphael 2007, 35). Hence, Raphael distinguishes the "special concept of the *impartial* spectator", i.e. the projected spectator that judges the agent (Raphael 2007, 33), from the more mundane "man without" (Raphael 2007, 36). Generally, though, Smith himself is often quite clear in distinguishing the two, for he refers to the "special concept of the impartial spectator", i.e. conscience, *not* as 'the impartial spectator' but rather the '*representative* of the impartial spectator' or the spectator '*within*'.

generally are as agents and spectators. We are often not ideal in these ways: we are often unaware of our biases like excessive self-love and uninformed about various details of the situation. So when we appeal to this ideal spectator, we come to a very different judgment than we would were we not trying to be completely impartial and fully informed.

Finally, the progression of Smith's discussion in *TMS* suggests that conscience is the product of our moral sentiments shaped by our experience with others. Prior to his discussion of conscience, Smith discusses how we form judgments of propriety and merit, of how we correct for our sentiments by becoming more impartial in our judgments, and of how we form a better understanding of how impartial spectators tend to judge. If conscience is the next step in this progression, then conscience is not just our judging ourselves from a different perspective. Rather, it is our judging ourselves as we imagine impartial spectators would judge us, given our experience as impartial spectators and as the object of the judgments of impartial spectators. So then to judge ourselves, we appeal to an understanding of impartial spectators that extend beyond ourselves. We appeal to an *ideal* spectator, one that holds qualities that actual spectators, including the one judging oneself, do not, including complete knowledge of the situation and complete impartiality.

In what follows, I will argue that this interpretation of conscience is mistaken. The main point of contention is the sense in which conscience is *supposed* and *impartial*. Treating conscience as an idealized prototype is, I think, inconsistent with Smith's overall theory, particularly his understanding of the sympathetic mechanism and its limits. So when we consider the whole of Smith's theory, we will see that:

[A] When we judge ourselves, we do not 'imagine' a spectator in the sense that we form in our mind a spectator with distinct faculties and principles by

which it judges; rather, we 'imagine' in the sense that we project ourselves into a different situation, thus becoming a spectator unto ourselves.

[B] Conscience can be partial or impartial to varying degrees; it need not be fully (ideally) impartial. When Smith speaks of conscience as 'ideal', he is not referring to the degree of impartiality, but rather its unique position of having access to more information about the situation and about the agent.

By clearing up the sense in which conscience involves adopting an imagined and impartial perspective on ourselves, we can also get a better understanding of how Smith's discussion of conscience fits into the theme of progression in *TMS*. Conscience is ideal in the sense that it has access to information that outside spectators often do not, particularly the agent's sentiments, which outside spectators have to infer. But conscience can still be mistaken, and can most definitely be infected by partiality. Hence, even after conscience is formed, it must be cultivated and refined to serve as a source of accurate moral judgment.

2.5 Interpretive Issue #1: 'Supposed' or 'Imagined' Spectators

Conscience is 'supposed' or 'imagined', as opposed to 'real' or 'actual'. But there are two different senses in which conscience can be imagined:

[S-Self] Conscience is 'supposed' in that the agent projects herself into the situation of a spectator, and judges herself, using only her own faculties and principles.

[S-Other] Conscience is 'supposed' in that the agent considers what another spectator would judge, given that spectator's faculties and principles.[73]

[73] The 'S' in S-Self and S-Other refers to 'supposed'. This distinction is, admittedly, oversimplified. There is a range of possibilities between S-Self and S-Other. We can adopt some of the faculties and principles of others without adopting them all. For instance, in imaginatively entering into the situation of the fictional character Professor Charles Xavier, I might not adopt any of his faculties and principles except for his telepathic abilities, or except for his desire to bring harmony between humans and mutants.

This distinction needs some clarifying. In the [S-Self] sense of imagining, the agent is *not* imagining a spectator as an independently operating entity, with its own set of faculties and principles. Rather, the agent merely determines the *situation* of another person (a spectator) and then projects herself into that situation before evaluating her own motives and conduct. Once she has imagined or determined the situation of a spectator, the agent does not merely consider what she would judge given some set of facts about her dispositions; rather, she *pretends* to be in that situation, imaginatively responding to various facets of that situation *as if* she were actually in that situation.

In the [S-Other] sense, however, the agent does not project herself, but rather her understanding of a particular spectator (whether a specific figure, like her father; a prototype that embodies all, or most, spectators; or even a version of herself), which operates from faculties and principles that are distinct from the agent's own. To be clear, the [S-Other] sense of imagining includes cases in which the agent forms an image of herself, which embodies what the agent believes of herself (e.g. her dispositions), and determines what such a person would judge from the situation of a spectator. In such cases, the agent is not engaging with the spectator's situation *as if* she were actually in the situation, but rather deducing what she would judge from facts about her dispositions, the same way that she would deduce what some other spectator, like a friend or a parent, would judge given that spectator's dispositions.

In the previous section, I presented the evidence for [S-Other], endorsed by, among others, Campbell, Darwall, and Griswold. One piece of evidence I gave was that Smith often speaks of viewing ourselves as *others* are likely to view us. The reference to other spectators

suggests that when we view ourselves, as conscience, we are judging ourselves using what we take to be other spectators' principles, not those which we take to be our own.

But such a reading conflicts with a key feature of Smith's account of moral sentiments: we only have unmediated access to our own faculties and principles, never to those of others. Smith is quite explicit on this point in *TMS* 1.1.1.1 (restated here):

> As we have no immediate experience of what other men feel, we can form no idea of the manner in which they are affected, but by conceiving what we ourselves should feel in the like situation.

And he makes the same point again in *TMS* 1.1.3.10:

> Every faculty in one man is the measure by which he judges of the like faculty in another. I judge your sight by my sight, of your ear by my ear, of your reason by my reason, of your resentment by my resentment, of your love by my love. I neither have, nor can have, any other way of judging about them.

We cannot ever view situations as others do in the sense of exercising *their* faculties and principles. We can only infer what others feel and think, either by considering how *we* would feel and think using our own faculties and assuming their faculties operate similarly, or by examining their verbal and bodily expressions and connecting those expressions to certain feelings and thoughts, as taught to us by experience.

Smith makes a similar point in Part I when he first discusses what happens when an agent's sentiments do not initially agree with the sentiments of spectators. It will be helpful here to quote at length:

> As they continually place themselves in his *situation*, and thence conceive of emotions *similar to* what he feels, so he is as constantly placing himself in *theirs*, and thence conceiving some degree of that coolness about his own fortune, with which *he*

> *is sensible that they will view it*. As they are constantly considering what they themselves would feel, if they actually were the sufferers, so he is as constantly led to imagine in what manner he would be affected if he was only one of the spectators of his own situation. As their sympathy makes them look at it, in some measure *with his eyes*, so his sympathy makes him look at it, in some measure, *with theirs*, especially when in their presence and acting under their observation: and as the reflected passion, which he thus conceives, is much weaker than the original one, it necessarily abates the violence of what he felt before he came into their presence, before he began to recollect in what manner they would be affected by it, and to view his situation in this *candid and impartial light* (*TMS* 1.1.4.8, emphases added).

The first point is that we as agents place ourselves into the situation of a spectator. We are imagining that we are in the situation of the spectator, and from that situation, we imagine what feeling *we* would come to. It is this imaginary change in situation that leads to our coming to a new "reflected passion", which is "similar to" that of the spectator, though "much weaker than the original one". At no point do we exercise our faculties in the manner that the spectator actually does. Instead, our faculties generally operate differently under different situations. In this case, the change in situation (from ours as agent to that of the spectator) results in a change in how *our* faculties operate, thus producing a different sentiment from our original one.

Furthermore, it is the change in situation, not in the adoption of different faculties or principles, to which Smith refers when speaking of viewing ourselves "with [the spectators'] eyes". It is through sympathy that we come to view ourselves as others do. But "sympathy" involves a spectator entering into the situation of an agent, or vice versa, to determine how the *spectator* would feel in the situation of the agent: "*[the spectators]* constantly consider what *they themselves* would feel" (emphases added). Thus, we only view the situation as others do "*in some measure*", for we are still viewing the situation using our own faculties and principles, not theirs. What I am suggesting, then, is that we should read Smith's claims

that we judge "with the eyes" (e.g. *TMS* 3.1.1.2) or "in the light" in which others view us (e.g. *TMS* 2.2.2.1) as our judging ourselves from the *situation* of others, but not with their faculties and principles. Similarly, when Smith speaks of how our judgments "must always bear some *secret reference,* either to what are, or to what, upon a certain condition, would be, or to what, we imagine, ought to be the *judgment of others*" (*TMS* 3.1.1.2), Smith is not claiming that we judge ourselves using what we take to be the principles of others. Rather, he is pointing out that *our* judgments of ourselves using the faculties and principle we typically use when evaluating the motives and actions of others often give us insight into how *others* do or might view us *from their situation*. Smith repeats this point later in the same paragraph when he writes how we approve ourselves by "placing ourselves in his [any other fair and impartial spectator's] *situation*" (emphasis added).

These points extend to Smith's discussion of conscience. Consider how he begins his account of conscience:

> The principle by which we naturally either approve or disapprove of our conduct, seems to be altogether *the same with* that by which we exercise the like judgments concerning the conduct of other people (TMS 3.1.1.2, emphasis added).

We judge ourselves *the same way* we judge others. But the manner in which we judge others is not to enter into the situation of the agent and consider how *others* would feel from the same situation. Rather, we judge an agent by imaginatively entering into the agent's situation and determine what *we* as spectators would feel from that situation. Similarly, when we divide ourselves to both agent and spectator, we as spectators judge us as agent by imaginatively entering into the situation of the agent and considering what *we as spectators* would feel in that same situation, not what others would feel. And it is this change in

situation, in our entering into the situation of an impartial spectator to judge ourselves, that we view ourselves "with the eyes" of others:

> We begin, upon this account, to examine our own passions and conduct, and to consider how these must appear to them, *by considering how they would appear to us if in their situation*. We suppose ourselves the spectators of our own behaviour, and endeavour to imagine what effect it would, *in this light*, produce upon *us* (*TMS* 3.1.1.5, emphases added).

Again, Smith stresses that we consider how we appear to *others* by entering into their situation and consider how *we*, not they, would feel. We see as others do, in *their* light, not by taking on their faculties and principles, but by taking on their *situation*. And it is because we enter into this different situation that we as spectator is not the same as we the agent: "But that the judge should, in every respect, be the same with the person judged of, is as impossible, as that the cause should, in every respect, be the same with the effect" (*TMS* 3.1.1.6). This difference between us as spectator and us as agent is not that we use different faculties and principles to make judgments when occupying these two different perspectives. Rather, our faculties and principles operate *differently* as a result of the change in *situation*.

2.6 Interpretive Issue #2: 'Impartial' Spectators

There is another feature that has been thought central to Smith's account of conscience: impartiality. Smith often refers to conscience as the 'supposed *impartial* spectator'. And in judging ourselves, we often do attempt to judge ourselves from the situation of an impartial spectator. But it is unclear exactly what role impartiality plays in conscience. Some have suggested that impartiality is essential for conscience:

[I-C] Conscience *must be* impartial, i.e. our judging ourselves from the situation of an impartial spectator.

If an agent is not judging herself from the situation of an impartial spectator, then an agent is not judging herself *as conscience*. Thus, conscience is not just any self-judgment; it is an agent's self-judgment as an *impartial spectator*. The support for this interpretation stems from an interpretation of Smith's more general view of moral judgments:

[I-M] Only impartial spectators can have moral sentiments (e.g., Macfie 1967, Griswold 1999, Carrasco 2004 & 2008, Sayre-McCord 2010, Schliesser 2011, and Fleischacker 1999).

Thus, if conscience does indeed issue moral judgments on the agent, then it must be an impartial spectator. I argue that neither of these views is correct, and that Smith does not adopt [I-C]. Smith's recognition that we are liable to self-deception, even in our judgments of ourselves, suggests that Smith makes room for conscience to judge partially. The confusion, I take it, is between moral judgments and *correct* moral judgments. Impartiality is required for forming *correct* moral judgments but not for forming *any* moral judgments.

Commentators have taken different routes in interpreting Smith as endorsing [I-M]. For instance, A. L. Macfie begins by noting a shift in Smith's language: from 'spectator' to 'impartial spectator'. In his presentation of the sympathetic mechanism, Smith refers to the 'spectator' as the one who judges (e.g. *TMS* 1.1.1.4-11, 1.1.3.1, 1.1.4.6-8). But after that, in discussing the amiable and respectable virtues, judgments of merit and demerit, and then later, conscience, Smith refers to the one judging as the 'impartial spectator' (e.g. *TMS* 1.1.5.4, 1.1.5.8, 1.2.3.8, 2.1.2.2-3, 2.2.2.1, 2.3.2.2-4) and the 'supposed impartial spectator' (e.g. *TMS* 3.1.2.32, 3.1.3.1, 6.2.1.22). According to Macfie, this shift indicates that in

presenting the sympathetic mechanism, Smith is merely presenting the psychological mechanism, present in all spectators, that gives rise to fellow-feeling. However, Smith does not gives his account of *moral* sentiments until later, particularly in Part II and III in which Smith refers explicitly to impartial spectators (Macfie 1967, 94).

Let me begin by pointing out that the shift is not as drastic as Macfie presents it. Smith refers to 'impartial spectators' with greater frequency in Part II and III of *TMS*. But Smith also continues to refer to 'spectator' without the qualification 'impartial' after Part I, and specifically *in discussions of moral sentiments* (e.g. *TMS* 2.1.5.6, 2.2.1.1-2, 2.2.1.4, 2.2.1.7, 2.2.2.2, 3.1.1.5-6, 3.1.3.21). So even if there is a shift in Smith's language, it is difficult to identify a specific point at which this shift occurs, and to draw any conclusions from when this shift occurs.

Still, Macfie is correct in pointing out that the sympathetic mechanism alone does not give rise to moral sentiments. The sympathetic mechanism only results in the spectator's actually feeling some semblance of what the spectator would feel in the situation of the agent. But what Macfie overlooks is that the description of the mere psychological mechanism ends at *TMS* 1.1.2. And in *TMS* 1.1.3, Smith begins his account of the judgments of propriety and impropriety:

> To approve of the passions of another, therefore, as suitable to their objects, is the same thing as to observe that we entirely sympathize with them; and not to approve of them as such, is the same thing as to observe that we do not entirely sympathize with them (*TMS* 1.1.3.1).

Smith would agree with Macfie that the feelings in the spectator arising from the sympathetic mechanism are *not* moral sentiments. However, in observing the agreement or disagreement between *that* sympathetic sentiment of the spectator and the original sentiment of the agent,

we form a moral sentiment. Thus, the judgments of propriety and impropriety, described in Part I, *are* genuine moral sentiments. Yet Smith never refers to impartial spectators in this section at all, but instead refers to "the sympathetic emotion of the spectator" (*TMS* 1.1.3.1), suggesting that impartial spectators are not necessary for moral sentiments.

D. D. Raphael gives a different explanation for the use of 'impartial spectator'. There are various passages in *TMS* in which Smith appears to be making generalizations about spectators. Consider, for instance, the following:

> But these, as well as all the other passions of human nature, seem proper and are approved of, when the heart of *every impartial spectator* entirely sympathizes with them, when *every indifferent by-stander* entirely enters into, and goes along with them (*TMS* 2.1.2.2, emphases added).

This passage is in Part II, but seems to refer back to Part I in which Smith discusses propriety. And so even though Smith does not always use 'spectator' with the qualification 'impartial', he seems to have had impartial spectators in mind all along. As Raphael puts it:

[Smith] knows that unanimity cannot always be guaranteed and so he sometimes introduces a slight qualification, as in "every impartial spectator" or "every indifferent by-stander" (*TMS* II.i.2.2); but, since the passage containing those phrases is almost immediately followed by "every human heart", the qualification counts for little. Smith in fact takes for granted that a spectator or bystander will be impartial just because he is not a party to the conduct judged (Raphael 2007, 17).

So even though Smith sometimes does qualify 'spectators' with 'impartial' or 'indifferent', this qualification holds little weight – contra Macfie's reading – because Smith seems to assume that the spectators for which his generalizations hold are impartial.[74] To some degree, Raphael is correct. All spectators tend to be at least more impartial than the agents that they judge because we generally do not feel for others the degree to which we feel for ourselves. The spectator, in entering into the situation of *another* and imagining how the spectator would feel, generally has a dulled response to the situation because the spectator *is not the agent,* not the one that is actually going through that situation:

> My companion does not naturally look upon the misfortune that has befallen me, or the injury that has been done me, from the same point of view in which I consider them. They affect me much more nearly (*TMS* 1.1.4.5).

In other words, because spectators are entering into the situation of someone else, and because generally people are more concerned about themselves than others, spectators will not be as concerned for the agent in the agent's situation as the agent is for herself. And so spectators are generally more impartial than the agents that the spectators judge.

There are, however, several problems with Raphael's conclusions. For one, even if it turns out that spectators are generally impartial, that fact alone does not show that partial spectators *cannot* make moral judgments. At most, that most spectators are more impartial about the actions of agents than are those agents themselves shows that most of those who

[74] Similarly, Campbell argues that because impartial spectators are judging from this shared position (similar to Hume's common point of view; see Cohen 1997 and Korsgaard 1999), Smith can speak of impartial spectators as *the* impartial spectator: "The spectator standpoint is the common ground which unites men because it is when they are in the position of non-involved spectators that they tend to agree most readily…. This process… explains why he thinks he is justified in talking of *the* impartial spectator" (Campbell 1971, 138). "*The* impartial spectator" is simply a way of speaking of the generalizations that appear to hold for all impartial spectators. So when Smith refers to conscience as *the* supposed impartial spectator, Smith is presenting conscience as the embodiment of all the generalizations we hold of *impartial* spectators.

make moral judgments are more impartial than the agents they judge. Or alternatively, it may be that Smith is giving a more limited account of moral judgments, restricting his discussion to how *impartial* spectators make moral judgments. But even if Smith were to assume that spectators are at least more impartial on these matters than those they judge, he is in no way committed to the stronger claim that they *must* be more impartial. Moreover, he surely need not assert that spectators are fully impartial. This would be an exceedingly uncharitable interpretation as Smith agrees with Hume that when a spectator judges the reactions, motives or behavior of someone he loves, he is much less impartial than when evaluating a stranger.

But setting this issue aside, there are problems with interpreting Smith in this manner, i.e. as using 'spectator' to refer to impartial spectators. For one, Smith is well aware that spectators can be partial, even if they are not a party to the conduct judged, and can be partial to varying degrees. If by 'impartial' Raphael means the spectator is devoid of *any* or *almost all* bias, to a larger extent, he is simply mistaken. Though a spectator may not be affected by self-love, she may very well be partial for other reasons, like particular habits and customs that make the spectator more or less likely to agree in sentiment with the agent (e.g. 1.2.2.1-6, 6.1.1.2). Spectators can be biased in various ways, depending on their relationship with the agent, e.g. family, friends, acquaintances, that affect their judgments without the presence of any self-interest. On the other hand, if by 'impartial', Raphael means the spectator is less partial than the agent because of the absence of self-love, then it is unclear how this greater impartiality provides the kind of generalization that Raphael claims Smith is making. Given the variance in degree of bias that could still exist with this minimal degree of impartiality, Smith's generalizations about how we judge would not hold true. Take a simple example like determining the proper punishment to an agent for a wrong done. Even if all spectators are

less partial than the agent, each spectator would come to different judgments depending, in part, on their relationship to the agent. Friends and family, because of past experiences, will be more sympathetic to the agent than a stranger, and even more so than friends and family of the victim, who would be more sympathetic to the victim.[75] So Raphael might be correct that spectators are generally more impartial since self-love does not bias their judgment the way that it does the parties involved. But this would not reflect Smith's use of 'impartial', which implies a much greater degree of impartiality.

Furthermore, there are many instances throughout *TMS* in which it would be a mistake to assume that 'spectator' without qualification is referring to an impartial spectator. Let me just point out one such instance, though I think there are many more. In *TMS* 2.2.3.11, Smith discusses a case in which a spectator "both punish[es] and approve[s] of punishment, merely from a view to the general interest of society, which, we imagine, cannot otherwise be secured." In such cases, the spectator is faced with multiple sentiments, founded on very different principles.[76] One sentiment is that which makes the severity of the punishment "appear necessary, and, for that reason, just and proper." This sentiment is grounded in a principle that causes us to look at the more "remote consequences", "either a considerable inconveniency, or a great disorder in the society." The consideration of the consequences on the society-at-large seems to suggest that the spectator is impartial. The problem is that Smith here is distinguishing the judgment resulting from consideration of society-at-large with judgments stemming from the spectator's sympathetic resentment (resentment that resonates with an impartial spectator), which makes the punishment seem

---

[75] Because of these sources of bias other than self-love, Campbell, for instance, interprets Smith as assuming the stronger sense of impartiality in making generalizations in, e.g., *TMS* 1.2.3.1 (Campbell 1971, 110).

[76] "The very different sentiments with which the spectator views those different punishments, is a proof that his approbation of the one is far from being founded upon the same principles with that of the other" (*TMS* 2.2.3.11).

"excessively severe." This contrast suggests that judgments stemming from considerations of the "remote consequences" of an action are not judgments of merit made from an impartial situation because they do not make use of the sympathetic mechanism. Rather, they are more akin to those resulting from following general rules which are used to correct judgments of *partial* spectators. This is not to say that the spectator is not impartial to any degree. Instead, since this particular spectator appears to be judging from multiple principles, and not all require impartiality, it makes sense to leave open the question of the spectator's impartiality. In general, we should pay close attention to the context, because though Smith often assumes impartiality for certain generalizations, he does also speak of people in general, both partial and impartial.

Finally, assuming impartiality from the beginning fails to appreciate a central theme in Smith's work: we *become* more impartial by correcting for our biases.[77] In Part I, Smith gives a detailed explanation of how we may be partial in our judgment, and then come to be more impartial through interaction with others. Whether or not we are partial, we only have one way to judge the faculties of others: by comparing theirs with ours. When we run into a disagreement with others, we attempt to come to an agreement by disengaging from our situation and entering into their situation because we usually desire to have corresponding sentiments (*TMS* 1.1.2.1). The key is that *both* the spectator and the agent try to see through the eyes of one another. Through this process, both become more impartial because by seeing through each others' eyes, they come to a *shared* situation, one that is not particular to one or the other. In that situation, they gain a degree of indifference which allows them to overlook any difference in opinion and have fellow-feeling: "so that, though our opinions may be

[77] Many have already recognized Smith's focus on moral development and education. For a more in-depth discussion, see Evensky (2005), Hanley (2009), Heilbroner (1982), and Macfie (1967).

opposite, our affections may still be very nearly the same" (*TMS* 1.1.4.5). A spectator can do so by adopting "the whole case of his companion with all its minutest incidents; and strive to render as perfect as possible, that imaginary change of situation" (*TMS* 1.1.4.6). The agent, likewise, must enter fully into the situation of the spectator, "conceiving some degree of that coolness about his own fortune, with which he is sensible that they will view it" (*TMS* 1.1.4.8). By doing so, the agent comes to "view his situation in this candid and *impartial* light," since one is able to place oneself in a more disinterested position, i.e. one that is not clouded by self-love and custom.[78] It is particularly telling that this discussion of becoming more impartial comes *after* the discussion on how people make judgments of propriety and impropriety. If impartiality of the spectator is assumed in explaining how people come to approve and disapprove, it is unclear why Smith would wait to discuss impartiality, and even less clear why he would feel a need to give an account of how people become more impartial.[79]

[78] It is easy to read *TMS* 1.1.4.8, which states that the agent comes to "view his situation in this candid and *impartial* light", as only making a claim about the agent's impartiality (e.g. Raphael 2007). But remember that this candid and impartial view is referring to the view of the spectator who not only engaged in this imaginative exercise of entering into the agent's situation, but also does so *fully*. Just as the agent may be partial, so might the spectator. And in cases that the spectator is partial, it is the spectator that needs to imaginatively enter into a different situation to come to a sentiment that agrees with the agent.

[79] Raphael points out that this first appearance of 'impartial spectator' occurs in the context of the virtue of self-command (*TMS* 1.1.5.4). According to Raphael, Smith only stresses impartiality of the spectator when discussing "the *effect on the agent* of the reactions of spectators" (also see Shaver 2007, 193); hence, Raphael stresses the connection between impartiality and the virtue of self-command, since it is only the impartiality of the one judged that is of interest (Raphael 2007, 34). Self-command, then, is "essentially to feel for ourselves only what we see others can feel for us", for the one judged to come closer to impartiality by viewing herself through the eyes of the spectator. I think Raphael is mistaken here. He is correct that impartiality is *first* introduced in the context of self-command, but he overlooks the fact that Smith is specifically speaking of the self-command of one's feelings of resentment. The excessively strong feeling of resentment, the one that the impartial spectator cannot go along with, is what Smith calls 'rage'. The more moderate feeling of resentment, the one felt by the impartial spectator, is what Smith calls 'that noble and generous resentment'. Smith simply stresses the person judged coming to a more impartial stance. However, as I have tried to show, the spectator can also come to a more impartial stance by bringing herself closer to the situation of the one judged. Hence, impartiality is connected to *both* the virtue of self-command *and* the virtue of humanity.

The only other time we find 'impartial spectator' in Part I is in *TMS* 1.1.5.8, which is not in the context of resentment. However, this passage harkens back to the discussion in *TMS* 1.1.4.5-8. Both passages discuss cases in which the spectator and the agent disagree in sentiment. In these cases of disagreement, both the

More recently, Maria Alejandra Carrasco has taken a quite different approach. Carrasco endorses both [I-C] and [I-M]: that conscience must be impartial and that only impartial spectators can make moral judgments, respectively. But she does so because she endorses an even stronger claim: moral sentiments require an appeal to *conscience*. Her argument relies on a distinction between the brute and uncultivated "natural sentiments" of partial spectators and the "moral" or "appropriate" sentiments of impartial spectators (Carrasco 2004, 87 and 2008, 136). Natural sentiments are our initial responses to an agent's motives or action. Such feelings are not adjusted in any manner, and are generally excessive or deficient in the eyes of other (impartial) spectators. Moral sentiments, on the other hand, are sentiments of an ideal impartial spectator, i.e. conscience. By considering what an impartial spectator would feel, we adjust our initial (natural) sentiments to align more closely with that of conscience. We come to feel the appropriate sentiment, the sentiment that all impartial spectators would feel. And without this refinement through the appeal to conscience, we simply do not have moral sentiments.

In defense of her interpretation, Carrasco points to a passage from *TMS* 3.1.3.28, in which we are torn between two sentiments:

> In such paroxysms of distress, if I may be allowed to call them so, the wisest and firmest man, in order to preserve his equanimity, is obliged, I imagine, to make a considerable, and even a painful exertion. His own natural feeling of his own distress, his own natural view of his own situation, presses hard upon him, and he cannot, without a very great effort, fix his attention upon that of the impartial spectator. Both views present themselves to him at the same time. His sense of honour, his regard to

---

spectator and the agent try to become closer in sentiment to one another. In *TMS* 1.1.5.8, Smith is pointing out that even after the spectator enters as much as possible into the situation of the agent, and the agent tries to lower the degree of sentiment, there may be no fellow-feeling because the sentiments of the agent are too overwhelming.

> his own dignity, directs him to fix his whole attention upon the one view. His natural, his untaught and undisciplined feelings, are continually calling it off to the other. He does not, in this case, perfectly identify himself with the ideal man within the breast.

Smith is presenting a conflict between our "natural feelings" and the feelings we have in virtue of adopting the perspective of an impartial spectator. Carrasco takes this distinction to be between natural sentiments, which are our "untaught and undisciplined feelings", and *moral* sentiments, which are disciplined. And in this context, these disciplined feelings are referring to those of conscience, which is "impartial, universal, and reflexive" (Carrasco 2004, 86). Thus, Carrasco concludes:

> I think it is clear that Smith does not believe that our notions of moral good and evil arise from our brute or natural sentiments but rather from our sentiments informed by the impartial spectator. Before the development of this "inmate within the breast" we simply do not make moral judgments…. The contrivance of our nature that enables us to identify good and evil certainly includes sentiments, but sentiments informed by the impartial spectator (Carrasco 2004, 87).[80]

Moral sentiments, then, are the sentiments we feel after considering what conscience, an impartial spectator, would feel, and aligning our sentiment with it. They are not simply any sentiment arising from the sympathetic mechanism, but rather one arising from the sympathetic mechanism applied to conscience's perspective to adjust our original sentiment.

[80] By "inmate within the breast" and "impartial spectator", Carrasco is referring to the supposed impartial spectator, i.e. conscience. Sayre-McCord provides a similar analysis in which judgments of deservedness are tied to the view of the supposed impartial spectator, which he calls 'Impartial Spectator' (Sayre-McCord 2010, 130). For our purposes, we only need to note that Carrasco is claiming that any moral judgment is one made from a *moral* sentiment, i.e. a sentiment informed by an impartial stance, and not a natural sentiment.

Eric Schliesser makes a similar distinction, but he views moral sentiments as "the cultivated feelings humans acquire from the local social institutions that acculturate them" (Schliesser 2011, 20). And so unlike Carrasco, Schliesser does not require the appeal to the supposed impartial spectator to form moral sentiments; instead, he requires the cultivation of sentiments from social settings to form moral sentiments. Though my responses are directed at Carrasco, they are equally applicable to Schliesser.

So then it is not until Part III, when Smith introduces conscience, that Smith gives an account of moral sentiments.

There are several issues with Carrasco's interpretation. The strongest, most direct evidence against Carrasco's reading is the following passage:

> [O]ur first *moral criticisms* are exercised upon the characters and conduct of *other people*; and we are all very forward to observe how each of these affects us (*TMS* 3.1.1.5, emphases added).

If Carrasco is correct, then our first moral sentiments occur after the development of conscience. But this passage is a summary of what he has presented thus far, and it occurs *before* his presentation of conscience. More importantly, Smith claims that our first moral judgments are of *others*. And given the context, he is referring to his account of how we judge *others* presented in Part I and II: "In the forgoing parts of this discourse, I have chiefly considered the origin and foundation of our judgments concerning the sentiments and conduct of others" (*TMS* 3.1.1.1). Conscience, on the other hand, is described as the means by which we judge *ourselves*: "I come now to consider more particularly the origin of those concerning our own" (*TMS* 3.1.1.1).

Her reading of *TMS* 3.1.3.28 is also problematic. In that passage, Smith is without a doubt distinguishing between our natural ("untaught and undisciplined") sentiments and our sentiments informed by conscience. Carrasco assumes that this distinction is between non-moral and moral sentiments. But Smith *never* makes such a claim, nor does he suggest such a distinction, for instance, through his use of 'moral' or 'morality.' Instead, the context suggests that Smith has a wholly different distinction in mind. In this chapter, Smith is discussing how conscience disciplines, refines, and corrects our sentiments:

> In the same manner, to the selfish and original passions of human nature, the loss or gain of a very small interest of our own, appears to be of vastly more importance, excites a much more passionate joy or sorrow, a much more ardent desire or aversion, than the greatest concern of another with whom we have no particular connexion…. Before we can make any proper comparison of those opposite interests, we must change our position. We must view them, neither from our own place nor yet from his, neither from our own eyes nor yet with his, but from the place and with the eyes of a third person, who has no particular connexion with either, and who judges with impartiality between us (*TMS* 3.1.3.3).

The "natural feelings" that Smith describes in *TMS* 3.1.3.28 are these "selfish and original passions of human nature." The key is that the original sentiment is what we feel *as an agent*, not as a spectator. Upon encountering some situation that affects us, we initially tend to feel a stronger degree of self-love than what would be agreeable to others. These "selfish passions" are described in detail in *TMS* 1.2.5. This excess of self-love, however, is felt by the one who is now under judgment by conscience, not by a spectator of another. In contrast, the feelings we have as conscience are what we feel *as spectator*, specifically as an impartial spectator of ourselves. We view ourselves from the situation of an impartial spectator, and feel very differently than we do as agent. And through this change in person, we correct for this "natural inequality of our sentiments", i.e. we feel something much closer to what all other spectators would feel.

This contrast parallels Smith's discussion of how we divide ourselves when judging ourselves: "I, the examiner and judge" and "I, the person whose conduct is examined into and judged of" (*TMS* 3.1.1.6). And this parallel is maintained in *TMS* 3.1.3, "Of the influence and authority of conscience", in which Smith describes people of differing degrees of self-command. The struggle is between viewing the situation as an agent, with all the original, natural sentiments that come with such a viewpoint, and viewing the situation as a spectator,

with the more refined sentiments that align much closer to those of others. The child with no self-command has all these natural sentiments without any restraint: "it endeavours always, by the violence of its outcries, to alarm, as much as it can, the attention of its nurse, or of its parents" (*TMS* 3.1.3.22). The person of weak self-command is only able to view herself as a spectator for a short duration, and is only "becalmed the moment [spectators] come into his presence" (*TMS* 3.1.3.23). She is unable to continuously view herself as a spectator. She is soon brought back to feel what she originally did *as agent*, to her original (natural) passions of "sighs and tears and lamentations" felt from "[h]is own view of his situation" (*TMS* 3.1.3.23). The person of strong self-command, the "man of real constancy and firmness", on the other hand, is able to constantly view herself as conscience, and feeling as she does from this viewpoint *as a spectator*:

> He has never dared to forget for one moment the judgment which the impartial spectator would pass upon his sentiments and conduct…. He does not merely affect the sentiments of the impartial spectator. He really adopts them. He almost identifies himself with, he almost becomes himself that impartial spectator, and scarce even feels but as that great arbiter of his conduct directs him to feel (*TMS* 3.1.3.25).

Unlike those of weaker self-command, this person constantly views himself as a spectator, feeling as a spectator does, never returning to the original feelings from the viewpoint as agent.

The contrast between what we feel as agent and what we feel as spectator is revealing. Carrasco is correct that the natural sentiments are not moral sentiments. But that they are not moral sentiments has nothing to do with the lack of appeal to conscience. Instead, it is because natural sentiments are sentiments felt *as agents*, and moral sentiments are sentiments felt by *spectators*. In Smith's account, moral sentiments are sentiments of

spectators in response to something felt or done by *agents*. They are the sentiments we feel when we as spectators judge others by entering into their situation and considering what we would feel from that same situation. And because unlike the agent's sentiments, the spectator's sentiments are a product of this imaginative exercise, the spectator's sentiments will always be different from the agent's:

> What [the spectators] feel, will, indeed, always be, in some respects, different from what [the agent] feels, and compassion can never be exactly the same with original sorrow; because the secret consciousness that the change of situations, from which the sympathetic sentiments arises, is but imaginary, not only lowers it in degree, but, in some measure, varies it in kind, and gives it a quite different modification (*TMS* 1.1.4.7).

Similarly, when we judge ourselves, we take on the roles of both agent and spectator. But only we *as spectators* are the ones who are engaging in this imaginative exercise, and who have moral sentiments. The distinction between natural and refined sentiments is not essentially one between non-moral and moral sentiments. Rather, natural sentiments are the initial, often partial, sentiments we feel *as agents*, and the refined sentiments are the sentiments we feel when we *as spectators* attempt to judge ourselves. These refined sentiments are moral sentiments in virtue of their being the sentiments of *spectators* judging agents, not because they result from conscience.

Finally, Carrasco is correct that Smith thinks we adjust our sentiments by viewing the situation in a more impartial light, and that this process of adjusting our sentiments by appealing to the impartial perspective is part of moral education. However, Carrasco thinks this process is necessary for moral sentiments. But Smith often speaks of moral sentiments of *partial* spectators, of spectators who fail to appeal to this impartial perspective. And when

such partial spectators do appeal to the impartial spectator, they *correct*, not form, moral sentiments. In other words, the appeal to conscience is not necessary for the formation of moral sentiments, but rather for the *correcting* of both our moral and non-moral sentiments. For instance, in *TMS* 1.3.3, Smith provides a detailed account of the various ways in which our "disposition to admire the rich and the great, and to despise or neglect persons of poor and mean condition" *corrupt* our moral sentiments. He states:

> This disposition to admire, and almost to worship, the rich and the powerful, and to despise, or, at least, to neglect persons of poor and mean condition, though necessary both to establish and to maintain the distinction of ranks and the order of society, is, at the same time, the great and most universal cause of the corruption of our moral sentiments (*TMS* 1.3.3.1).

This, of course, does not indicate that a spectator need not be impartial to make moral judgments. But his discussion here is important for understanding how he uses 'corruption'. In the "middling and inferior stations of life," acquisition of wealth and greatness often results from praiseworthy virtuous character: "real and solid professional abilities, joined to prudent, just, firm, and temperate conduct, can very seldom fail of success" (*TMS* 1.3.3.5). However, this is not so for the "superior stations of life," in which "success and preferment depend, not upon the esteem of intelligent and well-informed equals, but upon the fanciful and foolish favour of ignorant, presumptuous, and proud superiors" (*TMS* 1.3.3.5). Because we associate wealth with certain praiseworthy characters, we often approve of the wealthy, even when they lack virtue. The result is that people begin to seek out wealth while abandoning the cultivation of virtuous characters. It is this confusion that Smith calls a "mistake" (*TMS* 1.3.3.3). And it is the "most studious and careful observer" that does not make this mistake in *moral sentiments* and aims for what is "more correct" (*TMS* 1.3.3.2).

Later, Smith speaks explicitly of different kinds of partiality that are, like the disposition to admire the rich and powerful, 'corrupters' of moral sentiments. In *TMS* 3.1.3.41, he states: "The propriety of our moral sentiments is never so apt to be corrupted, as when the indulgent and partial spectator is at hand, while the indifferent and impartial one is at a great distance." Soon after, he specifies one kind of partiality: "Of all the corrupters of moral sentiments, therefore, faction and fanaticism have always been by far the greatest" (*TMS* 3.1.3.43). And in Part V, Smith points out another 'corrupter of moral sentiments':

> There are other principles besides those already enumerated, which have a considerable influence upon the moral sentiments of mankind, and are the chief causes of the many irregular and discordant opinions which prevail in different ages and nations concerning what is blameable or praise-worthy. These principles are custom and fashion, principles which extend their dominion over our judgments concerning beauty of every kind. (*TMS* 5.1.1.1).

And these distorted moral sentiments, born out of the bias of custom and fashion, produce perverse moral judgments (see *TMS* 5.1.2.12).[81] In each of these cases, Smith is pointing out that our *moral* sentiments can be corrupted by various partialities of the spectator. And by 'corrupted' moral sentiments, all Smith means is that they are incorrect, not that they are not moral sentiments at all. So there is a connection between moral sentiments and impartiality: moral sentiments are often *corrupted* when the spectator is partial, e.g. when influenced heavily by custom and fashion or by faction.

This treatment of corrupted moral sentiments fits in nicely with Smith's analogy between moral judgments and perceptual judgments. According to Smith, we make errors in our judgments of the dimensions of objects at a distance because by the "eye of the body",

[81] More specifically, custom and fashion do not significantly affect our judgments of the "general style of character and behaviour", but they do affect our judgments of the "propriety or impropriety of particular usages" (*TMS* 5.1.2.12).

"objects appear great or small, not so much according to their real dimensions, as according to the nearness or distance of their situation" (*TMS* 3.1.3.2). We *correct* this inaccurate judgment by imaginatively entering into a more impartial situation. Similarly, we make errors in our moral judgments because our sentiments are often distorted as a result of our particular situation. And so "we *remedy* the defects of both these organs pretty much in the same manner" (*TMS* 3.1.3.2). In both cases, we form judgments using what the "eye of the body" and the "natural eye of the mind" (i.e. moral sentiments) give us, and when these judgments are erroneous, we *correct* for them by entering into a more impartial situation.[82]

In sum, spectators, including those employing their conscience, need not be impartial to have moral sentiments. Generally, spectators will be more impartial than the agent they are judging, for the agent's reactions to her situation are affected by self-love in a way that does not affect the reactions of spectators evaluating the propriety of the agent's reactions and the merit of the motives and actions to which they give rise. But spectators can be just as partial, if not more so, than agents, and still form moral sentiments; it is just unlikely that their sentiments will consistently agree with those that would be felt by impartial spectators.

2.7 Interpretive Issue #3: 'Ideal'

Several times, Smith refers to conscience as an 'ideal': the "ideal man within the breast" (*TMS* 3.1.3.26, 3.1.3.28-9, 3.1.4.4) and the "ideal spectator" (*TMS* 3.1.3.38). The general assumption is that this ideal judge embodies all the characteristics needed to form

[82] Smith also treats general rules like the appeal to impartial spectators in that general rules are used to *correct* our initial sentiments, e.g. *TMS* 2.2.3.10: "Our disapprobation of his ordinary character and conduct does not in this case altogether prevent our fellow-feeling with his natural indignation; though with those who are not either extremely candid, or who have not been accustomed to *correct and regulate* their natural sentiments by general rules, it is very apt to damp it" (emphasis added). This passage also serves as an example of how Smith uses 'natural' to describe our moral sentiments in both partial and impartial situation. The 'natural indignation' is the indignation that rises to the level that an impartial spectator would feel, whereas the 'natural sentiment' that is corrected and regulated by general rules is the sentiment that would arises in a partial spectator.

ideal (correct) moral judgments. In particular, this ideal judge is understood to be completely impartial and fully informed. For example, in the introduction to *TMS*, Haakonssen describes conscience as follows:

> We tend to imagine how a spectator would judge us and our behaviour if he or she was not limited by prejudice, partiality, ignorance, poor imagination and lack of ordinary good will in the way in which the actual spectators of us, including we ourselves, are limited. We imagine an ideal judgment and an ideal judge" (Haakonssen 2002, xv).

Similarly, Carrasco claims that conscience, though initially nothing more than "social consensus", eventually develops an independence from this consensus and becomes the kind of ideal judge that Haakonssen describes (Carrasco 2004, 106-7). The evidence for attributing these features to Smith's conception of an ideal judge consists in various characterizations that Smith gives of conscience: "equitable judge" (*TMS* 3.1.1.2), "well-informed" (*TMS* 3.1.2.32), "awful and respectable judge" (*TMS* 3.1.3.25), "complete impartiality of an equitable judge" (*TMS* 3.1.4.3), and "the severe eyes of the most impartial spectator" (*TMS* 3.1.4.4). These descriptions leave us with the impression that conscience is ideal at least in that it is completely impartial and fully informed, endowed with all the capacities to judge, including superior imaginative and emotional abilities. As a result, conscience issues only ideal, i.e. correct, moral judgments.

Under this interpretation of 'ideal', conscience is a spectator that feels and judges differently than any particular spectator does. When we judge others as spectators or adopt the perspective of spectator in judging ourselves, we are rarely fully informed and fully impartial. If conscience has these properties, judgments of conscience are ideal in ways that our typical judgments often are not. Thus, when we appeal to conscience and imagine how

conscience judges us, we must not be exercising our faculties as *we* generally do to judge others, but rather exercising our faculties as *someone else*, an abstract entity or a prototype that has all the features of an ideal judge, would judge.

Though Smith does speak of conscience as an 'ideal', Smith also speaks of conscience as faulty in certain respects, which suggests Smith does not think that conscience is ideal in the ways Haakonssen suggests. The primary textual evidence for my position comes from Smith's discussion of self-deception in *TMS* 3.1.4. According to Smith, though we often attempt to judge ourselves as impartial spectators, we often fail to do so:

> So partial are the views of mankind with regard to the propriety of their own conduct, both at the time of action and after it; and so difficult is it for them to view it in the light in which any indifferent spectator would consider it (*TMS* 3.1.4.5).

It is clear from the first half of this passage that we can form partial moral judgments of ourselves. But it is less clear what the self-deception is. The claim might be that we fail to judge ourselves as an impartial spectator and so fail to judge ourselves *as conscience*, for conscience is necessarily our judging ourselves as *ideal* (and thus impartial) spectator. The self-deception, then, is that we believe we are judging as conscience even though we are not, that we are hearing the voice of conscience when it is not conscience that is speaking to us. On the other hand, the claim might be that we judge ourselves as conscience, believing conscience to be impartial when it is not. Under this interpretation, the self-deception is that we believe conscience to be impartial when it is not, that conscience is judging us impartially and with full knowledge of the situation when it is not doing so. Conscience, in other words, need not be ideal.

To settle this matter, then, we need to determine whether we are appealing to conscience when, unbeknownst to us, we judge ourselves as partial spectators. If conscience must be an ideal (thus fully impartial) spectator, then we must not be appealing to conscience when we judge ourselves as partial spectators. But if conscience need not be ideal, then we can still be appealing to conscience when we judge ourselves as partial spectators. At times, Smith seems to adopt the former:

> The violent emotions which at that time agitate us, discolour our views of things; even when we are endeavouring to place ourselves in the situation of another, and to regard the objects that interest us in the light in which they will naturally appear to him, the fury of our own passions constantly calls us back to our own place, where every thing appears magnified and misrepresented by self-love (*TMS* 3.1.4.3).

The passage here is similar to Smith's discussion of the different degrees of self-command in *TMS* 3.1.3. Those of weak self-command are unable to continually appeal to conscience. Their difficulty is not in judging correctly as conscience, but rather in maintaining that perspective of conscience, in remaining in the situation of an impartial spectator: "with a weak man, it is not of long continuance" (*TMS* 3.1.3.23). They soon abandon their view as conscience, and return back to their own situation. Those of great self-command, however, are able to maintain that perspective of an impartial spectator, to continually appeal to conscience, so much so that they *identify* with it (*TMS* 3.1.3.25). Smith might be presenting the same difficulty here: we are attempting to view ourselves as conscience, but fail to because "the fury of our own passions constantly calls us back to our own place", i.e. our own partial perspective on and reactions to our situation (colored as they are by self-love) and not that of conscience.

There is, however, a relevant difference between *TMS* 3.1.3 and 3.1.4. In *TMS* 3.1.3, Smith is discussing how we are often partial when an impartial spectator, either real or imagined, is not present: “The propriety of our own moral sentiments is never apt to be corrupted, as when the indulgent and partial spectator is at hand, while the indifferent and impartial one is at a great distance” (*TMS* 3.1.3.41). But in *TMS* 3.1.4, Smith explains how we can be partial *even when* an impartial spectator is present:

> In order to pervert the rectitude of our own judgments concerning the propriety of our own conduct, it is not always necessary that the real and impartial spectator should be at a great distance. When he is at hand, when he is present, the violence and injustice of our own selfish passions are sometimes sufficient to induce *the man within the breast* to make a report *very different from what the real circumstances of the case are capable of authorizing* (*TMS* 3.1.4.1, emphases added).

The problem is not that we fail to view ourselves as conscience; Smith is clear here that the “man within the breast”, i.e. conscience, is present and forming a judgment. The problem is that when our self-love is far too strong, then even conscience can form sentiments that are not called for by the “real circumstances”, i.e. the actual situation. Conscience, in other words, can be infected by self-love such that it issues partial judgments. This point is repeated later in *TMS* 3.1.4:

> Of the manner in which those objects would appear to another, of the view which he would take of them, we can obtain, if I may say so, but instantaneous glimpses, which vanish in a moment, and which, *even while they last, are not altogether just* (*TMS* 3.1.4..3, emphasis added).

Smith is making two points here. For one, as he spoke of in *TMS* 3.1.3, we sometimes have difficulty sustaining this viewpoint of an impartial spectator. Hence, we might only get “instantaneous glimpses, which vanish in a moment”. But Smith is also pointing out that,

even when we do judge ourselves from that situation, i.e. even as conscience, our judgments "are not altogether just". Smith, in other words, recognizes that even conscience can err in its judgment, and hence is not always ideal.

Later on, Smith states that we fail to judge ourselves correctly as conscience, even when we judge ourselves for something we have already done and our emotions regarding that situation have subsided:

> [W]hen the paroxysm of emotion, in the same manner as when the paroxysm of distress, is fairly over, we can identify ourselves, as it were, with the ideal man within the breast, and, in our own character, view, as in the one case, our own situation, so in the other, our own conduct, with the severe eyes of the most impartial spectator…. *It is seldom, however, that they are quite candid even in this case*…. It is so disagreeable to think ill of ourselves, that we often purposely turn away our view from those circumstances which might render that judgment unfavourable (*TMS* 3.1.4.4, emphasis added).

Again, it is not that we fail to judge as conscience. Smith is clear that we are identifying ourselves with "the ideal man within the breast", "with the severe eyes of the most impartial spectator". So we do not fail to enter into the situation of an impartial spectator to judge ourselves. Rather, *while in that situation* of an impartial spectator, we begin to distort our own beliefs and sentiments to justify our prior actions. In other words, we do not use our faculties and principles the way we generally would when judging others, i.e. we are not *candid* in our judgments of ourselves.[83] While in this imaginative exercise, we purposely

[83] To judge candidly is to judge in a manner consistent with how one's own faculties and principles generally operate (and in that sense, to judge *honestly*). Thus, when we judge ourselves candidly, we judge ourselves as we would generally judge others. Judging consistently in this manner is not necessarily the same as judging impartially, for a spectator can be impartial (i.e. hold no connections with any of the parties involved) and still fail to judge consistent with human faculties and principles. Thus, Smith supplements 'impartial' with 'candid' to stress that it is sometimes not enough simply to be an impartial spectator to make the correct or proper moral judgment: "[He] strives to regard himself, not in the light in which he at present appears, but in that in which he ought to appear… *if the sentiments of mankind were either altogether candid and equitable*, or even perfectly consistent with themselves. *The more candid and humane* part of mankind entirely go along with the effort

ignore certain features of the situation, or we "endeavour to exasperate anew those unjust passions which had formerly misled us; we endeavour by artifice to awaken our old hatreds, and irritate afresh our almost forgotten resentments" (*TMS* 3.1.4.4). In such cases, conscience is present, but it fails to judge as an ideal spectator by failing to judge with full impartiality or to judge without considering all features of the situation.

But if conscience need not be ideal, why does Smith refer to conscience as the 'ideal judge'? I want to suggest that all that Smith means by 'ideal' is that conscience is in the best possible situation to judge correctly. Unlike the agent, conscience is one step removed from the agent's situation and so less likely to feel too much self-love. And unlike outside spectators, conscience has better access to the agent's motives, at least the motives of which the agent is aware.[84] However, though conscience judges from an optimal position, it does not necessarily judge correctly, for even from this advantageous situation, it remains, as Smith notes, susceptible to the distorting influence of self-love.

The primary textual evidence for this reading concerns Smith's two distinctions: 1) between sentiments of agents and those of spectators, and 2) between our judgments of ourselves and other spectators' judgments of ourselves. I presented the first distinction in §2.6 in response to Carrasco's interpretation.[85] In that discussion, I argued that the spectator in question need not be conceived of as ideally impartial and so need not present his evaluations as those of his conscience. However, Carrasco is right that conscience tends to be

---

which he thus makes to support himself in his own opinion (*TMS* 2.3.3.6, emphases added; also see *TMS* 1.1.4.8, 3.1.3.28, 7.2.4.10).

[84] My interpretation aligns with that of Vivienne Brown :"The impartial spectator is ideally placed to do just this, as his location with respect to the moral agent provides impartiality, while his knowledge of the circumstances of the case enables him to make an informed judgment" (Brown 1994, 29). Raphael (1975, 95) and Evensky (2005, 36-37) also provide similar interpretations. However, there is little defense of this interpretation. So my aim is to provide richer textual evidence in support of this interpretation.

[85] My point there was to show that moral sentiments are those of spectators (whether other spectators or an agent's viewing herself from a spectator's perspective) and involve the comparing of sentiments, whereas natural sentiments are those of unreflective agents and involve no comparison of sentiments.

more impartial than the agent *in virtue of being a spectator*. The agent is directly affected by her situation; spectators are always one step removed from that situation. As a result, spectators are generally more impartial than the agent. Smith initially made this point in *TMS* 1.1.4.8, when he speaks of how the mere viewing ourselves from the situation of a spectator, not necessarily an impartial one, results in our viewing ourselves in a "candid and impartial light". By becoming a spectator, the agent leaves her "natural station", i.e. the situation of the agent, and views herself "at a certain distance" (*TMS* 3.1.1.2). She becomes a spectator that can now only consider the agent's situation by imaginatively entering into it as all spectators must. But through this imaginative exercise, though the spectator may "form some idea of [the agent's] sensations, and even feel something… which is not altogether unlike them", the spectator generally feels something that is "weaker in degree" (*TMS* 1.1.1.2). This weaker sentiment is not a product of the spectator's *impartiality* (Smith has yet to even mention impartiality at this point), but rather of the nature of the sympathetic process, of the difference between actually being in the situation as an agent and imaginatively being in the situation as a spectator. Hence, when we divide ourselves to spectator and agent to judge ourselves, our sentiments are modified: their character is less violent than they were when we reacted as an agent (*TMS* 3.1.1.6). In particular, conscience, like other spectators, lacks "the strongest impulses of self-love" and the "natural misrepresentations of self-love" (*TMS* 3.1.3.4) felt by the agent initially, for like any spectator, conscience is not fully absorbed in the situation of the agent.

While both an agent's conscience and actual spectators are more impartial than that agent, conscience and actual spectators differ in how informed they are of the agent's motives and actions. The distinction plays a central role in Smith's division in *TMS* 3.1.2

between praise and praiseworthiness (i.e. the deservedness of praise). The two are different *in kind*.[86] The former are others' judgments of us (i.e. the approval of actual spectators); the latter, our judgments of how others *ought* to judge us,[87] by which Smith means how others would judge *if they were better informed*:

> When he [the agent] views it in the light in which the impartial spectator would view it, he thoroughly enters into all the motives which influenced it, looks back upon every part of it with pleasure and approbation, and though mankind should never be acquainted with what he has done, he regards himself, not so much according to the light in which they actually regard him, as according to that in which they would regard him if they were better informed (*TMS* 3.1.2.5).

Conscience judges as, we presume, others would *had they been better informed*. In other words, the gap between the praise actually received and the praiseworthiness judged by conscience is a product of conscience's being better informed about the agent's actions or the agent's motives. Since conscience is the agent adopting a certain perspective on her actions and reactions, conscience has access to the agent's motives and actions of which the agent herself is aware. Actual spectators, on the other hand, do not always have this information because spectators are not observing the agent at every moment, and furthermore must *infer* the motives of the agent, e.g. based on facial and verbal expressions. Thus, actual spectators are liable to be "deceived" (*TMS* 3.1.2.4) or "ignorant" (*TMS* 3.1.2.5) in ways that conscience is less susceptible. This is not to say that conscience is not liable to deception or ignorance. We are often unaware of our own motives. We might mistakenly identify the sentiments that are driving our actions. For instance, I might be unaware of the jealousy that I feel towards a

---

[86] "Those two principles, though they resemble one another, though they are connected, and often blended with one another, are yet, in many respects, distinct and independent of one another" (*TMS* 3.1.2.2).

[87] "Praise and blame express what actually are; praise-worthiness and blame-worthiness, what naturally ought to be the sentiments of *other* people with regard to our character and conduct" (*TMS* 3.1.2.25).

colleague and how that jealousy is motivating me to work harder so that I can gain more accolades. I genuinely believe that I am driven by a desire for security so that I can better provide for my family, when in fact that desire is quite weak. However, because conscience does not have this additional barrier of having to infer the thoughts and sentiments of the agent, conscience has *some* epistemic advantage over outside spectators. As a result, conscience becomes the standard by which we measure ourselves, not the praise of actual spectators, for conscience does not issue judgments that are not rooted in ignorance or mistake, but correctly assesses the propriety of the agent:

> It is by no means sufficient that, from ignorance or mistake, esteem and admiration should, in some way or other, be bestowed upon us. If we are conscious that we do not deserve to be so favourably thought of, and that if the truth were known, we should be regarded with every different sentiments, our satisfaction is far from being complete. The man who applauds us either for actions we did not perform, or for motives which had no sort of influence upon our conduct, applauds not us, but another person (*TMS* 3.1.2.4).
>
> If the man without should applaud us, either for actions which we have not performed, or for motives which had no influence upon us; the man within can immediately humble that pride and elevation of mind which such groundless acclamations might otherwise occasion, by telling us, that as we know that we do not deserve them, we render ourselves despicable by accepting them (*TMS* 3.1.2.32).

In these cases, in which praise is not forthcoming from others because they are ignorant of our motives, we justly retain our belief in our praiseworthiness because of the inner praise issued by conscience, for it is the more accurate judgment, grounded on the truth about our motives and actions.

Conscience, then, is not ideal in that it is the perfect judge, one with no partiality and perfect knowledge of the situation such that it only forms correct moral judgments. Rather, it

is ideal in that it is in a privileged position to judge the agent: it is more impartial than the agent before her reflects in virtue of the agent's having adopted the perspective of a spectator at some distance from the agent's situation, and conscience is more well-informed than any actual spectators in virtue of being the agent herself, having direct access to all of the agent's motives and actions without the need to infer them from observations. But such a notion of an ideal spectator leaves open the door for self-deception and uncorrected sentiments still biased by the agent's self-love (discussed in *TMS* 3.1.4), for though conscience may be more impartial than the agent and more informed than actual spectators, it does not necessarily have complete impartiality and full information, and so may still form incorrect moral judgments.

2.8 Assessing Smith on Conscience

In the previous chapter, I presented an overview of the history of 'conscience' and highlighted some key features that any plausible account of conscience must explain:

[1] Conscience at least *seems* to provide moral knowledge, i.e. knowledge of what is good, right, or proper.
[2] Conscience involves judgments of *particulars*, that some particular feeling, intention, or action is good, right, or proper in some particular instance.
[3] Conscience is *not* infallible; we are capable of doubting our own conscience, and admit that our conscience can be mistaken, at least with regard to particular moral judgments.
[4] Conscience is private in that one has direct access only to one's own conscience.
[5] Conscience is like a judge or witness in that it makes us aware of what we have done with respect to some moral standard.
[6] Conscience motivates certain behaviors, specifically through emotions like guilt and shame, as well as tranquility and self-content.

Smith's theory is able to account for all these features. Conscience judges us from a situation that the agent herself believes to be the optimal situation from which to judge. And so conscience always at least appears to issue correct moral judgments, providing us with what seems to be knowledge of what is proper and improper. Those moral judgments are specific to the particular situation of the agent: conscience enters into the particular situation of the agent and judges the agent with regard to that situation only. Conscience does not issue more general judgments of what is or is not proper across a range of situations. Though conscience appears to provide moral knowledge, it is not infallible; conscience can be infected by self-love, the same way our judgments of others can be. Each conscience is specific to the agent, since conscience just *is* the agent judging herself, using all of her own faculties and principles. Yet, because conscience is the agent judging herself from a *different* situation, she appears to be like a spectator distinct from us who judges us and makes us aware of what we have or are about to do. Finally, we (generally) derive pleasure from the agreement of spectators, and displeasure from their disagreement. This sentiment extends to conscience – the most intimate of all spectators for it has access to our innermost motives. We feel pleasure when conscience approves of our actions, and displeasure when it disapproves. Through such feelings, we are then motivated to change our behavior accordingly so that we maintain conscience's approval or avoid its disapproval.

Aside from accounting for these features, Smith's view has some advantages over the alternatives I present in Chapter 1. Medieval philosophers and Hobbes grounded conscience in practical reason: conscience was a function of practical reason by which we derived knowledge of what to do in a particular situation by deriving the right action syllogistically from the first principles of morality (or in the case of Hobbes, definitions). The problem with

this picture of conscience is that it is inconsistent with our phenomenological experience of conscience. The phenomena that Medieval philosophers and Hobbes describes are really more explicit and deliberate forms of moral reasoning. However, conscience is often experienced without such deliberation. Conscience seems to *come upon us*, rather than operate under our deliberate control. Thus, we experience conscience as a watchful eye that we cannot escape, as the feelings of guilt that we cannot shake.[88]

Smith is able to avoid this issue because conscience does not issue moral judgments in such a rigid manner. Conscience is grounded in imagination, not reason.[89] Through imagination, we divide ourselves such that there are now two distinct judgments, one by we the agent and one by we the spectator. This divided self is consistent with our common imagery of conscience as an internal witness or judge, or as the representative of God within us.[90] We as the agent are imagining another being, a spectator, who actually judges us. Yet, unlike actual spectators, and like God, this imagined spectator has access to all of our past and present motives, and our past, present, and intended actions. Of course, Kant also describes conscience as an internal judge: conscience is that internal tribunal in which a judge convicts or acquits us of having acting in accordance or against our duties. However, in Kant's system, conscience only judges of whether or not we have lived up to our duties, and so provides us with information about the consistency between our moral beliefs and our actions. It does not seem to provide us with any knowledge of what to do beyond what we already believe we ought to do. For such knowledge, we need to turn instead to, e.g., the

---

[88] Fuss leveled the same criticism against Broad's account of the cognitive aspect of conscience (Fuss 1964, 114).

[89] Carrasco (2004) attempts to argue that Smith denies conscience as a function of *theoretical* reason, but not *practical reason*. Most commentators, however, take Smith at his word when he explicitly denies conscience as a function of reason in any sense (D. D. Raphael 2007, Darwall 1999, Kleer 1995, Otteson 2002).

[90] Smith refers to conscience as "the demigod within the breast", "those vicegerents of God within us", and "the substitute of the Deity". For more on Smith's religious commitments in *TMS*, see Hill (2001), Evensky (2005), Brown (1992-1993), and Schliesser (2008).

Categorical Imperative. In Smith's system, conscience does check for consistency in our duties (as we believe them to be) and our actions, but does so by forming its own judgment of what is right to do. In other words, conscience provides moral knowledge by issuing a *new* judgment that *corrects* our prior judgment.

Finally, because Smith builds up conscience from more basic faculties of the human mind, he does not have to rely on any theory of a distinct faculty of the mind. Both Butler and Freud relied on a distinct faculty (for Butler, conscience as a distinct judging faculty; for Freud, the superego) to account for conscience. As Langston points out, one of the reasons that conscience seems to have fallen out of favor in contemporary ethics is that doubt arose as to whether such a distinct faculty exists (Langston 2001). Smith, however, grounds his theory in a more basic, better understood set of human capacities: our capacity to feel and act from those feelings, and our capacity to imagine being (feeling and acting) in another person's situation.[91]

[91] Jan Horst Keppler provides an extensive treatment of how Smith builds up our moral faculties from more basic capacities in *Adam Smith and the Economy of Passions* (2010).

3

Cultivating Conscience: The Amiable Virtues

There is nothing more dangerous than the conscience of a bigot.

George Bernard Shaw

Conscience is commonly described as a moral guide: it directs us toward the right motive and action by either commanding or convicting us to perform it. Hence, we often appeal to its dictates in deciding how we ought to feel and act toward others, or in justifying ourselves to others. Yet, few would argue that conscience is infallible in its moral judgments. Bonaventure and Aquinas were well aware that conscience errs, leading us to do what is sinful all the while believing we are obeying God.[92] Mill seemed to think that conscience can serve as a moral guide, but only after it has received proper education and training.[93] Similarly, Kant claimed that though we cannot determine when we experience conscience and so have no duty to obey conscience, we have a duty to develop our conscience so that it correctly presents our moral duties and convicts (or acquits) us accordingly.[94] So there has long been a general recognition that conscience does err, and that there was a need to properly educate it if it is to be a reliable moral guide. The difficulty, however, is in

[92] Bonaventure (1980, 120); Aquinas (1980, 132).
[93] Mill (*U* 3.8-11).
[94] Kant (*MM* 6:401).

providing a plausible account of how to improve our conscience given that we are liable to self-deception: we believe conscience judges correctly when it does not.

Smith recognized that conscience can and does err, and that we need safeguards for protecting ourselves against an erroneous conscience. He was especially worried about how we can correct conscience in the face of self-deceit, which prevents us from realizing that conscience has erred in the first place. Smith's explicit account for how we correct our erroneous conscience is that through experience, we form general rules of what is proper or improper in a given situation, and that such rules can correct for our erroneous conscience.

In this chapter, I argue that though general rules are crucial for the moral development of an agent, the articulation and acceptance of such rules is limited in its ability to combat self-deceit and to correct for an erroneous conscience. General rules help an agent act properly but only in matters of justice; they do not do much to mitigate the errors that conscience makes in other areas of judgment. I suggest an alternative account of how we cultivate conscience: since conscience uses all the same faculties and principles we use to judge others, we can only improve conscience by improving how we judge others generally. We need to develop our capacity to evaluate ourselves *as spectators*, thereby acquiring what Smith calls the 'amiable virtues'. Smith does not focus on this aspect of moral development, and instead focuses on self-command. As a result, Smithean scholarship has focused on how to better feel and do what is proper through the use of conscience, and has largely ignored the question of how to improve conscience. This oversight, I believe, is a serious mistake. Smith places conscience at the core of a virtuous person, but it can only serve the role that Smith gives it if conscience is highly developed. My aim then is to sketch a clearer picture of how

we can cultivate conscience within Smith's theory of moral development, with a focus on the importance of encountering diversity.

3.1 How Does Conscience Err?

An errant conscience is a conscience that fails to form the correct moral judgment of the agent who is judging herself. Conscience might judge the agent to have felt the proper degree of resentment when she did not, or to be blameworthy for an action that was proper. Though Smith does not use the term 'errant conscience', it is clear that he thinks that our moral judgments of ourselves can err:

> In order to pervert the rectitude of our own judgments concerning the propriety of our own conduct, it is not always necessary that the real and impartial spectator should be at a great distance. When he is at hand, when he is present, the violence and injustice of our own selfish passions are sometimes sufficient to induce the man within the breast to make a report very different from what the real circumstances of the case are capable of authorising (*TMS* 3.1.4.1).[95]

However, Smith is less clear about the different ways in which conscience can err. I want to begin by highlighting some primary ways in which conscience errs, ways that are consistent with Smith's moral theory. I am drawing here on the different ways that we err in our judgments of others. Smith recognizes these kinds of errors. Furthermore, it is wholly consistent with his account of conscience that conscience fall prey to the same kinds of errors.

First, we can be uninformed about the agent's situation, either completely unaware or simply mistaken of certain facts of the case. Thus, we come to an incorrect judgment of what we would feel in that situation and, as a result, incorrectly approve or disapprove of the

[95] See Chapter 2 for a detailed account of Smith on conscience's fallibility.

agent. Smith first suggests this possibility in explaining how we sympathize with others. He notes that certain passions and actions, like the "furious behaviour of an angry man," serves to "disgust and provoke us against them" even though we are uninformed of what provoked the agent's anger (*TMS* 1.1.1.7). It is possible that we simply withhold judgment when we are not fully informed of the situation. When we are unsure of the situation, and aware that we are, "[g]eneral lamentations… create rather a curiosity to inquire into his situation" (*TMS* 1.1.1.9). Yet even then, we do not sympathize with the agent, and until we are informed otherwise, "our sympathy with the grief or joy of another…. is always extremely imperfect" (*TMS* 1.1.1.9). Furthermore, though Smith speaks here of our being uninformed, he thinks we can also be mistaken about the situation, e.g. when a person is falsely accused (*TMS* 2.1.5.6). In all these cases, if I know of the details of your situation, I may be more inclined to agree with you since I may be affected by those details in the same way. However, because I am unaware of those details or misinformed about them, I do not affected by those details and so feel quite differently from you. My imperfect sentiment then becomes the basis for my errant moral judgments.

To some degree, conscience is less susceptible to this kind of error because conscience and the agent are one and the same person, so that an agent's conscience has access to all the same information that the agent does. In particular, conscience has access to the agent's motives and action, about which other spectators are often uninformed. Other spectators often praise "from ignorance or mistake" (*TMS* 3.1.2.4), failing to praise even when praise is deserved if they should never be acquainted with what [the agent] has done" (*TMS* 3.1.2.5). Conscience, however, not only praises but also judges the praiseworthiness of the agent, of the praise deserved "if the truth were known" (*TMS* 3.1.2.4), "if [the spectators]

were better informed" (*TMS* 3.1.2.5). Thus, conscience does not seem to suffer from the same problem of false beliefs or ignorance that spectators often face when judging others.[96]

However, even though conscience is better informed of the agent's motives and actions, it may not be better informed about all the other aspects of the situation. Because conscience and the agent are one and the same, conscience only has access to information about the situation that the agent has. Thus, if the agent is uninformed, so is conscience. And the agent can be uninformed about many aspects of the situation, beyond her own motives and actions. For instance, suppose I am furious at your stepping on my foot because I am uninformed of the fact that you had a muscle spasm that resulted in your stepping on my foot. If I remain unaware of this fact as I appeal to conscience, then conscience will judge me to be perfectly proper in feeling resentment towards you, even though it is improper for me to feel resentment given you had no ill-motive. It is not uncommon for agents to be uninformed of some aspect of the situation, and as a result feel and act improperly. And if the agent is uninformed, conscience, which is really just an aspect the agent's mind, will also be uninformed in the same way.

Second, we can form incorrect moral judgments when we fail to account for the inconsistency in our current state and the way our faculties and principles generally operate. Smith, for instance, recognizes that the spectator's current state can influence how the spectator feels upon entering into the agent's situation:

[96] As I stated in the previous chapter, the claim is not that people (and their conscience) are always fully aware and correct about their own motives. We might have hidden, unconscious motives, or might be simply mistaken about which sentiment is motivating our actions. Still, conscience does have one epistemic advantage over outside spectators: it does not need to make inferences to the motives of which the agent is aware. Because conscience does not need to infer what the agent (knowingly) feels, conscience has one less possible source of error.

> We may often approve of a jest, and think the laughter of the company quite just and proper, though we ourselves do not laugh, because, perhaps, we are in a grave humour, or happen to have our attention engaged with other objects. We have learned, however, from experience, what sort of pleasantry is upon most occasions capable of making us laugh, and we observe that this is one of that kind (*TMS* 1.1.3.3).

Smith thinks that when we are attuned to our current state, e.g. "grave humour", then we can adjust our moral judgments accordingly based on our experience of how we have felt in situations like the agent's. However, if we are unaware of our current state, then we cannot account for this abnormality in forming a judgment of others. If I am downtrodden and unaware of it, then when I enter the situation of a person in great spirit, I will not be able to feel nearly the same because my current state prevents me from imaginatively feeling anything so cheerful.

Similarly, conscience can be unaware of its own state and so fail to account for peculiarities that may affect its sentiments when judging others. This occurs because, again, the agent and conscience are one and the same person using the same faculties. Thus, the agent's sentiments often "carry over" over to (or unduly bias) his or her conscience. If the agent is in a grave mood, then conscience is exercising its faculties from this grave mood. If the agent does not adjust for her current mood, then neither will her conscience. The mere change in situation, of an agent to that of a spectator, will not necessarily change a person's mood from grave to balanced. The change in situation might mean that the conscience's sentiments are less pronounced in some respects since sentiments arising from the sympathetic mechanism generally does not rise to the phenomenological level or intensity with which they are felt by the agent with whom one is sympathizing (*TMS* 1.1.1.2). So the

change in situation might mitigate the effects of these peculiarities by dulling the sentiments, but it will not eliminate them altogether.

Third, we form incorrect moral judgments because of the various corrupting influences on our moral sentiments. Smith claims that various natural principles of human nature, if not properly regulated, lead us to approve of what is not proper, and disapprove of what is proper. In other words, these influences corrupt us *as spectators*, as the judge of others and ourselves. For instance, we have a natural disposition to admire those who display great wealth and power, regardless of whether or not they are virtuous. As a result, we come to approve of and follow in the steps of those with wealth and power, even if they lack all the virtues. For the same reasons, we disapprove of those in "poor and mean condition" even if they are virtuous and simply fell into bad fortune (see *TMS* 1.3.3 titled "Of the corruption of moral sentiments, which is occasioned by this disposition to admire the rich and the great, and to despise or neglect persons of poor and mean condition"). Custom and fashion also can corrupt our moral sentiments, though not to the degree that the disposition to admire the wealthy and powerful can.[97] Custom and fashion often compound the degree to which we approve or disapprove of another: "When custom and fashion coincide with the natural principles of right and wrong, they heighten the delicacy of our sentiments, and increase our abhorrence for every thing which approaches to evil" (*TMS* 5.1.2.2). Also, the differences in custom and fashion among the "different professions and states of life" and the "different situations of different ages and countries" lead to wide variance in what is considered proper and what virtues we deem most praiseworthy (*TMS* 5.1.2.4-11). Though Smith claims that

[97] "Since our sentiments concerning beauty of every kind, are so much influenced by custom and fashion, it cannot be expected, that those, concerning the beauty of conduct, should be entirely exempted from the dominion of those principles. Their influence here, however, seems to be much less than it is every where else" (*TMS* 5.1.2.1; also see *TMS* 5.1.2.12).

our moral sentiments are not "very grossly perverted" by custom and fashion, he does admit that they can and are corrupted to some degree by custom and fashion (*TMS* 5.1.2.13).[98]

The question is whether conscience is immune to these corrupting influences, or at least better at tempering these dispositions so that the conscience of the wealthy and powerful do not approve of their actions beyond what they deserve (nor the consciences of the poor and dispossessed unduly disapprove of their own actions) and our consciences do not (in general) approve or disapprove of sentiments and actions as a product of custom or fashion. Smith does not take up this issue directly, but his account of these corrupting influences suggests that conscience does not necessarily mitigate their effects. For Smith, conscience corrects for our moral sentiments by allowing us to view ourselves as an impartial spectator. Smith also claims that these corrupting influences are present in impartial spectators. For instance, if two people with the same intent and action produce different results, we would judge the person who confers greater benefit on us more meritous, even though they are not deserving of greater merit. More importantly, our judging as impartial spectators does not diminish the corrupting influence of fortune: "[t]his consideration, they imagine, should [be so] *even in the eyes of the impartial spectator*" (*TMS* 2.3.2.2).[99] Similarly, Smith claims that custom affects how we judge the degree to which a particular quality is "*blamable or praise-worthy*" (*TMS* 5.1.2.7). For Smith, *conscience*, as opposed to actual spectators, is that which

[98] Smith claims that we need to account for the differences in "rank and profession" when we judge the propriety of others: "when we bring home to ourselves their situation in this particular respect, we must be sensible, that every occurrence should naturally affect them more or less, according as the emotion which it excites, coincides or disagrees with the fixt habit and temper of their minds" (*TMS* 5.1.2.5). For instance, given the role of a clergy person, we "cannot expect the same sensibility to the gay pleasures and amusements… which we lay our account with in an officer" (*TMS* 5.1.2.5). It might seem as if he is claiming that there is no impropriety arising from the differences in custom and fashion. However, Smith is not making such a claim. He is pointing out that though "custom has allotted" certain manners for certain professions, the propriety of those manners for those engaged in those professions are "independent of custom" (*TMS* 5.1.2.5).

[99] Though Smith claims that this irregularity of our sentiments has some benefits (*TMS* 2.3.3.3), he claims that such irregularity is nonetheless "unjust" (*TMS* 2.3.2.2), and that the proper sentiment is the sentiment that a spectator would feel if it were "either altogether candid and equitable, or even perfectly consistent with themselves" (*TMS* 2.3.3.6).

forms judgments of blameworthiness and praiseworthiness (*TMS* 3.1.2). So though Smith does not focus on the effects of these corrupting influences on conscience (and instead on the effects of excessive self-love), he does seem to acknowledge that conscience is not immune to these corrupting influences.

Fourth, we form incorrect moral judgments by being partial: we fail to account for the particular relation that we have with the agent. The way we judge is influenced by the relationship we have with the agent. The closer we are to that agent, the more likely we are to agree with her sentiments. Parents generally do not judge their children improper because they care much more about their children than strangers, and so tend to side with their children and agree with their sentiments, even though they would not do the like for strangers (*TMS* 3.1.3.22). Similarly, friends judge fellow friends more favorably than they would a stranger because they care more deeply about friends than strangers, though not to the degree that a parent does (*TMS* 1.1.4.9). These biases are not born of any fact about the situation and the agent's response to it. Rather, these biases are solely a product of the relationship between the spectator and agent: the closer the relationship to the agent, the more likely it is that the spectator agrees with the agent, regardless of the propriety of the agent's sentiments. If the spectator does not account for the particular relation to the agent, then the spectator is susceptible to errant moral judgments. This is not to say that partial spectators will always make an errant moral judgment. For example, we can correctly judge the propriety of the actions of our enemies, against whom we are biased (*TMS* 2.3.1.5). However, these natural sources of bias do mean that spectators will need to adjust their moral judgments in recognition of the partial relation that they bear to certain agents.

Though conscience is aimed at judging as an impartial spectator, who presumably has no relation to the agent, conscience is particularly vulnerable to errors from partiality because it is directed at the very agent whose conscience it is. Hence, conscience is vulnerable to "caring" more about the agent than other people because conscience shares the closest possible relationship to to the agent: a relationship with itself. The danger, then, is the same danger that agents face in assessing what is proper to the situation. Agents feel an overly strong self-love, or preference for themselves, when they are affected by the situation. Similarly, though judgments of conscience are the products of an agent's attempt to adopt the perspective of a spectator on her own actions, judgments of conscience are still products of that agent's mind, and are therefore affected by a greater degree of self-love than the judgments of a numerically distinct spectator. This partiality presents itself in two primary ways. For one, when we consult conscience prior to acting so that we can determine what is proper to do, "the eagerness of [the agent's] passion will seldom allow us to consider what we are doing, with the candour of an indifferent person" (*TMS* 3.1.4.3). It is difficult for conscience to fully disassociate itself from the agent. So conscience is often in danger of adopting the agent's sentiments, feeling the same sentiment by becoming affected by the situation in all the same ways as the agent. Thus, though conscience can disagree with the agent, it is also vulnerable to agreeing with the agent's more immediate (less reflective) reactions to her situation and approving of her more often than do the evaluations of others, whether or not the agent is proper.[100]

[100] This form of partiality may seem similar to the previous kind of error. In both cases, the agent's sentiments are transferred over to conscience, which affects how conscience judges. However, there is a key difference. When conscience does not account for the agent's unusual mood, it fails to address an influence that is independent of the agent's relationship with the spectator, e.g. the depressive episode that the agent undergoes alone. However, when conscience does not account for partiality, it fails to address the influence of the *relationship* itself.

Conscience is also vulnerable to error because when the agent appeals to conscience to judge some prior action of hers, conscience has difficulty disapproving of that action because doing so would mean disapproving of itself. Unlike the prior case, the agent's sentiments does not infect conscience's judgments. Once the action is completed, the agent's sentiments subside. Yet even then, "[i]t is seldom… that they are quite candid" because "[i]t is so disagreeable to think ill of ourselves, that we often purposely turn away our view from those circumstances which might render that judgment unfavourable" (*TMS* 3.1.4.4).[101] A strong sense of self-love does not impose itself on conscience as a result of its sharing its faculties with the agent. Rather, conscience strongly favors the agent, just as a mother favors her child over other children, or just as a person favors a friend over a stranger. We wish to look at those closest to us in a more favorable light, and so conscience distorts or ignores various aspects of the situation, including the motives and actions of the parties involved, to come to a more favorable judgment of the agent. This is not to say that conscience cannot set aside this bias to judge as a more impartial spectator. However, the relation between conscience and agent leaves conscience vulnerable to such bias, which, left unchecked, results in an errant moral judgment.

This partiality that conscience has for the agent, both before and after the agent commits the act, constitutes the essence of self-deception. For Smith, self-deception is not merely conscience's unknowingly erring in its moral judgment, which is the case when conscience is simply misinformed about the situation. Instead, self-deception requires that conscience distorts its sentiments, either distorting (or ignoring) various aspects of the

[101] Smith repeats these points in *TMS* 6.3.1.22: "The principle of self-estimation may be too high, and it may likewise be too low. It is so very agreeable to think highly, and so very disagreeable to think meanly of ourselves, that, to the person himself, it cannot be well doubted, but that some degree of excess must be much less disagreeable than any degree of defect."

situation or igniting old passions to form a more favorable moral judgment of the agent. Conscience "turn[s] away our view from those circumstances which might render that judgment unfavourable (*TMS* 3.1.4.4). Or it views the situation the way the agent does, "where every thing appears magnified and misrepresented by self-love" (*TMS* 3.1.4.3). Conscience might also "foolishly and weakly, endeavour to exasperate anew those unjust passions which had formerly mislead us", to "awaken our old hatreds, and irritate afresh our almost forgotten resentments" (*TMS* 3.1.4.4). Self-deception, in other words, is conscience's failure in being honest with itself and the agent. By allowing self-love to infect its judgments through these distortions, conscience is not judging with the "candour" of an impartial spectator, by "pull[ing] off the mysterious veil of self-delusion, which covers from his view the deformities of his own conduct" (*TMS* 3.1.4.4). We lie to ourselves about the motives and actions of others and ourselves, facts of the case, and the consequences of our actions. Moreover, conscience genuinely believes itself to be judging as an impartial spectator, unaware that it is being dishonest in this way since "[t]he passions… all justify themselves, and seem reasonable and proportioned to their objects, as long as we continue to feel them" (*TMS* 3.1.4.3). So even though conscience may purposely distort its sentiments or the situation, it does not believe itself to be doing so, and instead believe that it judges as an impartial spectator.[102]

[102] There is much more to be discussed regarding self-deceit in Smith's theory. I limited my discussion here to how self-deception serves as a barrier to correcting our own sentiments. For more detailed discussions of Smith on self-deception, see Hanley (2009), Mitchell (1987), and Schliesser (2003). For a helpful contrast between Smith and Butler on self-deception, see Martin (1986).

3.2 Correcting an Erroneous Conscience: The Explicit Account

Immediately after introducing the problem of self-deceit, Smith gives an account of how we overcome it. Smith's explicit account is that from our repeated observations of our own moral judgments of others, we find that "all actions of a certain kind, or circumstanced in a certain manner, are approved or disapproved of" and from those observations, we are thusly able to form general rules about what is proper in certain situations (*TMS* 3.1.4.8). These general rules are initially a product of "*[o]ur* continual observations upon the conducts of *others*" (*TMS* 3.1.4.7, emphases added). But those rules are further confirmed when we hear "every body about us express the like detestation" (*TMS* 3.1.4.7). Once these general rules are "universally acknowledged and established, by the concurring sentiments of mankind", these rules become the standard for judging one another (*TMS* 3.1.4.11). But more important for our purposes, it is only when these rules become "fixed in our mind by habitual reflection" that they become effective in correcting our erroneous conscience, in correcting the "misrepresentations of self-love concerning what is fit and proper to be done in our particular situation" (*TMS* 3.1.4.12). To be "fixed in our mind" is for us to develop the "awe and respect with which he has been accustomed to regard it" (*TMS* 3.1.4.12), "what is properly called a sense of duty" (*TMS* 3.1.5.1). It is only once we develop this sense of duty that the general rules effectively correct for our erroneous conscience, because it is only then that we feel the "agonies of doubt and uncertainty", "terror", and the "fury of his desires to violate it" "at the thought of violating so sacred a rule" (*TMS* 3.1.4.12). Hence, "[w]ithout this sacred regard to general rules, there is no man whose conduct can be much depended on" (*TMS* 3.1.5.2).[103]

[103] For an excellent in-depth discussion of the role of general rules in combatting self-deceit, particularly on matters of justice, see Khalil (2009) and Haakonssen (1981).

I want to emphasize that though our sense of duty is confirmed and sometimes heightened by agreement with others in our judgments of obligation, unless *we* adopt these general rules as reflections of *our* better judgment, then those general rules do not have any effect in correcting our erroneous conscience. It is because *we* formed these rules by reflecting on *our* moral judgments that we come to respect them. The mere fact that others consistently make moral judgments of a certain kind in certain circumstances does not result in the kind of awe and respect for general rules that Smith thinks is central to our sense of duty. If it is to be our sense of duty, and not the mere fear of the disapproval or punishment of others, that is to motivate us to correct our conscience, then we must find the general rules to be a reflection of *our own* moral judgments. Hence, Smith writes:

> At the very time of acting, at the moment in which passion mounts the highest, he hesitates and trembles at the thought of what he is about to do: he is secretly conscious to himself that he is breaking through those measure of conduct which, in all *his* cool hours, *he* had resolved never to infringe, which *he* had never seen infringed on by others without the highest disapprobation, and of which the infringement, *his own* mind forbodes, must soon render him the object of the same disagreeable sentiments (*TMS* 3.1.4.12, emphases added).

General rules, at least the ones we apply to correct our conscience, are rules formed from *our own judgments*, not from the judgments of others.

Our own moral judgments are, of course, heavily influenced by those of others.[104] Through the years, how we judge others might be shaped by the judgments others make of us and of third parties. When others agree with our moral judgments, then we are more confident of our moral judgments; when others disagree, we sometimes change our initial moral judgments to better align with those of others. But unless we endorse their judgments,

[104] Even this might be an understatement given the kind of influence that parents, friends, and educators have on our moral development.

unless *we* judge the same, then we do not feel reverence or awe for the rules that might be extracted from *their* judgments.

Finally, the respect and awe that we have for general rules can be "further enhanced by an opinion… that those important rules of morality are the commands and laws of the Deity, who will finally reward the obedient, and punish the transgressors of their duty" (*TMS* 3.1.5.3). In regarding these general rules as the laws of the Deity, "they necessarily acquire a new sacredness" (*TMS* 3.1.5.12):

> How vain, how absurd would it be for man, either to oppose or to neglect the commands that were laid upon him by Infinite Wisdom, and Infinite Power! How unnatural, how impiously ungrateful not to reverence the precepts that were prescribed to him by the infinite goodness of his Creator, even though no punishment was to follow their violation (*TMS* 3.1.5.12).

Aside from the heavy rhetoric, his point is that the observing of general rules is a product of the esteem with which we hold those rules. At first, we do so because of our own commitments to those moral judgments, for we are the ones who made or endorsed those moral judgments.[105] This is what Smith calls our "natural sense of duty". It then becomes reinforced by the agreement with others. And finally, we gain an even greater sense of reverence and awe for those rules when we accept them as the laws of God, as not simply general rules that I lay down for myself, or those that others lay down for themselves, but what the Almighty lays down for all. This enhanced reverence for the general rules provides

[105] It is possible, or even likely, that we endorse general rules without building them up from our own moral judgments. For instance, parents often teach children these general rules without any (explicit) rationale, and children view those rules with respect. In such cases, children are deferring their judgments to their parents. Smith does not discuss such instances in *TMS*, but his overall point still holds true: our "natural sense of duty" is grounded in our agreeing with and giving esteem to the general rules.

us with an even stronger motive for acting in accordance with those rules, even when our conscience initially judges it proper to do otherwise.

3.3 Limits of General Rules

Though Smith explicitly appeals to general rules as the means for correcting erroneous conscience, I think it would be a grave mistake to think that the adoption of general rules is the only means, or even the primary or best means, by which we correct conscience. First of all, general rules are formed inductively from our particular moral judgments, and so are vulnerable to error as with any other inductive reasoning. All that is required for forming general rules is consistency in our moral judgments, and there may very well be consistency in erroneous judgments. Now, Smith does not think there will be errors of the sort with regard to the "general style of conduct or behaviour" (*TMS* 5.1.2.14) for if certain basic rules of justice are not in place and followed to a substantial degree, then a society would simply collapse (e.g. *TMS* 2.2.3.6, 5.1.2.16). But that does not mean that people are somehow immune to error when it comes to general rules. Smith seems to leave open the door for erroneous general rules when he admits that we can, through custom, have laws, a kind of general rule, that may have been reliable for determining proper action in the past, but are not so reliable today. For instance, infanticide was practiced in the "earliest period of society" because during those times, a person is *frequently* unable to provide support for both herself and her child. And in such cases, it is proper to abandon the child, and to do so customarily since such instances were so frequent. But in "all the states of Greece, even among the polite and civilized Athenians", such a custom, such a rule of action, would be perfectly improper, for these states are not faced with such frequent hunger (*TMS*

5.1.2.15). Any general rules allowing for infanticide in these states would be "destructive of good morals", capable of "establishing, as lawful and blameless, particular actions, which shock the plainest principles of right and wrong" (*TMS* 5.1.2.14). So though it seems unlikely that the most general of the principles of action will turn out to be erroneous, Smith recognizes that we can make errors in the continued adoption of at least some general rules of action that have persisted despite a change in context that renders them maladaptive.

Secondly, even if we were to form correct general rules, these rules are limited in two ways: we can only form accurate general rules regarding matters of justice, and the general rules that combat self-deceit only apply to conduct, not motives.[106] Let me address each limitation in turn. We form a wide range of different moral judgments. However, only judgments of what is demeritous, i.e. rules of justice, have the constancy required for the formation of applicable general rules. All other moral judgments, i.e. judgments of propriety, impropriety, and merit, vary too much for us to make any inductive inference:

> The general rules of almost all the virtues, the general rules which determine what are the offices of prudence, of charity, of generosity, of gratitude, of friendship, are in many respects loose and inaccurate, admit of many exceptions, and require so many modifications, that it is scarce possible to regulate our conduct by a regard to them…. There is, however, one virtue of which the general rules determine with the greatest exactness every external action which it requires. This virtue is justice. The rules of justice are accurate in the highest degree, and admit of no exceptions or modifications, but such as may be ascertained as accurately as the rules themselves, and which generally, indeed flow from the very same principles with them (*TMS* 3.1.6.9-10).[107]

[106] I am expanding on Griswold's claim that general rules do not serve as a substitute for moral judgment and so there is a need for moral education, which "will teach the observer how to judge" (Griswold 1999, 214).
[107] Smith repeats the same points by drawing an analogy to the written language: "The rules of justice may be comparable to the rules of grammar; the rules of the other virtues, to the rules which critics lay down for the attainment of what is sublime and elegant in composition. The one, are precise, accurate, and indispensible. The other, are loose, vague, and indeterminate, and present us rather with a general idea of the perfection we ought to aim at, than afford us any certain and infallible directions for acquiring it" (*TMS* 3.1.6.11).

There is enough constancy in our judgments of, say, murder and theft that we can lay down general rules for such cases (that it is demeritous to murder under any, or most, circumstances; that it is demeritous to steal under most circumstances). But for most other matters, like what is proper to give to charity, or what actions are deserving of reward, we have difficulty forming general rules with much precision because our judgments in such cases vary more from situation to situation. There is far too much variance in what is proper to do from one situation to the next, such that any general rules that we can form will still "admit of ten thousand exceptions" (*TMS* 3.1.6.9). Thus, these rules, if we are able to form any such rules, will not provide much guidance on what is proper to do:

> If your benefactor attended you in your sickness, ought you to attend him in his? or can you fulfill the obligation of gratitude, by making a return of a different kind? If you ought to attend him, how long ought you to attend him? The same time which he attended you, or longer, and how much longer? If your friend lent you money in your distress, ought you to lend him money in his? How much ought you to lend him? When ought you to lend him? Now or to-morrow, or next month? (*TMS* 3.1.6.9).

Loose and inaccurate rules are not of much help in answering these questions, so those rules, Smith claims, should not be given much regard: "To affect, however, a very strict and literal adherence to them would evidently be the most absurd and ridiculous pedantry" (*TMS* 3.1.6.9; also see *TMS* 3.1.6.8). We are much better off assessing each case on its own merit, rather than on a general rule.

Additionally, the general rules that Smith claims corrects for our erroneous conscience only direct our conduct, not our motives or sentiments. In his discussion of the general rules that correct for self-deceit (*TMS* 3.1.4), Smith qualifies every instance of 'general rules' with 'conduct', 'what to do', and 'action', for instance:

> Our continual observations upon the *conduct* of others, insensibly lead us to form to ourselves certain general rules concerning what is fit and proper either *to be done or to be avoided* (*TMS* 3.1.4.7, emphasis added).

> The general rule, on the contrary, is formed, by finding from experience, that all *actions* of a certain kind, or circumstanced in a certain manner, are approved or disapproved of (*TMS* 3.1.4.8, emphasis added).

> Those general rules of *conduct*, when they have been fixed in our mind by habitual reflection, are of great use in correcting the misrepresentations of self-love concerning what is fit and proper *to be done* in our particular situation (*TMS* 3.1.4.12, emphasis added).

These general rules, in other words, only extend to what we ought to do. They override any sentiment we have by commanding us to act in some way without any regard for how we ought to feel or what our motive ought to be.

Smith makes the same qualifications when describing the sense of duty. This sense is not merely a feeling of obligation born of these general rules that correct conscience; it is a feeling of obligation *to act* even when we have strong motives to act otherwise: "The regard to those general rules of *conduct*, is what is properly called a sense of duty, a principle of the greatest consequence in human life, and the only principle by which the bulk of mankind are capable of directing their *actions* (*TMS* 3.1.5.1, emphases added). And later, he stresses that the importance of this sense of duty is to direct one's *actions*, not one's sentiments: "That habitual reverence which your former experience has taught you for these, enables you to *act*, upon all such occasions, with nearly equal propriety, and hinders those inequalities of temper, to which all men are subject, from influencing your *conduct* in any very sensible degree" (*TMS* 3.1.5.2, emphasis added). The sense of duty, then, is a respect for certain general rules of *conduct*, rules that direct us to *act* in certain ways even when we have

motives to do otherwise. The sense of duty does not direct us to *change* our sentiments in accordance with some general rule. In other words, the sense of duty is not some application of a general rule regarding how we are to feel, such that we are obeying these rules when we change our sentiments to fit the sentiment determined appropriate by the general rule. Rather, the sense of duty serves as a *competing motive* to combat the influence of other motives in directing action; it is one of principle of action among many that influence an agent's actions. Sometimes, acting solely from the sense of duty is proper; other times, doing so is not, and instead requires that the sense of duty is accompanied by other motives, like benevolence (*TMS* 3.1.6.2). The sense of duty is itself a sentiment that competes or harmonizes with other sentiments to bring about some action, not a sentiment of obligation to have some other sentiment.

That general rules are only of conduct is a particularly worrisome when we are focusing on correcting moral judgments because Smith thinks that moral judgments are judgments about the principles of action, the motives or sentiments that bring about action, and that this is so even when we are making judgments of merit and demerit. Hence, he writes that even when someone performs an action that benefits (or harms others), we do not think gratitude (or resentment) is proper unless we also find the agent's sentiments proper:

> It is to be observed, however, that, how beneficial soever on the one hand, or how hurtful soever on the other, the actions or intentions of the person who acts may have been to the person who is, if I may say so, acted upon, yet if in the one case there appears to have been no propriety in the *motives* of the agent, if we cannot enter into the *affections* which influenced his conduct, we have little sympathy with the gratitude of the person who receives the benefit (*TMS* 2.1.3.1, emphases added).

Smith's claims here are consistent with his aim in *TMS*, which is not to produce rule-followers, but rather to cultivate virtues (*TMS* 7.3.Intro.3). The virtuous person does the proper action, but does so with the proper sentiments. The proper motive may at times strictly be the reverence for general rules. But in most cases, the reverence for general rules alone will be insufficient for proper motive. Smith is explicit about this in *TMS* 3.1.6, titled "In what cases the sense of duty ought to be the sole principle of our conduct; and in what cases it ought to concur with other motives". In that chapter, Smith describes two considerations that determine when acting solely from a sense of duty would be proper:

> [F]irst, upon the natural agreeableness or deformity of the sentiment or affection which would prompt us to any action independent of all regard to general rules; and, secondly, upon the precision and exactness, or the looseness and inaccuracy, of the general rules themselves (*TMS* 3.1.6.2).

As for the first consideration, the type of sentiment affects whether or not the sense of duty alone is the proper motive. The actions that generally arise out of benevolence, like giving money to those in dire need, must be accompanied by the benevolent affections as much as the sense of duty (*TMS* 3.1.6.4). In contrast, the actions that generally arise from the "malevolent and unsocial passions", like revenge, need not, and in many cases *should* not, be accompanied by such passions, but instead strictly from a sense of duty (*TMS* 3.1.6.5). As for the second consideration, Smith claims that the more loose and inaccurate the general rule is, the less proper it is to act solely from a felt obligation to obey the rule. Such loose and inaccurate rules are those of *"almost all of the virtues"* (*TMS* 3.1.6.9, emphases added). The only exception is the rules of justice (*TMS* 3.1.6.10-11). So even if general rules are able to correct our actions, and may even provide an additional motive, i.e. the sense of duty, they

cannot correct our sentiments nor can they correct judgments in matters other than justice—judgments that depend on identifying the proper motive or response to a given situation.

Of course, Smith does not deny that we can appeal to general rules to form judgments of propriety, judgments of what we ought to feel. For instance, he writes: "the general rules derived from our preceding experience of what our *sentiments* would commonly correspond with, correct upon this, as upon many other occasions, the impropriety of our present emotions" (*TMS* 1.1.3.4; also see *TMS* 2.2.3.10). He mentions these general rules for our sentiments again in Part III: "those general rules which our moral faculties observe in approving or condemning whatever sentiment or action is subjected to their examination" (*TMS* 3.1.5.6). But in this instance, he is making a point about some feature of all general rules of morality: these general rules are *prescriptions* in the way the rules of a sovereign are, so they can be properly called *laws*. His point is not about the correction of self-deceit.[108]

Given these restrictions on general rules, general rules seem to only be able to correct for conscience in a small range of cases. Conscience issues a wide range of different moral judgments. Conscience not only judges the propriety of our actions that may harm others, like physical assault; it also judges the propriety of our actions that benefit others, like how much we ought to give to charity. Conscience not only judges what we ought to do; it also judges how we ought to feel in doing those actions. General rules will not be able to correct conscience in any of these instances.[109]

---

[108] Similarly, in *TMS* 3.1.4.8, Smith leaves open his discussion of general rules to include both.

[109] Another possible problem with general rules as the means for correcting conscience is that general rules may be susceptible to the same self-deceit that conscience is. Some argue that Smith's account allows for general rules, once formed, to be used in the sympathetic process to judge others and ourselves (e.g. Griswold 1999, Carrasco 2004). We use the sympathetic mechanism to enter into the situation of the agent, and then consider how we would feel and act from that situation given that we have the general rules that we do. If conscience does use general rules, then the moral judgments we form through general rules are apt to self-deceit, since conscience, biased by strong self-love, can misapply such rules while the agent believes conscience to have judged correctly.

3.4 Correcting Conscience by Correcting Our Moral Judgments of Others

In the previous chapter, I argued that we judge ourselves using all the same faculties and principles we use to judge others. If this interpretation is correct, then we have another means for correcting conscience: *we can correct how we judge others*. Because we judge ourselves using the same principles we use to judge others, if we do not judge others correctly, then we will fail to judge ourselves correctly. Likewise, if we judge others correctly, then we will judge ourselves correctly. The strategy for correcting conscience then is to develop as spectators, cultivating what Smith calls the 'amiable virtues', which is the capacity to enter more fully into the situation of the agent so that we can form a more accurate judgment of what is proper in her situation.

Smith himself does not appeal to this process of correcting our judgments of others as a means for correcting an erroneous conscience. However, this process is embedded into Smith's account of moral development. Smith draws a distinction between how we develop as an agent and how we develop as a spectator. Given that conscience is us as *spectators*, not agents, we need to look to his account of the development of spectators to address how we are to improve conscience.

Smith first points to this distinction between agents and spectators in describing how we resolve disagreements about what is proper:

I think that this approach is mistaken. Though Smith is not explicit on this point, he seems to treat general rules as one of four distinct processes for forming moral judgments: "When we approve of any character or action, the sentiments which we feel, are, according to the foregoing system, derived from four sources, which are in some respects different from one another" (*TMS* 7.3.3.16). The first two involve the sympathetic mechanism, the third is general rules, and the fourth is the beauty found in utility. The suggestion is that the way in which general rules form moral judgments is one that does not involve the sympathetic mechanism. It may be that our sense of duty, which is a sentiment, can be considered in using the sympathetic mechanism. However, general rules themselves are not sentiments, but rather principles from which we can derive more particular judgments through reasoning, not the imaginative process.

> In all such cases, that there may be some correspondence of sentiments between the spectator and the persons principally concerned, the spectator must, first of all, endeavour, as much as he can, to put himself in the situation of the other, and to bring home to himself every little circumstance of distress which can possibly occur to the sufferer. He must adopt the whole case of his companion with all its minutest incidents; and strive to render as perfect as possible, that imaginary change of situation upon which his sympathy is founded (*TMS* 1.1.4.6).

> [A]s nature teaches the spectators to assume the circumstances of the person principally concerned, so she teaches this last in some measure to assume those of the spectators. As they are continually placing themselves in his situation, and thence conceiving emotions similar to what he feels; so he is as constantly placing himself in theirs, and thence conceiving some degree of that coolness about his own fortune, with which he is sensible that they will view it (*TMS* 1.1.4.8).

There are a couple points I want to highlight about this process. First of all, neither the spectator nor the agent is necessarily judging correctly about what is proper. The proper sentiment is often somewhere between what the agent feels and what the spectator feels:

> Two different sets of philosophers have attempted to teach us this hardest of all lessons of morality. One set have labored to increases our sensibility to the interests of others; another, to diminish that to our own. The first would have us feel for others as we naturally feel for ourselves. The second would have us feel for ourselves as we naturally feel for others. Both, perhaps, have carried their doctrines a good deal beyond the just standard of nature and propriety (*TMS* 3.1.3.8).[110]

Smith does not favor the spectator's sentiments over the agent's, nor the agent's over the spectator's. The standard of propriety is to be found somewhere in between the two, in a

[110] Smith is reiterating a point he made earlier in *TMS* in which he stresses that propriety is found neither in the excessive self-love that agents tend to feel, nor the deficient benevolence that spectators tend to feel: "And hence it is, that to feel much for others and little for ourselves, that to restrain our selfish, and to indulge our benevolent affections, constitutes the perfection of human nature; and can alone produce among mankind that harmony of sentiments and passions in which consists their whole grace and propriety. As to love our neighbour as we love ourselves is the great law of Christianity, so it is the great precept of nature to love ourselves only as we love our neighbour, or what comes to the same thing, as our neighbour is capable of loving us" (TMS 1.1.5.5).

moderate sentiment that both the agent and spectator can agree to.[111] We come to that standard through this process of both the spectator and agent adjusting their sentiments through the imaginative process. So it would be a mistake to think that only the agent comes to revise her errant moral judgment. Both the spectator and agent are correcting for their errant moral judgments through this interaction.

Second, though both the spectator and agent err in that they fail to feel what is proper to the situation, their errors are different in an important respect. The agent errs by failing to control her sentiments to what is proper to that situation. The spectator, on the other hand, errs by failing to fully account for and respond to all the details of the situation. Whereas the agent's error is an error in self-control, the spectator's error is an error in imagination. When the spectator enters into the situation of an agent, the spectator needs to imagine being in the same situation as the agent to come to the correct moral judgment. Doing so requires the spectator to hold the correct account of the situation, like the intent and actions of the people involved. It also requires the spectator to engage with the situation as the spectator actually would in such a situation. In other words, the spectator needs to fully immerse herself into this imaginative exercise, not only being aware of and correct about all the different situational factors, but also giving those factors due consideration in being affected by those factors to the proper degree.[112] If the spectator fails in one of these respects, then the spectator will feel differently than she would if she were in such a situation.

---

[111] For an excellent discussion of the similarities between Smith and Aristotle on the doctrine of the means, see Broadie's "Aristotle, Adam Smith and the Virtue of Propriety" (2010). Others have made similar points (see Berns 1994, Griswold 1999, Vivenza 2001, and Montes 2004.

Though Smith claims that propriety is often found somewhere between the sentiment of the agent and of the spectator, he does not, as Vincent Hope argues (Hope 1989), think that propriety is found in the consensus of spectators, evident in Smith's discussion of the virtues as determined not by a consensus, but rather the few virtuous people who judge very differently from the general public (*TMS* 1.1.4 and 3.1.2).

[112] I want to reiterate that to immerse herself in an agent's situation in this way does not entail that the spectator takes on the agent's dispositions. The spectator continues to exercise her own faculties and principles in

Finally, because of the difference in how they err, the spectator and agent correct their moral judgments differently. The agent is correcting for her initial moral judgment by entering into the situation of the spectator *for the first time*. The spectator, on the other hand, has already entered into the situation of the agent to form her initial moral judgment. So to correct the errant initial judgment, the spectator needs to enter into the situation of the agent *more fully*, accounting for "every little circumstance of distress", "adopt[ing] the whole case of his companion with all its minutest incidents", and "striv[ing] to render as perfect as possible, that imaginary change of situation". The spectator simply has no other recourse, other than general rules, for correcting her initial moral judgment because there is no other way for the spectator to form a moral judgment.

This difference between how the agent and spectator correct for their moral judgments is also reflected in the different virtues:

> Upon these two different efforts, upon that of the spectator to enter into the sentiments of the person principally concerned, and upon that of the person principally concerned, to bring down his emotions to what the spectator can go along with, are founded on two different sets of virtues. The soft, the gentle, the amiable virtues, the virtues of candid condescension and indulgent humanity, are founded upon the one: the great, the awful and respectable, the virtues of self-denial, of self-government, of that command of the passions which subjects all the movements of our nature to what our own dignity and honour, and the propriety of our own conduct require, take their origin from the other (*TMS* 1.1.5.1).

The "amiable virtues" are virtues of spectators; they involve the spectators' capacities to enter fully into the situation of an agent and feel closer to what the agent feels.[113] The "awful

---

forming a judgment of what is proper to feel and do in the given situation. The immersion simply allows the spectator to more accurately imagine her own response to the given situation.

[113] An alternative interpretation of the "amiable virtues" is that these virtues consist in the disposition of spectators to feel as the agent does: "How amiable does he appear to be, whose sympathetic heart seems to re-echo all the sentiments of those with whom he converses, who grieves for their calamities, who resents their

and respectable" virtues, on the other hand, are virtues of agents; they involve the agents' capacities for controlling their sentiments so that those sentiments are of a degree and kind to which impartial spectators find agreeable. Again, Smith does not favor one virtue over the other in attaining perfect propriety. Instead, he claims that the "perfection of human nature", the "harmony of sentiments and passions in which consists their whole grace and propriety", is found where these two virtues meet. The agent arrives at this correct moral judgment by "restrain[ing] our selfish" passions, and the spectator does so by "indulg[ing] our benevolent affections" (*TMS* 1.1.5.5).[114]

To be clear, the aim is not to simply come to an agreement between the spectator and agent, nor is the proper sentiment always found somewhere between the spectator and agent. It may be that the spectator is a virtuous person who meets the standard of "complete propriety and perfection" (*TMS* 1.1.5.9), and so feels the proper sentiment. Thus, only the

---

injuries, and who rejoices at their good fortune!" (TMS 1.1.5.2). However, this interpretation is inconsistent with Smith's views on virtues, which are dispositions that lead to proper sentiments. Agents generally do not feel the proper sentiment, and so if a spectator was disposed to feel as the agent feels, then the spectator would not feel the proper sentiment as well. Such is the case whenever a partial spectator sympathizes with the agent, like when parents sympathize with their children (TMS 3.1.3.13; 3.1.3.22) and when friends sympathize with friends (TMS 1.2.2.1).

[114] It seems possible under Smith's theory that the spectator can correct his own sentiment by appealing to a more impartial spectator (as an agent might), rather than enter more fully into the situation of the agent. We might, for instance, appeal to a third party with no vested interest in matters of international conflict (Smith opens the door for such a possibility when he points out that our bias tends to be toward neighboring nations, not to ones more distant; see *TMS* 6.2.2.5). Alternatively, we might appeal to a virtuous person who displays universal benevolence (*TMS* 6.2.3.3). Similarly, the agent can correct her own sentiment by entering more fully into the situation of the spectator, and taking into deeper consideration the situation of that spectator (Smith does not discuss this possibility explicitly, but his theory does allow for such imaginative exercises).

These alternative means of correcting for our sentiments seems to leave open the possibility for agents to correct errant sentiments by developing the amiable virtues, and for spectators to do so by developing the virtue of self-command. Thus, it seems that there is no virtue that is particular to the role of the spectator or the agent, as I have suggested. However, drawing such a conclusion is, I think, misguided. Smith's point is not that the agent can only correct for her own sentiments by viewing herself from the situation of the spectator, or that the spectator can his own by entering more fully into the agent's situation. Rather, these different virtues reflect the different barriers that agents and spectators tend to face in coming to the proper sentiment. The agent's primary, though not only, barrier to proper sentiment is excessive self-love, which generally heightens her sentiments and so requires her to have greater control her sentiments. The spectator's primary, though not only, barrier is his lack of awareness or understanding of the agent's situation, which generally dampens his sentiments and so requires him to engage more fully with the agent's situation. So there are virtues that seem better suited for agents, and those that seem better suited for spectators.

agent may need a correction of sentiment (or vice versa). Alternatively, the spectator might enter into the situation of the agent such that the spectator agrees with the agent even though the agent feels improperly. This is often the case when the spectator is partial towards the agent (e.g. *TMS* 3.1.3.22), or when an entire group is partial, e.g. toward its own nation, though in agreement with one another (see *TMS* 6.2.2.3). The goal then is not to come to an agreement for agreement's sake. Rather, agreement is often, though not always, an indicator that one is closer to the proper sentiment.

Thus far, I have attempted to draw a distinction between spectators and agents in the way they err, as well as the ways they correct for those errors. This division, however, is a bit more complicated than I have presented it. In discussing the problem of self-deceit, it is clear that Smith is particularly worried about the role of self-love in distorting conscience's judgments:

> The violent emotions which at that time agitate us, discolour our view of things…. every thing appears magnified and misrepresented by *self-love* (*TMS* 3.1.4.3, emphasis added).
>
> Nature, however, has not left this weakness, which is of so much importance, altogether without a remedy, nor has she abandoned us entirely to the self-delusions of *self-love* (*TMS* 3.1.4.7, emphasis added).

This partiality, this excessive self-love, follows from the fact that though conscience only emerges if we place ourselves into the situation of a spectator, conscience and the agent are still one and the same. As a result, conscience tends to be biased, in the form of self-love, in favor of the agent. The problem is that Smith seems to leave the task of controlling self-love to the *self-command*, that results when we moderate our immediate emotional reaction to an event by incorporating our understanding of how others are reacting to it, where these

reactions are invariably less violent or forceful when uncolored by self-love. Little role is left for the amiable virtues to play in this arena. Though the amiable virtues might allow us to better comprehend and experience the situation of those with whom we are familiar, self-command is what allows us to prevent excesses in our sentiments, particularly self-love. Smith frequently discusses the importance of self-control, the importance of taking the steps necessary to "humble the arrogance of [our] self-love, and bring it down to something which other men can go along with" (*TMS* 2.2.2.1; also see *TMS* 1.1.5.3). In doing this, we align our sentiments with those of an impartial spectator, who is without such bias (*TMS* 1.1.5.4).[115] Smith's account of moral development emphasizes the importance of self-command in affecting the proper degree of self-love (*TMS* 3.1.3.21-26). He argues that young children lack self-command, and thus feel whatever they do to violent degrees without any concern for the propriety of their feelings. However, a child is constantly exposed to others who fail to share her sentiments (either because the adults or peers in question are not directly affected by the situation provoking the child's sentiments or not affected in the same way as is she) and actively disapprove of the child's violent reactions. Out of a desire to avoid this form or censure and a desire to bring herself in harmony with the minds of her caretakers and friends, the child will learn to control her emotional reactions and avoid excessive sentiments and feelings, stemming, as they do, from an excessive self-love (*TMS* 3.1.3.22). Even as adults, we need to continue develop a stronger self-command so as to avoid excessive sentiments for longer periods. Initially, we will continue to respond

[115] According to D. D. Raphael, self-command is "essentially to feel for ourselves only what we see others can feel for us," for the agent comes closer to impartiality by viewing herself through the eyes of the spectator (Raphael 2007, 34). Raphael also points out that the first appearance of 'impartial spectator' occurs in the context of the virtue of self-command (*TMS* 1.1.5.4). For Raphael, there is a close connection between impartiality and self-command: self-command is the means by which we become impartial since we feel much closer to what an impartial spectator would.

immediately with improper sentiments. But as our self-command gets stronger, we become quicker at "checking" those immediate responses by more quickly considering how we would appear from an impartial stance (*TMS* 3.1.3.23-24). With practice and habituation, we begin to simply respond immediately the way that we would as impartial spectators. In a way, we no longer appeal to conscience to correct for our sentiments because we respond in the first instance as conscience would. It is only in this last stage that we have become the "man of real constancy and firmness, the wise and just man", what Smith later refers to as "the virtuous person" (*TMS* 3.1.3.35), who is able to maintain control over his sentiments in "all occasions" (*TMS* 3.1.3.25):

> He has been in the constant practice, and, indeed, under the constant necessity, of modelling, or of endeavouring to model, not only his outward conduct and behaviour, but, as much as he can, even his inward sentiments and feelings, according to those of this awful and respectable judge. He does not merely affect the sentiments of the impartial spectator. He really adopts them. He almost identifies himself with, he almost becomes himself that impartial spectator, and scarce even feels but as that great arbiter of his conduct directs him to feel (*TMS* 3.1.3.25).

The virtuous person does not need to "check" her immediate responses by appealing to conscience because her immediate responses are already aligned with consciences'. So the virtuous person is not suppressing who she is by overriding her initial responses with that of conscience; she is not replacing her judgments with those of conscience. Rather, she identifies with conscience in that she just responds to situations the way that conscience would.

For Smith, however, self-command corrects for partiality in agents, not in spectators including conscience. In all his appeals to self-command, Smith focuses on our development as moral agents, not as spectators. As Raphael points out, Smith first stresses impartiality

"only when he came to theorize about the *effect on the agent* of the reactions of spectators" (Raphael 2004, 34; also see Shaver 2006).[116] The agent, upon facing disagreement from spectators, aims to view herself with impartiality so as to correct her initial moral judgment. To do so, the agent needs to develop self-command, "a virtue of 'the person principally concerned' [i.e. the agent] and is the result of an endeavour [by the agent] to restrain natural emotion and to lower its pitch to that which the ordinary (not the especially humane) spectator feels by sympathy" (Raphael 2004, 34). In all other instances that Smith discusses self-command, he again stresses the self-command of agents in response to spectators, never the self-command of those spectators themselves (e.g. *TMS* 3.1.3.22-27, 6.1.1.11, 6.3.1.1-5). So it should be no surprise that the "great school of self-command" is a school *for the agent*, not for spectators. In this school, the agent "studies to be more and more master of itself" in response to her peers who "have no such indulgent partiality", so that she can "gain their favour" and "avoid their hatred and contempt" (*TMS* 3.1.3.22). Given that self-command is a virtue developed *by the agent,* in response to spectators, self-command seems a more apt virtue for us as agents in response to the supposed impartial spectator.

Also, self-command cannot correct for conscience in cases of self-deception. Self-deception blinds us from our own partiality: we believe ourselves to judge correctly when we do not. Hence, we are unaware that conscience has erred, that we have incorrectly judged what is proper to feel in the given situation. Hence, even if we have strong enough self-command to moderate our sentiments to the proper degree, self-deception prevents us from being able to recognize that our sentiments are not proper in degree. In other words, we cannot directly address an errant conscience with self-command by identifying self-deception

[116] "[W]hat noble propriety and grace do we feel in the conduct of those who, in their own case, exert that recollection and self-command which constitute the dignity of every passion, and which bring it down to what others can enter into!" (*TMS* 1.1.5.3).

in the moment it occurs and producing a new, corrected moral judgment. We can reflect on conscience's past moral judgments and recognize that we deceived ourselves. But we must do so after we are no longer deceiving ourselves. In the moment of self-deception, we have no means for identifying that deception since we are its victims.[117]

The process of correcting conscience by improving our judgments of others does not face the same problem of self-deception. The problem of self-deceit was that conscience judges incorrectly as a result of the infection of excess self-love, and we are unaware that conscience has made such an error. We can circumvent the problem of self-deceit by simply correcting for the infection of self-love, such that self-deception does not (or is less likely to) arise. The amiable virtues are essentially doing just that: they improve all our moral judgments by improving our general capacity as spectators, whether of others or of ourselves, to account for all the different features of the situation, including our relation to the agent, to come to a more accurate moral judgment.

So if I am correct, there is a sharp division between the roles of self-command and of the amiable virtues: self-command produces the proper sentiments and actions in the agent, and the amiable virtues produce correct moral judgments on the part of spectators. Given that conscience is our acting as spectator, not as agent, the virtues properly suited to improve conscience are the amiable virtues. This is not to say that self-command plays not role in the cultivation of conscience. As I will detail in the next section, the two set of virtues are intricately tied to one another, such that those with one set of virtues generally have the other. However, if we are to become better judges, we should not focus primarily on improving

[117] Hence, Smith presents the problem of self-deceit (in *TMS* 3.1.4) *after* his discussion of the development of self-command (in *TMS* 3.1.3).

self-command, but rather focus on improving the imaginative capacities that give rise to our moral judgments.

3.5 How to Cultivate the Amiable Virtues

So far, I have claimed that to improve the judgments of conscience, we need to improve as spectators, and that to improve as spectators, we need to cultivate the amiable virtues. Smith, unfortunately, does not provide any clear guide for how we are to acquire these virtues. He does discuss moral development generally (in *TMS* 3.1.3.21-28), in which he focuses on the virtue of self-command, and of the character of the different virtues (in *TMS* 1.1.5, 2.2.1-3, and all of *TMS* Part 6). Still, Smith's account leaves us with some insight into how we can cultivate the specifically amiable virtues. In this section, I attempt to piece together different parts of *TMS* to describe at least one important way in which we can cultivate the amiable virtues: *experience with diversity*. Through experience with diversity, we gain a better understanding of the different influences on our own faculties and so can better account for these influences when we imaginatively enter into the agent's situation. As a result, we come to a more accurate, more impartial moral judgment of others and ourselves.

In *TMS* 1.1.4.6, Smith claims that spectator can come closer to feeling the proper sentiment when entering into the agent's situation by:

> endeavour[ing], as much as he can, to put himself in the situation of the other, and to bring home to himself every little circumstance of distress which can possibly occur to the sufferer. He must adopt the whole case of his companion with all its minutest incidents; and strive to render as perfectly as possible, that imaginary change of situation upon which his sympathy is founded (*TMS* 1.1.4.6).

The amiable virtues, I argued, consist in the disposition to engage in this imaginative exercise perfectly, resulting in the proper sentiment and the correct moral judgment of the agent. But how do we attain this perfection? Smith does not say much about how to cultivate specifically the amiable virtues. However, in his discussion of the development of self-command, he gives some clues as to how we are to attain the amiable, and all other, virtues.

Smith begins his discussion of the development of self-command by claiming that we obtain self-command "from that great discipline which Nature has established for the acquisition of this *and of every other virtue*; a regard to the sentiments of the real or supposed spectator of our conduct" (*TMS* 3.1.3.21, emphasis added). Smith thinks that not only self-command but also "every other virtue" is attained by the appeal to "the sentiments of the real or supposed spectator". The appeal to the sentiments of others does not mean that we align our sentiments with theirs or that we take their sentiments to be perfectly proper. As I argued before (in §3.4), proper sentiment is generally found somewhere between what the agent feels and what the spectator feels. Rather, we appeal to the sentiments of others to gauge the propriety of our own sentiments. We all suffer from the same blind spot when in comes to our sentiments, actions, character, and all other products of our faculties: they always present themselves to us as perfectly proper. As such, we cannot rely solely on our own judgment to determine when we are erring. We need to rely on those of *others* as a comparison point to bring to our awareness of the possibility that we might be in error. Their agreement confirms the propriety of our sentiments; their disagreement indicates the possibility that *someone* feels improperly.

The cultivation of virtues occurs primarily in the face of disagreement because disagreement indicates the need for someone to change their sentiments, change the

operation of their faculties. Disagreements do not indicate *who* feels improperly or *the source* of the improper sentiment. It is left to the agent and spectator to make those determinations. However, disagreement is necessary for both the agent and the spectator to begin considering whether she is the one feeling improperly and thus the source of this impropriety or whether the other party is to blame.

Smith first introduces this process of growth through disagreement in *TMS* 1.1.4.5, in which disagreement in sentiments leads the *spectator* to consider possible sources of this disagreement. The spectator might feel differently because of “the different degree of attention, which our different habits of life allow us to give”, or maybe because of “the different degree of natural acuteness in the faculty of the mind to which they are addressed” (*TMS* 1.1.4.2). Alternatively, the disagreement in sentiments may arise because the spectator “do[es] not view [the agent’s situation] from the same station” (*TMS* 1.1.4.5). We do not seek to correct our sentiments when we are in agreement because agreement merely confirms what we already feel. We only seek these possible sources of error when we face disagreement, when the propriety of our sentiments are put into doubt.

Smith presents this same process again in *TMS* 3.1.3.22-25, this time on the part of the agent’s cultivation of self-command. Infants do not develop any self-command because they are primarily in the care of parents who are apt to agree with the infant’s sentiments, with the exception of the infant’s anger (*TMS* 3.1.3.22). Once a child is old enough to attend school and interact with peers, who do not share the same penchant as parents to match the sentiment of the child, the child now faces disagreements and must learn to adjust her sentiments, thus entering the “great school of self-command” (*TMS* 3.1.3.22). Through the continual experience of facing such disagreements and resolving those disagreements by

adjusting their sentiments, children learn to better approximate the proper emotional reaction to the situations they face. Later in life, we become aware of our impropriety through disagreement with ourselves in the form of conscience, and adjust our sentiments in light of this disagreement to become the "man of real constancy and firmness" (*TMS* 3.1.3.25). At each stage, we turn to others for agreement or disagreement to help us determine whether we are feeling properly or not. Agreement confirms the propriety of our sentiments; disagreement indicates we may need to adjust them.

Disagreement alone, however, does not lead to the cultivation of virtues. We need some motivation for resolving these disagreements if we are to come to experience a better (or more proper) reaction to our triumphs and disappointments. Disagreements simply indicate that there are some differences between individuals that give rise to those disagreements, and without augmentation, the realization that our reactions disagree with those of another might do nothing more than give rise to moral disapproval of him or her. To cultivate virtues, we need to be motivated to *change* our sentiments in light of their differences with those of others. For Smith, the primary motive to do so is the pleasure of mutual sympathy: "[N]othing pleases us more than to observe in other men a fellow-feeling with all the emotions of our own breast; nor are we ever so much shocked as by the appearance of the contrary" (*TMS* 1.1.2.1). This desire for mutual sympathy is quite strong: "A man is *mortified* when, after having endeavoured to divert the company, he looks round and sees that nobody laughs at his jest but himself. On the contrary, the mirth of the company is *highly agreeable* to him, and he regards this correspondence of their sentiments with his own as the greatest applause" (*TMS* 1.1.2.1, emphases added).[118] It is because we care so

[118] The desire for mutual sympathy can vary in degree. This desire is not quite so strong when we are considering matters of taste. However, Smith is clear that when the situation deeply affects us, we have a strong

deeply about attaining this mutual sympathy that we do not leave disagreements as they are, but rather seek to resolve those disagreements by adjusting our sentiments to come to an agreement.

Finally, in the process of resolving disagreements, we need to be drawn *towards* the proper sentiment. A disagreement is resolved by coming to feel the same sentiment to the same degree. However, the correspondence of sentiments does not insure that we are both feeling properly. Two people can share in the same improper sentiment. For instance, if two people are prone to excessive rage, they may approve of each others' rage, even when improper. Hence, Smith writes, "This natural disposition to accommodate and to assimilate, as much as we can, our own sentiments, principles, and feelings, to those which we see fixed and rooted in the persons whom we are obliged to live and converse a great deal with, is the cause of the contagious effects of both good and bad company" (*TMS* 6.3.1.17).[119] Even whole populations are susceptible to improper sentiments when judging others. Hence, Smith distinguishes between praise, which may or may not be proper, and praise of that which is praiseworthy, which is proper. Spectators may praise the agent, grounded in their agreement in sentiment with the agent, even though the agent may not be deserving of that praise because her sentiment is improper.

---

desire for mutual sympathy, even though that mutual sympathy is at the same time more difficult to attain: "With regard to those objects, which affect in a particular manner either ourselves or the person whose sentiments we judge of, it is at once more difficult to preserve this harmony and correspondence, and at the same time, vastly more important…. [I]f you have either no fellow-feeling for the misfortunes I have met with, or none that bears any proportion to the grief which distracts me; or if you have either no indignation at the injuries I have suffered, or none that bears any proportion to the resentment which transports me, we can no longer converse upon these subjects. We become intolerable to one another. I can neither support your company, nor you mine. You are confounded at my violence and passion, and I am enraged at your cold insensibility and want of feeling" (TMS 1.1.4.5).

[119] Smith presents the parallel case on matters of taste in *TMS* 5.1.1.2: "Where the conjunction is improper, custom either diminishes, or takes away altogether, our sense of the impropriety. Those who have been accustomed to slovenly disorder lose all sense of neatness or elegance. The modes of furniture or dress which seem ridiculous to strangers, give no offence to the people who are used to them."

So to cultivate virtues, we need to resolve disagreements in a way that draws the spectator and agent *toward* the proper sentiment. Commentators have recognized two key suggestions made by Smith on this point. One is the appeal to virtuous people, who feel the proper sentiment perfectly. We, who are not virtuous, do not agree perfectly in sentiment with virtuous people. However, the sentiments of virtuous people elicit feelings of wonder and surprise in us, which lead us to admire them (e.g. *TMS* 1.1.4.3, 1.2.1.12). This admiration leads us to want to be like them. The virtuous person becomes "the man who directs and conducts our own sentiments" (*TMS* 1.1.4.3). In seeking out the virtuous, we are not simply mimicking their sentiments, but rather attempting to exercise our faculties as they do. We admire how they "have attended to many things which we had overlooked" and "have adjusted them to all the various circumstances of their objects". We feel wonder and surprise "at their uncommon and unexpected acuteness and comprehensiveness" (*TMS* 1.1.4.3).[120] Our admiration is for the way in which virtuous people come to the proper sentiment, in the way they attune to the details of the situation and account for them in coming to their moral judgments. Furthermore, the natural response to admiration is praise, and we naturally desire both to be praised and to be worthy of that praise (i.e. to have qualities deserving of praise, as the virtuous person has) (*TMS* 3.1.2).[121] For all these reasons, we are naturally motivated to be like the virtuous person, the person who is perfectly proper in all situations, as spectator and as agent.

[120] Smith repeats these points on our admiration for aesthetic and intellectual virtues: "It is the acute and delicate discernment of the man of taste, who distinguishes the minute, and scarce perceptible differences of beauty and deformity; it is the comprehensive accuracy of the experienced mathematician, who unravels, with ease, the most intricate and perplexed proportions" (*TMS* 1.1.4.3).
[121] Though people have a natural desire for both praise and praiseworthiness (*TMS* 3.1.2.1), they may vary on the degree to which they desire each: "In this respect, however, men differ considerably from one another. Some seem indifferent about the praise, when, in their own minds, they are perfectly satisfied that they have attained the praise-worthiness. Others appear much less anxious about the praise-worthiness than about the praise" (*TMS* 3.1.2.28).

Virtuous people can lead us toward the proper sentiments and the virtues. However, it is difficult to rely on virtuous people because there simply are not many, if any, of them. As Smith acknowledges, the genuinely virtuous person, the person with "complete propriety and perfection", might be impossible to find, for "in those difficult situations, no human conduct ever did, or ever can come up to; and in comparison with which the actions of all men must for ever appear blameable and imperfect" (*TMS* 1.1.5.9). The best we can hope for is to find someone who, though not perfectly proper, is closer to that perfection than what is commonly found in society. Moreover, even if we encounter those that are nearer to this perfect virtue, we can be led astray because our feelings of wonder and admiration are liable to corruption. Who we feel wonder and admiration for depends in part on our experiences, on the kinds of people we spend time with:

> The man who associates chiefly with the wise and the virtuous, though he may not himself become either wise or virtuous, cannot help conceiving a certain respect at least for wisdom and virtue; and the man who associates chiefly with the profligate and the dissolute, though he may not himself become profligate and dissolute, must soon lose, at least, all his original abhorrence of profligacy and dissolution of manners. The similarity of family characters, which we so frequently see transmitted through several successive generations, may, perhaps, be partly owing to this disposition, to assimilate ourselves to those whom we are obliged to live and converse a great deal with (*TMS* 6.3.1.17).[122]

Smith admits that we often confuse praise with praise-worthiness. We confuse those who are virtuous from those who seem virtuous, and so end up admiring those that are not deserving

[122] In this passage, Smith also indicates that biology may also contribute to our dispositions: "The family of character, however, like the family countenance, seems to be owing, not altogether to the moral, but partly too to the physical connection. The family countenance is certainly altogether owing to the latter" (*TMS* 6.3.1.17).

of such admiration (*TMS* 3.1.2). We "regard with foolish wonder and admiration" the "violence and injustice of great conquerors" who are far from virtuous (*TMS* 6.1.1.16).[123]

An alternative means of attaining the virtues is to appeal to impartial spectators. Smith repeatedly appeals to impartial spectators as the standard for the proper sentiment. For instance, when we feel the improper sentiment or the proper sentiment to an improper degree, Smith suggests that we seek out impartial spectators to adjust our sentiments to what is more proper:

> Are you in adversity? Do not mourn in the darkness of *solitude*, do not regulate your sorrow according to the indulgent sympathy of your intimate friends; return, as soon as possible, to the day-light of the world and of society. *Live with strangers*, with those who know nothing, or care nothing about your misfortune….
>
> Are you in prosperity? Do not *confine* the enjoyment of your good fortune to your own house, to the company of your own friends, perhaps of your flatterers, of those who build upon your fortune the hopes of mending their own; frequent those who are *independent* of you, who can value you only for your character and conduct, and not for your fortune…. (*TMS* 3.1.3.39-40, emphases added).

Smith's suggestion in both cases is the same: spend time away from those who are partial, away from yourself ("solitude"), "intimate friends", and "flatterers", and instead spend time with "strangers" and "those who are independent of you". In other words, seek out those who are impartial, who do not have any particular ties to you. Of course, in this passage, Smith is discussing how as agents, we need to seek out impartial spectators to adjust our sentiments.

---

[123] There is an additional danger in relying on the looser standard of propriety when we are judging ourselves. When we judge ourselves along the looser standard, we focus on how we compare to others, who might be less virtuous along this standard, we begin to lose sense of our weakness and imperfection. We become "assuming, arrogant, and presumptuous" (*TMS* 6.3.1.27). This problem does not occur when we judge ourselves on the stricter standard of perfect propriety, since under this standard "the wisest and best of us all… see nothing but weakness and imperfection" (*TMS* 6.3.1.24). Thus, we find that the "wise and virtuous man directs his principal attention to the first standard; the idea of exact propriety and perfection" (*TMS* 6.3.1.25), so as not to fall under excessive self-admiration.

However, the same process also helps spectators improve their judgments of others. By comparing our own judgments to those of impartial spectators, we become aware of any partiality we might have toward the agent and come to adjust our sentiments accordingly. Griswold stresses this aspect of Smith when he writes:

> To that end, we must learn how to sympathize properly, and therefore how to think imaginatively about another's and our own situation. The imagination must be trained rightly. Judgment about the ethical course of action to be followed in a given situation requires imagination, since we must represent to ourselves possible courses of action, consider the intended effects of our choice on others and thus view the situation from their standpoint, and indeed review our own motivations from the standpoint of an impartial spectator (Griswold 1999, 214).

We are not born with the disposition to judge others correctly. We learn to do so by improving our capacity to sympathize with others, by understanding how to assess the agent's situation correctly by coming to the proper sentiment. And we learn to do so by appealing to "the standpoint of an impartial spectator" because one of the key errors we make as spectators is to allow our partiality to corrupt our moral sentiments. When we compare our sentiments as spectators to those of impartial spectators, we can begin to recognize how our sentiments might suffer from such corrupting influences. We use the impartial spectator's viewpoint to learn how to account for our partiality when judging others. As we do when looking to virtuous people as moral guides, we learn not only what sentiment to feel and to what degree, but also how to evaluate the agent's situation, how to imagine and sympathize with the agent, as the impartial spectator does.

Though I agree that impartial spectators can help us correct for our partialities when we judge others, they do little else for helping us improve as spectators. Appeal to impartial spectators is especially important for Smith because Smith treats partiality as the primary

source of errant moral judgments. Thus, he stresses the need to rely on impartial spectators throughout *TMS*, in all stages of moral development. However, it would be a mistake to think that partiality is the *only* source of errant moral judgments. As I argued earlier, there are several other sources of our improper sentiments as spectators, and it seems that impartial spectators are just as susceptible to these other sources of error as partial spectators. For instance, impartial spectators do not necessarily have better information about the agent's situation *simply in virtue of being an impartial spectator* because they suffer from the same epistemological shortcomings as partial spectators. Also, they are not necessarily better at accounting for various inconsistencies in the operation of their own faculties because being impartial does not adjust for how aware you are of your present physical and emotional state. Finally, they do not help us account for the corrupting influences of fortune, custom, and fashion. Impartial spectators, like partial spectators, are just as susceptible to giving fortune too much weight in forming moral judgments, or failing to account for differences in custom in coming to a moral judgment. So though I accept the importance of appealing to impartial spectators given the significant impact of partiality on our moral judgments, I am also suggesting that the appeal to impartial spectators is very much incomplete. We need some other means of correcting for the other kinds of errors we make as spectators, errors that Smith is well aware of, if we are to cultivate the amiable virtues.

There is a crucial third means for attaining virtues, one that is consistent with Smith's theory (though he does not suggest it): *experience with diversity*. Earlier, I argued that we can only gain awareness of the corruption of our own moral sentiments and those of others through disagreement. Yet each disagreement only brings to light some of the errors we make. For example, we might disagree in our moral judgment of the agent because I am a

close friend of the agent, whereas you are not. This disagreement brings to light a particular kind of partiality of which I was not aware of before: friendship. Yet it does not bring to light the other kinds of errors that I may often make. I might, for instance, also be biased toward people of a certain ethnicity or race in ways that others might not be. Until I encounter a disagreement with others arising from this particular kind of error, I will not be aware of my errant moral sentiment. But once I face such a disagreement, I am now prompted, as Smith suggests, to consider my sentiment from different perspectives, e.g. the person I may have offended or a third-party that is of neither my ethnicity nor the offended. Also, I might tend to form moral judgments hastily without considering all the facts of the situation. When this process leads to a disagreement with someone else, and that person points out some features of the situation that I have not considered, then I am prompted to reconsider my own judgment in light of this new information. I might also be prompted to correct for my tendency to overlook details of the situation by thinking more carefully about all the different facets of the case before coming to a judgment. So if we want to account for *all* the different kinds of errors in our moral judgments, we need to seek out a wide variety of disagreements with others. But we can only encounter such disagreement if we engage with people that differ in different ways, e.g. in partiality, in intellectual capacities, in fortune, in custom, in fashion. The more diversity we encounter, the more aware we become of how we, as well as others, err in moral judgments.

In seeking out people of different backgrounds, we need not necessarily distinguish between virtuous, impartial, and partial spectators. We can learn from disagreements with any person, including a partial spectator. For instance, Smith recognizes that sometimes partial spectators might be much closer to the proper sentiment than an impartial spectator

when judging the agent: “We expect less sympathy from a common acquaintance than from a friend: *we cannot open to the former all those little circumstances which we can unfold to the latter*” (*TMS* 1.1.4.9, emphasis added). Partial spectators might be more likely to agree with the agent in virtue of their relationship with the agent. However, because partial spectators also know the agent better, partial spectators are more likely to have information about the full circumstances that impartial spectators do not. As a result, partial spectators might in at least one way be *more* accurate than impartial spectators.

There is a danger that through repeated experience with partial spectators, we will develop dispositions that lead us away from the proper sentiment; their biases become our own. However, this danger is less of a worry if we diversify our experience because the differences in different partial spectators offset the effects of some particular partial spectator. We do not improve as spectators if we are exposed only to our parents who tend to be quite partial in their judgments toward us. However, if we are exposed to a range of different partial spectators, e.g. our parents, our friends, and our colleagues, each different in the particular degree of partiality and in talents and background, we will be less likely to be habituated to one particular kind of bias or one kind of error. What one spectator fails to account for, another does not. For instance, I may have one friend who is very attentive to details and so is able to recognize the finer details of an agent’s situation. Another friend may be better at recognizing the influence of fortune on our sentiments. Given the differences in the abilities of my different friends, my disagreements with each arise from different sources. With one friend, I become more aware of my failure to account for some detail of the situation, and with the other, I become more aware of my disposition to be influenced too much by fortune. So even though some people might feel more proper sentiments than

others, we can still learn something from the disagreements we have with any of them. Thus, by interacting with a range of different people who provide different insight into the operations of our faculties, as well as those of others, we begin to develop a richer understanding of the many ways we tend to err in our moral sentiments, and can adjust for all those errors.

Though Smith does not suggest or endorse this process of encountering diversity to cultivate amiable virtues, he does stress the importance of experience with various situations in moral development. For instance, as Griswold points out, Smith emphasizes the importance of literature in moral development, in both *TMS* and in *WN*[124], in that literature allows us to navigate through the complexities of different situations, and in doing so, hone our capacities to imaginatively enter into the agent's situation:

> Smith's continual references to literature, especially tragedies, are surely meant to help us to grasp how so difficult an achievement as moral education is to be attained. For study of literature teaches us how ethically complex human situations can be, how to stretch the moral imagination so as to size up the relevant factors, how to carry on a conversation about the competing claims of the dramatis personae (Griswold 1999, 215).

It is not simply engaging with any literature but literature that "makes visible to us different shades of 'circumstance, character, and situation' that no 'casuistic rules' could ever define" that leads us to improve as spectators (Griswold 1999, 216).[125] We improve as spectators

[124] "Literature also plays an important and variegated role in *The Theory of Moral Sentiments.* Not only plays, novels, and poems but tragedies, in particular, intrigue Smith. Together they completely overshadow his relatively rare references to properly philosophical texts (putting aside *TMS* VII). The notion that we are to understand literature and drama as sources for moral theory and moral education is clearly and strikingly evident in *The Wealth of Nations* as well (V.i.g.14-15)" (Griswold 1999, 216).
[125] Here, Griswold is quoting Smith's praise of Voltaire's *Mahomet* as a source of moral instruction (*TMS* 6.2.1.22).

because we are practicing how to engage with a range of *different* situations, each with its own unique set of factors to consider.

Smith himself seems to engage in this kind of moral development in his discussion of the influence of custom and fashion. In Part V of *TMS*, Smith begins with what appears to be disagreements in sentiments: what a person of one custom deems proper, a person of a different custom deems improper. We tend to judge inaccurately in such cases because we fail to account for how our custom and fashion influence our judgments of others. We recognize how we, or the other person, is influenced in this way when we disagree with one another (*TMS* 5.1.2.6). We resolve these disagreements by better understanding the other person's customs, by better understanding the thoughts and sentiments from which these customs arose, by better understanding the utility of these customs and seeing the beauty in that utility, and so forth. Once we do, we can come to a more accurate moral judgment. Smith engages in this very process in his discussion of infanticide. We tend to disapprove of infanticide. Yet, there are certain customs that find infanticide perfectly permissible. Such a disagreement leads Smith to consider how such customs arose, from what sentiments and dispositions. Depending on what we come to find about the differences in customs, we may end up approving of the agent:

> The extreme indigence of a savage is often such that he himself is frequently exposed to the greatest extremity of hunger, he often dies of pure want, and it is frequently impossible for him to support both himself and his child. We cannot wonder, whether, that in this case he should abandon it. One who, in flying from an enemy, whom it was impossible to resist, should throw down his infant, because it retarded his flight, would surely be excusable; since, by attempting to save it, he could only hope for the consolation of dying with it (*TMS* 5.1.2.15).

In such cases, we find that the differences in custom are rooted in the same principle: people who share the proper degree of self-love and of concern for an infant may differ in their behavior if they differ sufficiently in their circumstances. Thus, though the practices of two communities might differ, we may nevertheless find that their sentiments align. Or alternatively, we might find that even after accounting for differences in custom, we find that the agent acts improperly because his customs are not rooted in the requisite concern:

> In the latter ages of Greece, however, the same thing was permitted from views of remote interest or conveniency, which could by no means excuse it…. When custom can give sanction to so dreadful a violation of humanity, we may well imagine that there is scarce any particular practice so gross which it cannot authorise. Such a thing, we hear men every day saying, is commonly done, and they seem to think this a sufficient apology for what, in itself is the most unjust and unreasonable conduct (*TMS* 5.1.2.15).[126]

My point is that whether or not we agree with a particular custom, we cannot form an accurate moral judgment of an agent unless we take into account the customs under which he acts, how these customs differ from our own, and how our customs influence our moral judgments. It is a mistake to judge the agent as if he were acting under our customs when he is not. Our customs might distort our moral sentiments, or the agent's customs might distort his moral sentiments and the actions that follow upon them. But without having some understanding of these different customs, what purpose they serve, what sentiments motivate those practices, and so forth, we will not be able to accurately judge the agent. And we can only come to such an understanding of other customs if we are exposed to them, so that we

[126] Smith also engage in this process of encountering disagreements arising from difference in customs and correcting (or affirming) his moral sentiments with a better understanding of the customs of others, when he discusses the difference in custom among different professions (see *TMS* 5.1.2.5-6).

can at least make sense of the "particular turn or habit of imagination" they entrain (*TMS* 1.2.2.1).

3.6 Concluding Remarks

The general focus of this discussion of Smith's theory of moral development is on the role of self-command, general rules, and impartial spectators. Yet these three aspects of his account are limited in important ways. Self-command and the appeal to impartial spectators are necessary for the agent to feel and act properly, but unless we already judge correctly as impartial spectators, neither self-command nor their best, sincere attempts to be impartial will lead reflective agents toward from the proper sentiment. General rules are useful for promoting the proper course of action on matters of justice, but they are of little use in all other cases. Instead, I have argued that we need to improve as spectators by cultivating the amiable virtues. These virtues require us to do more than account for our own partiality; we need to account for errant beliefs about the situation of the agent we are judging, inconsistencies in the operations of our own faculties, and various biases (especially those of custom and ethnicity) that distort our moral sentiments. Only once we improve as spectators, will conscience, i.e. the supposed impartial spectator, serve as the proper moral guide.

My suggestion for cultivating the amiable virtues is to engage with a diverse range of people in different situations, to use the disagreements we face with others as a way to better understand how we come to our own moral sentiments. The benefit of this approach is that we can develop as spectators in many ways beyond impartiality. We can improve our attentiveness to the details of the agent's situation. We can recognize the influence of custom, fashion, and fortune on our moral judgments. We can gain better awareness of the influence

of our current physical and emotional state on our moral judgments. By using others as a point of contrast, we are better able to understand *ourselves* as both spectators and agents. And with an improved self-awareness, we are better able to account for all the ways our faculties might lead us to an improper sentiment.

In some ways, my suggestion is similar to Hobbes' prescription in *Leviathan* "*nosce teipsum*, read thyself". Hobbes claimed that we can know about the faculties and principles of others by looking inward and reflecting on the operations of our own faculties and principles.[127] However, my claim is that we need to look outward to others to gain a better understanding of ourselves. When we look inward toward ourselves, we are blind to our own errors. Conscience is especially susceptible to this blindness because we dislike judging ourselves negatively. So if we are to correct for an errant conscience, we need to rely on those around us, those who can provide different insight through their disagreements with us.

[127] "[F]or the simultitude of the thoughts and passions of one man to the thoughts and passions of another, whosoever looketh into himself and considerth what he doth, when he doth *think, opine, reason, hope, fear,* &c, and upon what grounds, he shall thereby read and know, what are the thoughts and passions of all other men upon the like occasions" (*Leviathan* Intro.3).

4

The Authority of Conscience

Conscience does not offer itself to shew us the way we should walk in, but it likewise carries its own authority with it, that it is our natural guide, the guide assigned us by the author of our nature: it therefore belongs to our condition of being, it is our duty, to walk in that path and follow this guide, without looking about to see whether we may not possibly forsake them with impunity.

Bishop Joseph Butler from *Fifteen Sermons* 3.5

There is a higher court than courts of justice and that is the court of conscience. It supercedes all other courts.

Mahatma Gandhi from "A Puzzle and Its Solution"[128]

Conscience is frequently characterized as having authority over us, as binding us to its commands such that we ought to obey them. However, it is not always clear what this authority consists in. There are several features of our phenomenological experience of conscience that are often associated with conscience's authority. We feel compelled to act in accordance with conscience's dictates, and when we fail to do so, conscience torments us, especially through feelings of guilt and remorse. Conscience can also speak in different "volumes", from prodding us incessantly to remaining silent. Even when conscience speaks loudly, we may struggle to obey it. We feel torn about what to do, and at times are willing to

[128] M. A. Gandhi. "A Puzzle and Its Solution" in *The Gandhi Reader*, ed. Homer A. Jack (1994, 194).

act against our own conscience. Finally, conscience presents itself as *the* moral authority, as the final word on what is right or wrong to do. This experience of conscience has led to characterizing conscience as the internal representative of God or the innate awareness of *the* moral law.

There have been different attempts at explaining the authority of conscience. Aquinas and Bonaventure, for instance, thought conscience's authority is derived from God: conscience simply presents us with God's dictates applied to particular situations through practical reasoning. Since we ought always to obey God, we ought always to obey conscience. Butler, on the other hand, argued that conscience has authority in virtue of its very position and function as the judging faculty. So we ought to obey conscience for the constitution of human nature demands that we do. Others, like Mill and Freud, grounded conscience's authority in its motivational force: conscience is authoritative insofar as it is efficacious in bringing about action in accordance with its dictates. Thus, we ought to obey conscience only to the degree that we feel compelled to do so.

Smith, however, is not very clear on what it is for conscience to be authoritative, even in the chapter of *TMS* titled "Of the influence and authority of conscience". But given some similarities between passages in Smith's *TMS* and Butler's *Fifteen Sermons*, several commentators have suggested that Smith adopts Butler's account of conscience's authority: conscience ought to be obeyed because it was placed in the position of authority, the position to rule, within the human constitution for it is the only faculty that judges all others. In this chapter, I argue that this interpretation is mistaken. Unlike Butler, Smith does not view authority as distinct from mere power. Rather, authority is a particular kind of power: the power to procure obedience through respect resulting from the distinction of rank.

Conscience has such an authority because we recognize conscience as a superior moral judge, i.e. respect it. The degree to which we ought to obey conscience, then, is dependent on the strength of this respect we have for conscience.

Smith, however, fails to capture what is central to our experience of conscience as authoritative: conscience presents itself as *the* moral guide, as the final judge of what is proper. It is the voice of God or the moral laws written in our hearts, regardless of how motivated we are to obey it. Hence, even when we act against conscience, we always do so *believing that we are acting improperly*. So it is not that we recognize conscience as a better judge; rather, we cannot view conscience as anything but the *best* judge. Still, I believe Smith's *theory* can account for why conscience must be viewed as the moral guide. Since conscience is an exercise of the agent's faculties, the agent is committed to conscience's judgments the same way she is committed to her own judgments. Because there is no other means for the agent to judge other than the use of her own faculties, all her judgments seem correct to her. So the agent cannot dismiss conscience's judgments as mistaken because they are her own judgments and thus will always seem correct to her.

4.1 Butler and Smith on the Authority of Conscience

Commentators have been long aware of the similarities between Smith's and Butler's account of conscience. In particular, people frequently note the similarities between Butler's *Fifteen Sermons* 2.14-15 and Smith's *TMS* 3.1.5.5:

> Thus that principle, by which we survey, and either approve or disapprove our own heart, temper, and actions, is not only to be considered as what is in its turn to have some influence…. [Y]ou cannot form a notion of this faculty, conscience, without taking in judgment, direction, superintendency. This is a constituent part of the idea, that is, of the faculty itself: and , to preside and govern, from the very economy and

constitution of man, belongs to it…. [T]hat this faculty was placed within to be our proper govenour; to direct and regulate all under principles, passions, and motives of action (*Sermons* 2.14-15).

[Our moral faculties] carry along with them the most evident badges of this authority, which denote that they were set up within us to be the supreme arbiters of all our actions, to superintend all our sense, passions and appetites, and to judge how far each of them was either to be indulged or restrained…. No other faculty or principle of action judges of any other…. But it is the peculiar office of those faculties now under our consideration to judge, to bestow censure or applause upon all the other principles of our nature…. It belongs to our moral faculties, in the same manner to determine when the ear ought to be soothed, when the eye ought to be indulged, when the taste ought to be gratified, when and how far every other principle of our nature ought either to be indulged or restrained. What is agreeable to our moral faculties, is fit, and right, and proper to be done; the contrary wrong, unfit, and improper. The sentiments which they approve of, are graceful and becoming: the contrary, ungraceful and unbecoming. The very words, right, wrong, fit, improper, graceful, unbecoming, mean only what pleases or displeases those faculties (TMS 3.1.5.5).

Butler is stressing a point he argues for in *Sermons* 1-3. The very idea of conscience contains within it "judgement, direction, superintendency" over the other faculties. Thus, conscience is the "proper govenour" of our other faculties, and so ought to be obeyed. In this sense, conscience has an authority distinct from mere power, since conscience ought to be obeyed regardless of whether or not conscience can procure obedience from the other faculties. The passage from *TMS* seems to echo these very points. Our moral faculties, and only those faculties, "superintend" and "judge" the other faculties. Thus, they seem to have been "set up within us to be the supreme arbiters" over the other faculties.[129] They have authority over the other faculties in virtue of their role or position as the judge of the other faculties.

The similarity between these passages have led some commentators to argue that Smith thinks conscience ought to be obeyed because of its place within human nature as the

[129] Haakonsen points out this similarity Between Butler and Smith in a footnote in *TMS* 3.1.5.5. Also see Brown (1994), Darwall (1999), Forman-Barzilai (2010), Raphael (2007).

judging faculty. Stephen Darwall, for instance, makes this point explicitly in a footnote in his article "Sympathetic Liberalism: Recent Work on Adam Smith" (1999):

> But why should we follow conscience? What gives judgments of propriety *their* authority? Here Smith follows closely Butler's idea that conscience's authority is tied up with the ancient (stoic) doctrine that virtue "consists in following nature" (*S* Preface.13). It "cannot be doubted," Smith writes, that our "moral faculties… were given us for the direction of our conduct in this life" and "were set up within us to be the supreme arbiters of all our actions" (*TMS* 3.1.5.5). By fashioning us in this way, nature (or God) promulgates the rules these faculties recognize, including prominently the rules of justice, as "laws." Conscience is God's "viceregen[t]… within us," so its "torments" inherit God and nature's authority (*TMS* 3.1.5.6). (Darwall 1999, 150).[130]

According to Darwall, then, Smith thinks that we ought to obey our conscience because conscience was placed in the *position of authority* over our other faculties. Conscience is our governor, our superintendent, our director, our authority figure which is to be obeyed, in virtue of what it is, i.e. a judging faculty.

Darwall's claims are buttressed by another similarity between Butler and Smith. Smith at one point refers to conscience as "reason": "It is reason, principle, conscience, the inhabitant of the breast, the man within, the great judge and arbiter of our conduct" (*TMS* 3.1.3.4). Recall that Butler used 'reason' and 'conscience' interchangeably; they referred to one and the same judging faculty. Butler therein identifies conscience with Plato's account of that faculty. Smith seems to support this identification by stating that Plato "very properly called" the judging faculty "reason" (*TMS* 7.2.1.3), the faculty that shares the same role as Smith's conscience: it is that faculty "by which we judge of the propriety and impropriety of

[130] More recently, similar interpretations of Smith on conscience's authority have been made by Khalil (2005) and Özler and Gabrinetti (2013).

desires and affections" (*TMS* 7.2.1.3) and to bring the passions that are "too vehement" to the proper degree, as determined by that faculty (*TMS* 7.1.2.44).

Furthermore, like Plato (and Butler), Smith claims that virtue consists in conscience exercising control over the various passions. In *TMS* 3.1.3, Smith provides a detailed picture of a person developing from a child with little self-command to a virtuous person, who exercises the greatest of self-command. The virtuous person is someone for whom conscience is ever-present, controlling the passions and bringing them down to the proper degree. Thus, insofar as we ought to be virtuous, we ought to let conscience be our guide and control all our sentiments. Smith's account seems to be a mirror image of Plato's, in which the virtues consist in the passions operating "under the direction of reason", in reason "check[ing] and restrain[ing] the inferior and brutal appetites" (*TMS* 7.2.1.7; also see *TMS* 7.2.1.8-9).[131]

Given the similarities between Smith, Butler, and Plato, it is tempting to view Smith as endorsing the view that conscience is by nature the rightful ruler over the other faculties. That is, conscience ought to be obeyed, regardless of the strength of its motivating force. It is good/right/proper that our sentiments align with conscience's, for virtue consists in each of the different parts of human nature "confin[ing] itself to its proper office, without attempting to encroach upon that of any other" and conscience's, or reason's, proper office is to command the other faculties.

[131] Fleischacker relies on Smith's identification of conscience with reason to draw a comparison between Smith and Kant: "If we recall… [Smith's] rhetorical identification of 'reason' with 'principle' and 'conscience,' it will be clear that 'reason' for him is analogous to what Kant would call 'judgment,' specifically 'determinant judgment': it determines which specific situations or feelings belong under which general rule of propriety" (Fleischacker 1991, 266). Though an interesting proposal, it seems more likely that given Smith's knowledge of Plato and Butler, he is using the term 'reason' in the same sense as Plato and Butler, which is not the same as judgment in Kant's sense. For Plato and Butler, the operation of conscience need not involve any determination of what general rule applies to a particular situations.

4.2 Distancing Smith from Butler: Smith on Conscience and "Moral Faculties"

There are, however, reasons for distancing Smith from Butler on the authority of conscience. For one, though Smith does use 'conscience' and 'reason' interchangeably, and gives conscience the role that Butler (and Plato) give to reason, Smith also explicitly denies that conscience is itself a faculty, at least in Butler's sense:

> The word conscience does not immediately denote any moral faculty by which we approve or disapprove. Conscience supposes, indeed, the existence of some such faculty, and properly signifies our consciousness of having acted agreeably or contrary to its directions (*TMS* 7.3.3.15).

Conscience is not, strictly speaking, a faculty in Butler's sense: it is not one of several parts of the human constitution.[132] Rather, conscience is our awareness of our having acted properly or not. Conscience "supposes… the existence of some such faculty", whatever that faculty may be, since there must be *some* faculty that produces the immediate motivations and thoughts that are the objects or targets we judge when engaging our consciences. However, conscience is not one of the various powers of the mind. It is a product of those powers and faculties that we use to evaluate ourselves, but it is not itself a separate power or proper object of evaluation.

Furthermore, the faculty on which conscience is founded is not itself a distinct moral or judging faculty. Smith, like Hume, is committed to the sympathetic mechanism as the source of moral judgments:

> [T]here is no occasion for supposing any new power of perception which had never been heard of before: Nature, they imagine, acts here, as in all other cases, with the strictest economy, and produces a multitude of effects from one and the same cause;

[132] Given his view that conscience is not itself a faculty, Smith is committed to rejecting an important part of Plato's theory of the parts of the soul.

> and sympathy, a power which has always been taken notice of, and with which the mind is manifestly endowed, is, they think, sufficient to account for all the effects ascribed to this peculiar faculty (*TMS* 7.3.3.3).[133]

The sympathetic mechanism is our capacity to enter into the situation of another person and imaginatively feel as we would from that same situation. This mechanism explains how we come to form moral judgments of others and ourselves (in the form of conscience), so it is a capacity by which we judge. However, unlike Butler's conscience, the sympathetic mechanism is not a power of the mind *distinct from* the passions. Rather, the sympathetic mechanism is an exercise *of* our passions, achieved through acts of the imagination. We imaginatively enter the situation of another, and we feel from this imagined situation: "we conceive ourselves enduring all the same torments… and thence form some idea of his sensations, and even *feel* something which, though weaker in degree, is not altogether unlike them" (*TMS* 1.1.1.2, emphasis added). The sympathetic mechanism is not distinct from our appetites and passions. Rather, the sympathetic mechanism is *engages* our appetites and passions. It compares our sentiments with those of others. Even when these mechanisms are engaged in their role as conscience and so directed upon our own sentiments and actions, they require that we divide ourselves to compare our sentiments as conscience to our sentiments as agent. So even if the sympathetic mechanism is the faculty by which we judge, "to bestow censure and applause upon all the other principles of our nature" (*TMS* 3.1.5.5), it is not one that is independent of the appetites and passions, as Butler claimed it to be.

Furthermore, Smith does not give authority to conscience or to the sympathetic mechanism, but instead to moral *faculties* (*TMS* 3.1.5.5). His use of the plural form of

[133] This passage comes in the context of Smith's presentation of moral theories that claim sympathy is the origin of all moral sentiments (e.g. Hume's moral theory).

'faculty' is not an accident. Smith thinks that there are four different sources of moral judgments, and not all of them involve the exercise of the sympathetic mechanism:

> When we approve of any character or action, the sentiments which we feel, are, according to the forgoing system, derived from four sources, which are in some respects different from one another. First, we sympathize with the motives of the agent; secondly, we enter into the gratitude of those who receive the benefits of his actions; thirdly, we observe that his conduct has been agreeable to the general rules by which those two sympathies generally act; and, last of all, when we consider such actions as making a part of a system of behaviour which tends to promote the happiness either of the individual or of the society, they appear to derive a beauty from this utility, not unlike that which we ascribe to any well-contrived machine" (TMS 7.3.3.15-16).[134]

The first two sources of moral judgments directly rely on the sympathetic mechanism: through the sympathetic mechanism, the spectator is able to enter into the situation of the agent (or recipient), and imaginatively feel what the spectator would from the agent's situation. The spectator compares that imagined sentiment with the sentiment of the agent (or recipient) to judge what is proper or meritous. The third source, however, does not rely directly on the sympathetic mechanism. We form general rules through the faculty of *reason*[135], through induction on our past moral judgments. These past judgments were formed using the sympathetic mechanism, but the general rules themselves are not. Similarly, the fourth source does not use the sympathetic mechanism directly. We do not approve of a system of behavior because we sympathize with the happiness of those who benefit from such a system. We approve of a system because we find *beauty* in the utility and perfection

[134] This passage reflects Smith's general agreement with Hume on the mechanisms by which we judge sentiments, actions, and character proper or improper, virtuous and vicious (see *THN* 3.3.1.27-30). The one point of disagreement is that whereas Hume claims we approve of systems because it is useful (see Hume's account of justice as an artificial virtue in *EPM* Section 3), Smith claims that we approve of systems because we find what is useful to be beautiful (*TMS* Part IV). For more on similarities and differences in the moral theories of Hume and Smith, see Martin (1990), Morrow (1923), Raphael (1972/3), and Ross (2010).

[135] In this context, Smith seems to use "reason" to refer strictly to theoretical reason, not to practical reason.

displayed in that system, which is formed from the "love of system", from the regard to "the beauty of order, of art and contrivance" (*TMS* 4.1.1.11).

The different sources of our moral judgments indicates that there is no one particular distinct faculty responsible for our moral judgments. We may *in the first instance* have to rely on the sympathetic mechanism to form moral judgments. However, there are several different faculties that can give rise to our moral judgments. 'Moral faculties' is referring to *all* of these different faculties that give rise to moral sentiments, including imagination, reason, and love of systems. For instance, *reason,* which forms general rules through inductive inferences, also forms moral sentiments by deducing what is proper through these general rules. Thus, it should not be too surprising that Smith gives authority to moral faculties in the chapter titled "Of the influence and authority of *general rules*" (*TMS* 3.1.5). Smith gives authority to all faculties that gives rise to moral sentiments, including reason, which does so by means of formulating general rules. He does not give authority exclusively to the sympathetic mechanism, or conscience, or general rules; rather, he gives authority to *all* of the faculties responsible for them.

4.3 Distancing Smith from Butler: Smith on "Authority"

The most crucial difference between Butler and Smith is, I argue, lies in their views on what authority is. In *Sermons*, Butler stresses the distinction between authority and mere power. Power is the motivating force of the various parts of human nature. If the passions exert more force than reason or the appetites, then the agent will act on those passions. For Butler, authority is distinct from motivating force. Authority demands obedience, regardless

of the beliefs and desires of the person commanded. Thus, for Butler, reason is the faculty that ought to be obeyed, even if the other faculties disregards, or overpowers, reason.

Smith, I believe, does not endorse this sharp distinction between power and authority. For Smith, authority is a special kind of motivating force, not something distinct from it. Authority is the power to procure obedience from others through subjugation. This subjugation is not necessarily achieved with threats or coercion, but rather through the natural sentiment of respect, i.e. reverence in recognition of some distinction in rank. So authority is essentially dependent on the subjugated, because authority only exists if the subjugated acknowledges the distinction of rank and experiences the accompanying desire to obey those of higher rank. Authority is *given* by the subjugated through the natural desire to obey those of higher rank; it is not taken by the authority figure through force. So in claiming that conscience has authority, Smith is merely claiming that conscience has the capacity to procure obedience from the agent through the agent's own recognition that conscience is of a higher rank than the faculties that give rise to thoughts, motives and actions impugned by conscience and the agent's consequent desire to obey conscience (or repent for past actions) in response to that recognition.

My interpretation of Smith on authority does not find any clear support in *TMS*, in which Smith rarely uses the term 'authority', much less describe it in any detail.[136] However, Smith provides an extensive discussion of authority in *Lectures on Jurisprudence* and *Wealth of Nations*. In *LJ*, Smith argues that government does not obtain its authority through the

[136] Smith only uses the term 'authority' on a few occasions in *TMS*, most notably the chapter titles "Of the influence and authority of conscience" (*TMS* 3.1.3) and "Of the influence and authority of the general rules of morality, and that they are justly regarded as the laws of the Deity" (*TMS* 3.1.5). Interestingly, he only speaks of the influence of fortune and the influence of custom and fashion. Yet, it becomes clear in his discussions in *WN* and *LJ*, as we shall see, that at least fortune does confer authority.

forming of contracts (*LJA* v.116-119).[137] Instead, our obligation to the government, i.e. the "principle or duty of allegiance" is founded on two principles: the "principle of authority" and principle of "common or generall interest" (*LJA* v.119).[138] Smith describes the principle of authority as follows:

> [W]e see that every one naturally has a disposition to respect an established authority and superiority in others, whatever they be. The young respect the old, children respect their parents, and in generall the weak respect those who excell in power and strength. Whatever be the foundation of government this has a great effect (*LJA* v.119).

Authority is founded on a natural disposition to *respect* "established authority" and the "superiority in others".[139] Respect, for Smith, is the reverence we naturally feel towards others with certain superior qualities. It is a certain regard we give others in recognition of their superiority in these respects. It is through respect of what is represented as superior in some way that authority is established. An older person has authority in virtue of the greater *respect* that she is given on account of her supposed greater wisdom (*WN* 5.1.2.7). A commander has authority over a military officer because the officer acknowledges the

---

[137] In this section, Smith presents several cases that can be construed as a contract between the government and its citizens. I present one notable case: "Again, of all the cases where one is bound to submit to the government that of an alien comes the nearest to a voluntary or tacit contract. He comes into the country not asleep but with his eyes open, inlist(s) himself under the protection of this government preferably to all others; and if the principle of allegiance and obedience is ever founded on contract it must be in this case. *Yet we see that aliens have always been suspected by the government, and have always been laid under great dissabilities of different sorts and never have any trust or employment in the state; and yet they have shewn more strong and evident signs of an inclination to submit to the government than any others; and the obligations they are under to obedience are to those of a native subject as that of one who voluntarily enlists into the fleet compared to that of a pressed man. So that upon the whole this obedience which every one thinks is due to the sovereign does not arise from any notion of a contract*" (LJA v.118-119, emphasis added).

[138] For our purposes, we need not be concerned with the principle of "common or generall interest". For more on Smith's principle of "common or generall interest" and its relation to the principle of authority, see Khalil (2005, 63-65).

[139] "Established authority" refers to those who are in a higher position within an established system of rank, e.g. king, lords. "Superiority in others" refers to four sets of superior qualifications, outlined in *WN*: [1] "the superiority of personal qualification, of strength, beauty, and agility of body; of wisdom and virtue; of prudence, justice, fortitude, and moderation of mind"; [2] "the superiority of age"; [3] "the superiority of fortune"; and [4] "the superiority of birth" (WN 5.1.2.5-9).

superiority of the commander with regard to both military prowess (a "natural" basis for respect) and institutional power (a form of respect arguably grounded in the soldier's self-interest or fear of punishment). This authority depends upon or constitutively involves the respect the officer affords his commander. When the authority is natural, or unquestioned, the officer "submits, without reluctance, to the authority of a superior by whom he has always been commanded" (*WN* 5.1.2.9). Because a person has authority in virtue of the respect that others give her, authority is ultimately dependent on the *responses* of people to those they know or believe to have superior qualifications:

> A man of rank and fortune is, by his station, the distinguished member of a great society, who attend to every part of his conduct, and who thereby oblige him to attend to every part of himself. *His authority and consideration depend very much upon the respect which this society bears to him*. He dares not do anything which would disgrace or discredit him in it; and he is obliged to a very strict observation of that species of morals, whether liberal or austere, which the general consent of this society prescribes to persons of his rank and fortune (*WN* 5.1.3.12, emphasis added).

A person with authority "attend[s] to every part of his conduct", and "dares not do anything which would disgrace or discredit him" because in doing so, he would lose the respect of others, and so lose his natural authority. In the case of the military officer, "when another family, in whom [the officer] had *never acknowledged any such superiority*, assumes a dominion over them" (*WN* 5.1.2.9, emphasis added), the officer does not obey, or at least does not obey out of *respect* for the family.

There are two kinds of respect, and two corresponding kinds of authority: *natural authority* and *derived authority*.[140] Natural authority originates from our natural dispositions to form distinctions in rank and respect people of certain qualities, like the distinction in age

[140] These are my terms, not Smith's.

and in virtue. We do not need some established system of rank for one to have natural authority over another. For instance, Smith recognizes the natural authority that teachers, in virtue of their "superior virtue and abilities", have over their pupils:

> If the opinion of their own utility could not draw scholars to them, the law neither forced anybody to go to them, nor rewarded anybody for having gone to them. The teachers had no jurisdiction over their pupils, nor any other authority besides that natural authority which superior virtue and abilities never fail to procure from young people towards those who are entrusted with any part of their education (*WN* 5.1.2.44).

The authority granted teachers is distinct from any political authority they might be granted over their students by the law or any benefits students imagine they can receive through learning from the teacher.[141] Rather, it is an authority arising solely from the natural disposition to give respect to those the student know or believe to be more virtuous or skilled than they. Similarly, Smith describes how judicial authority is, in its earlier stages, a kind of natural authority: a "great shepherd or herdsman" who has "natural authority over all the inferior shepherds or herdsmen" can, as a result of this superiority, "naturally procure him some sort of judicial authority" (*WN* 3.1.5.12).

Derived authority, however, originates from our habituation to some established order, like a government or church hierarchy. Unlike natural authority, derived authority involves an acceptance of (and with that respect for) the established ranking, independent of our natural dispositions to rank people higher or lower based on certain qualities. Thus, we

---

[141] We may tend to give greater respect to those with qualities that are more useful to us, e.g. the person of superior wisdom (who can teach us) or the person of superior strength (who can protect us). However, we do not respect others *in virtue of* the usefulness of those qualities, evident in our respect for people who display such qualities but do not benefit us in any way. We might obey the teacher because we find the teacher's instruction beneficial for us, but that obedience does not necessarily mean we respect the teacher or that the teacher has any authority over us. We simply obey because we find it in our self-interest.

can have no natural disposition to respect the lazy, incompetent prince who is the heir to the throne, but may still give that prince derived authority through our recognition of the established order, which ranks the prince above all others except the king.[142] In some cases, there might be a natural authority that is later replaced with derived authority. For instance, though a father has natural authority over his child, the father can be granted additional authority or lose authority by the state:

> The power of the father over his children was at first, in Rome, altogether absolute. His authority extended to three different heads: 1st, he had the power of life and death over his children; 2ndly, that of selling him; and 3rdly, that every thing which was acquired by him was to belong to the father, he being considered as incapable of property. But these powers seem to have been pretty soon mitigated, in the same manner as that which the husband had over his wife (LJA iii.81).

Similarly, judicial authority is initially a kind of natural authority (see discussion above on *WN* 3.1.5.12; also see *LJA* iv.22). Thus, the judge was someone who the disputing parties both respected in virtue of some natural superiority. However, when it becomes more and more difficult to resolve disputes, we simply appoint someone who is now granted authority independent of whether or not the disputing parties respect that particular person. Rather, the respect is given to the established order, which places the judge in the position to resolve differences, and this authority begins to extend beyond any degree of respect that is naturally conferred on the judge (*LJA* iv.15-16).

[142] In the example of the military commander earlier, the military commander might have one kind of authority (natural or derived) without the other. The commander can, for instance, punish the disobedient officer in virtue of the commander's derived authority (as commander), even if the commander lacks any natural authority (e.g. lacks military prowess). Or the officer might obey the commander out of respect for the commander's rank within the military system even if the officer lacks any natural respect for the commander (e.g. questions the commander's judgments).

Though in *WN* 5.1.2 and *LJ*, Smith discusses authority in the context of political systems, he seems to be presenting his more general view of what it is to have authority, both natural and derived. Consider the range of examples he gives in *WN* and *LJ*:

[i] older-younger (e.g. *WN* 5.1.2.7; *LJA* iii.78-87)
[ii] husband-wife (e.g. *LJA* iii.1-87)
[iii] landlord-tenant (e.g. *WN* 1.11.1.36)
[iv] clergy-congregation (e.g. *WN* 5.1.3.12)
[v] educator-student (e.g. *WN* 5.1.2.17; 5.1.2.41; 5.1.2.44)
[vi] master-servant (e.g. *WN* 4.7.1.3; *LJA* iii.87-147)
[vii] ruler-subject (e.g. *WN* 1.4.1.5; 3.2.1.12; 3.3.1.2; 3.3.1.10-12; *LJA* iii.122)
[viii] judge-judged (e.g. *WN* 3.1.5.12)
[ix] mother country-colony (e.g. *WN* 4.7.1.2)

Some of these examples are of political authority (e.g. ruler-subject, judge-judged, mother country-colony) in which the relations between superior and subordinate exist within an established system of governance. However, many of these examples of authority are devoid of any political system or relations. For instance, an older sibling has authority over a younger one, the parent over the child, the elder in the community over the younger, independent of the recognition or existence of any political system (unless families and communities are thought of as polities). In a community without any recognized system of rule, the older person will have authority over the younger because, according to Smith, the younger has a natural disposition to respect the elder. These distinctions arising from distinction in age may contribute to the formation of political systems, e.g. established system of seniority, but they need not, and can operate independently of any political system. The same holds for a range of other relationships: husband-wife, clergy-congregation, educator-student, and virtuous-nonvirtuous. In all these relations, Smith speaks of one having

authority over the other independent of any political order. Moreover, he does not claim that they necessarily lead to the formation of a political order.

Respect is not simply an admiration for the qualities (including that of holding a position within a system of rank) in another. Respect is a motive to obey someone or something out of admiration for the *superiority* one believes another has in the area at issue. Thus, in respecting another, we place the other person in a higher rank than ourselves in the relevant respect.[143] Additionally, in making these distinctions of rank, we place people in an authority-to-subordinate relationship. For instance, Smith writes of the superiority of birth and fortune:

> Birth and fortune are evidently the two circumstances which principally set one man above another. They are the two great sources of personal distinction, and are, therefore, the principal causes which naturally establish authority and subordination among men (WN 5.1.2.11, emphasis added).

Rank is an ordering of individuals from most superior to least. It is not simply recognition of differences in qualifications, but recognition of status or position in relation to others in virtue of those differences in qualifications. To be of a higher rank is to be in a position above another, and the establishing of one person over another just is to establish one person as authority and another as subordinate. There can be multiple authorities of different degrees based on rank. For instance, for common citizens, both the clergy and the civil magistrate can have authority over us, because common citizens recognize the superiority of both. Yet in some cases, the clergy has stronger authority over those citizens than the civil magistrate:

[143] With "plain and palpable quality" like superiority of age, there is "no dispute" about the ranking (WN 5.1.2.7), but with the "invisible" qualities like qualifications of the mind, the ranking is "always disputable, and generally disputed" (WN 5.1.2.6).

> But the authority of religion is superior to every other authority. The fears which it suggests conquer all other fears. When the authorized teachers of religion propagate through the great body of the people, doctrines subversive of the authority of the sovereign, it is by violence only, or by the force of a standing army, that [the prince] can maintain his authority (WN 5.1.3.17).

Though the civil magistrate maintains some authority through the threat of physical force, the clergy are respected more by the people and so are given greater authority.

Furthermore, to place a person in a position of authority is to enter into a relationship of obedience with that person, in which the person of higher rank, i.e. the person of authority, commands and the person of lower rank, i.e. the subordinate, obeys. Thus, in a system of rank, the person of authority demands with an expectation that the subordinate obeys, and the subordinate finds it proper to obey:

> One is born and bred up under the authority of the magistrates; he finds them demanding the obedience of all those about him and he finds that they always submit to their authority; he finds they are far above him in the power they possess in the state; he sees they expect his obedience and sees also the propriety of obeying and unreasonableness of [dis]obeying (*LJA* v.119-120).

The expectation of obedience by both parties is built into the nature of the authority-subordinate relationship because *that the superior person ought to be obeyed* is built into the notion of respect: "Respect and deference to the monarchy, the idea they have that there is a sort of sinfullness or impiety in dissobedience, and the duty they owe to him, are what chiefly influence them" (*LJA* v.121). This is not to claim that the subordinate will always obey. The subordinate has a natural disposition to obey those whose authority they acknowledge, but the degree to which the subordinate will depend on all the other motives present in the

subordinate in a given situation.[144] However, the subordinate will always consider the obedience to a recognized authority proper. Without this consideration, there is no authority-subordinate relationship.

A possible worry for Smith is that we seem to be able to give respect to one another without ranking one above another, without entering into this authority-subordinate relationship. For instance, my colleague and I have a mutual respect for one another, though we both recognize each other to be of more or less equal intellectual abilities, and in no way view our relationship as that between an authority and a subordinate. Similarly, two fierce competitors seem to be able to respect one another without ever recognizing the other as superior in any way.

Smith does not address the possibility of mutual respect, and it is unclear how Smith *would* address it. But given his commitment to the claim that respect involves some recognition of superior quality or rank, one possible response is that our mutual respect might be founded on recognition of *different* superior qualities, or rank along *different* systems. Even if my colleague and I view one another as equal in intellectual capacities (and even experts in the same area of study), I respect my colleague in that I recognize her as capable of providing deep insight that I might not come to myself, and vice versa. Because I recognize this capacity that I may not have, I give my colleague due consideration.[145] We can also account for mutual respect among equally talented competitors in this manner. Two

---

[144] For instance, Smith recognizes utility as one motive for someone to obey another. However, he is clear that obedience out of a recognition of rank is independent of motives of utility. Hence, he writes: Nature would teach us to submit to them for their own sake, to tremble and bow down before their exalted station, to regard their smile as a reward sufficient to compensate any services, and to dread their displeasure, though no other evil were to follow from it, as the severest of all mortifications (*TMS* 1.3.2.3).We obey because of our "admiration for the advantages of [superior people's] situation", i.e. our respect for the other person's higher rank, not because of "any private expectations of benefit from their good-will" (*TMS* 1.3.2.3).

[145] For the same reason, I do not give the same kind of consideration to someone who I believe lacks the capacity to provide any insight beyond what I am already capable of obtaining myself.

competitors might recognize they are equally talented in playing a particular sport, but recognize that they are talented in different ways. Or they might be equally talented in every respect, but they also recognize one another as ranked above all others. Thus, they might recognize each other's superior talents *in relation to others* (and so show each other respect on those grounds), but not in relation to one another.[146] This line or response can also explain how individuals can be in an authority-subordinate relationship in one setting but not in another. For instance, if you and I are friends, but I am also your employer, then I am your authority in the context of a workplace setting, but not your authority outside of work. Likewise, my expectation that you obey, and your recognition that you should obey, only exists within the work setting.[147]

To summarize, in Smith's picture, to say that a person "has authority" or "is in a position of authority" is to claim that such a person stands in a particular relation to those that she has authority over: she has the capacity to procure the obedience of those who are disposed to obey out of respect, out of the recognition of (or belief in) her superiority in some important respect. The superior person may not be able to garner the obedience of the subordinate at all times, but the subordinate is motivated to some degree to obey because the subordinate always recognizes the superior as someone that is to be obeyed (at least on those matters on which the superior is a recognized authority). Thus, authority is a kind of motivating force, one that arises from respect, from the recognition of the superiority of another. More important for our purposes, Smith's recognition of natural authority suggests

[146] Smith might have had such considerations in mind when he speaks of "respect of our equals" and "rank and credit among our equals" in *TMS* 6.1.1.3-4. It is clear that the rank and respect is with regards to distinctions in fortune, character, and conduct. However, it is unclear what Smith means by "equals". One possibility is that they are equals in their recognized position within society, e.g. citizens, or equals in age, which is the way Smith uses "equals" in describing the moral development of children (see *TMS* 3.1.3.22).

[147] My primary aim here is to give a plausible interpretation of Smith on authority. Admittedly, this account of mutual respect needs further explication, and may not accord with what some mean by 'mutual respect' (e.g. respect for one another as human beings).

that authority is not granted solely in virtue of one's rank, position, or role. For Butler, one can have authority in virtue of one's position or role, regardless of the subordinate's beliefs or motivation (e.g. the passions ought to obey reason regardless of whether the passions submit to reason or whether the agent with those passions feels motivated to act in accordance with reason). However, for Smith, authority is inextricably tied to the subordinate's motivations because authority depends on the subordinate's respect and disposition to obey. The authority cannot be an authority without having this particular kind of influence over its subordinate.

4.4 Smith on the Authority of Conscience

As I have interpreted Smith, to have authority is to have a particular kind of power or influence over another: the power to gain the respect of others, which is to have the recognition of rank, i.e. the recognition as someone whom it is proper to obey. Smith, I believe, understands the authority of conscience, as well as all other moral faculties, in the same way. The agent respects her conscience, and in doing so, the agent gives conscience authority, recognizing conscience as something that is to be obeyed. Conscience's authority, then, is akin to the virtuous person's authority. Just as someone gives respect (and so authority) to a virtuous person by ranking the virtuous person as a moral superior, an agent gives respect (and so authority) to conscience by ranking conscience above one's other motives, faculties and thoughts as a moral superior.

Though Smith does not explicitly make any such claim, he presents the correction of moral judgments in a manner that suggests that they occur along a system of rank. In the first three parts of *TMS*, Smith presents the correction of our moral judgments in stages: the agent

corrects her judgments by first appealing to *spectators* (*TMS* 1.1.4), then to more virtuous people (*TMS* 1.1.5), then to *conscience* (*TMS* 3.1.1-3), and finally to *general rules of morality* (*TMS* 3.1.4-5). One way to understand this progression is as a description of moral development, a description of how we rely on different tools at different times in our development to form our moral sentiments. Yet Smith also suggests that some ways of coming to our moral judgments are better than others, where "better" is measured in terms of accuracy of a certain sort. For instance, though no one has ever lived up to the standard of perfect propriety, we are able to recognize that the conduct of some comes closer to that perfection than the conduct of others. Thus, in attempting to judge accurately by appealing to others, we are better off adjusting our sentiments in response to the judgment of the person closer to this perfection than one farther from it. We are essentially recognizing the superiority in character of some over others, including ourselves, when we are attempting to correct for our errant moral judgments. Though we might initially rely on any spectator, with experience we begin to differentiate between the characters of others and begin to identify some as better guides, better exemplars of virtues, for directing our own sentiments.

Smith appears to have something like our appeal to the virtuous person in mind when he speaks of the superiority of conscience as a moral judge. Conscience is generally in a better position to judge accurately than the agent and other partial parties. Because conscience views the situation as an impartial spectator would, it is able to make a "proper" comparison between the interests of different parties and thus come to a more accurate moral judgment:

> [I]t is only by consulting this judge within, that we can ever see what relates to ourselves in its proper shape and dimensions; or that we can ever make any proper comparison between our own interests and those of other people (*TMS* 3.1.3.1).

> It is from him only that we learn the real littleness of ourselves, and of whatever relates to ourselves, and the natural misrepresentation of self-love can be corrected only by the eye of this impartial spectator. It is he who shows us the propriety of generosity and the deformity of injustice; the propriety of resigning the greatest interests of our own, for the yet greater interests of others, and the deformity of doing the smallest injury to another, in order to obtain the greatest benefit to ourselves (*TMS* 3.1.3.4).

Conscience has the advantage of viewing the agent's situation from a position that avoids certain "misrepresentations". As a result of this advantage, conscience is a better judge of what is proper than the agent and any other partial parties.

Furthermore, conscience also has access to information that other spectators do not, especially the agent's sentiments. Whereas others need to infer the agent's sentiments, conscience has more immediate access to the agent's sentiments because conscience and the agent are one and the same person. Hence, Smith recognizes that conscience is more likely to praise what is praiseworthy (what actually deserves praise), than other spectators, even impartial ones who more often praise what does not deserve it. In arguing this, Smith recognizes that conscience is in some ways a better judge of what is proper than other spectators, and even the public-at-large. It is conscience that corrects for the errors of other spectators, not the other way around:

> If the man without should applaud us, either for actions which we have not performed, or for motives which had no influence upon us; the man within can immediately humble that pride and elevate the mind which such groundless acclamations might otherwise occasion, by telling us, that as we know that we do not deserve them, we render ourselves despicable by accepting them. If, on the contrary, the man without should reproach us, either for actions which we never performed, or for motives which had no influence upon those which we may have performed; the man within may immediately correct this false judgment, and assure us, that we are

> by no means the proper objects of that censure which has so unjustly been bestowed upon us (*TMS* 3.1.2.32).

Because of these advantages, conscience generally comes to more accurate moral judgments, and so aligns itself more closely with the virtues. Hence, Smith describes the motive to obey conscience as the "love of what is honourable and noble, of grandeur, and dignity, and superiority of our own characters", as opposed to the love "of our neighbour" or "of mankind" (*TMS* 3.1.3.4). In attempting to attain the virtues, we look to conscience, and not to our neighbors or other people, because we are essentially recognizing conscience as a superior moral guide.

Smith treats the general rules of morality in a similar fashion in relation to conscience. We recognize that even conscience is not immune to perversion (*TMS* 3.1.4.1). Hence, we need a safeguard for when conscience does err. Though limited to matters of justice and action, the general rules of morality can provide these safeguards, preventing us from forming inaccurate moral judgments as a result of our heightened passions in the given moment (*TMS* 3.1.4.7; 3.1.4.12). Again, we do not arbitrarily rely on general rules as one of many ways of forming our moral judgments. We use general rules as a means for correcting "the misrepresentations of self-love concerning what is fit and proper to be done in our particular situation" because we recognize it as superior to conscience in forming moral judgments. And unlike in his discussion of conscience, Smith is explicit that the influence of the general rules is rooted in our *respect*, our *reverence*, for them:

> [T]hat *reverence* for the rule which past experience has impressed upon him, checks the impetuosity of his passion, and helps him to correct the too partial views which self-love might otherwise suggest, of what was proper to be done in his situation. If he should allow himself to be so far transported by passion as to violate this rule, yet,

> even in this case, he cannot throw off altogether the awe and *respect* with which he has been accustomed to regard it (*TMS* 3.1.4.12, emphases added).

More importantly, these rules have influence over us because *we* give them authority by recognizing them as proper and just. *We* are the ones laying down these rules for ourselves as “inviolable”, “to abstain from them upon all occasions” (*TMS* 3.1.4.12). And it is *our* regard for these rules that constitute our very sense of duty or obligation toward those rules (*TMS* 3.1.5.1). Furthermore, we give these rules greater respect than we give conscience. Unlike conscience, the general rules are given “sacred regard”, and our reverence for them is “still further enhanced by an opinion… that those important rules of morality are the commands and laws of the Deity” (*TMS* 3.1.5.3). Thus, we regard the general rules as superior to conscience, that when the two come to competing judgments from the two, we regard the general rules to be that which we are to obey, not conscience, because we give the general rules greater respect.

What we find in Smith, then, is that we do not treat all means of forming moral judgments in the same way. We recognize some moral judges as better than others. We do so with others, differentiating them based on how closely they come to the standard of perfect propriety. We also do so with the different ways we form our own moral judgments: general moral rules are superior to conscience, and conscience is superior to ourselves as agents. In forming these distinctions, we are recognizing rank in the ways we form moral judgments. And as I argued earlier, to recognize rank is to respect, and the higher the rank, the greater the respect. Thus, the greatest respect, as Smith emphasizes, goes to the general rules of morality, and the next greatest to conscience.

To be clear, my claim is not that certain means of correcting our moral judgments are actually better than others. Rather, my claim is that *we as agents* acknowledge a ranking in the different ways of forming and correcting our moral judgments. Impartial spectators are superior to partial ones. Conscience is superior to other spectators. General rules of morality are superior (in matters of justice and action) to conscience. We do not correct our moral sentiments that arise through these different means arbitrarily. Rather, we appeal to some to correct others because we recognize some as superior to others in the realm of moral judgment. Conscience falls within this ranking. For Smith, it is superior as a moral judge to other spectators, both partial and impartial, but not superior to the general rules of morality on matters of justice and action. My suggestion, then, is that Smith treats conscience as having the same kind of authority over the agent that the virtuous person has over those who respect her. Just as we recognize the virtuous person as a superior moral judge, we recognize our conscience as a superior moral judge. We recognize conscience as having higher rank than other spectators and ourselves as agents. In doing so, we give conscience respect, and so authority over us. As with all other authority, the degree of authority depends on the degree of respect given. The more we respect conscience, the more authority conscience has over us (and the more influence we give it).

4.5 Why Should I Obey My Conscience?

Thus far, I have presented Smith's account of what conscience's authority consists in. Though interesting in its own right, we are ultimately interested in giving an explanation of what this authority is because we want to identify what *justifications* there are, if any, for obeying conscience. We want to know whether we should obey our conscience, and if so,

why we ought to do so. This question of justification *Why should I obey my conscience?* is ultimately what Bonaventure, Aquinas, and Butler were attempting to address.[148] Their accounts of the nature of conscience's authority were merely aimed at addressing why we should always obey conscience.

Though Smith does not seem interested in addressing the question of justification, his account does provide one response to this question: *we ought to obey conscience insofar as we respect conscience, i.e. insofar as we recognize conscience as a superior*. The very authority that conscience has is dependent on the degree to which we respect conscience. To respect something is to subordinate oneself to another, to *make* another a superior over oneself.[149] To enter into such a relationship is nothing more than to recognize another, i.e. the superior, as something to be obeyed. Hence, in virtue of respecting conscience, we have already settled the matter of whether or not we ought to obey conscience: since we respect conscience, we view it as that which is to be obeyed.

The question of justification, then, simply does not make sense for the agent who already respects conscience. Since the agent already gives respect to conscience, she has already subordinated herself to conscience as something to be obeyed. The reason for obeying conscience thus is embedded in the sentiment of respect. We have an obligation to obey conscience because *we* oblige ourselves to conscience by respecting it. The question of

---

[148] I am drawing on Korsgaard's discussion of the normative question "Why should I be moral?" in *Sources of Normativity* (Korsgaard 1996, 9). According to Korsgaard, this question is seeking some justification, not description, for being moral. Similarly, the question "Why should I obey my conscience?" is seeking a justification for our acting as conscience demands. A mere description of how conscience operates alone will be insufficient. Hence, Bonaventure and Aquinas tried to ground our reason for obeying conscience in our commitment to God's commands, and Butler in our commitment to our human nature.

[149] As noted earlier, it is possible to respect another along one particular quality or system of rank, but not in another, and so give another authority in some respects (or in certain contexts) but not in others. The difficulty with conscience is that it is recognized as a superior in conscience's only role: that of a moral judge. Thus, it seems less likely, though not impossible, that the agent and conscience can have mutual respect for one another.

justification only makes sense for someone who has already lost respect for conscience.[150] This is, I believe, what is occurring when we accept Freud's account of conscience. If Freud is correct and conscience is nothing more than an internalization of parents and society, then we have respect for conscience only to the degree that we have respect for parents and society. If we do not respect our parents and society, then we have no respect for conscience. Conscience, in that case, seems to have little weight in our moral life; we do not feel obligated to it at all, we do not appeal to it for any moral guidance, and we dismiss its judgments as nothing more than a feeling or product of experience. It is for such a person that the question "Why should I obey conscience?" makes sense, and for whom the question needs an answer.

It is possible that the agent obey conscience for reasons other than respect. For instance, an agent might recognize that her conscience issues judgments that often align with the judgments of the general public. Because the agreement of the general public can provide various advantages (e.g. in business transactions, in avoiding public humiliation or condemnation), the agent might act in accordance with conscience to gain those advantages. However, there is a difference in obeying conscience out of respect and obeying conscience for pragmatic reasons. When we obey conscience out of respect, we are recognizing conscience as a superior. But we do not necessarily have to give any such recognition to conscience when we are acting for pragmatic reasons. Because there is no such recognition, the agent is not obeying conscience *as an authority*. And once we exclude conscience's authority from the question of justification, then we are not providing the reason for obeying conscience that is central to our experience of conscience. Conscience is no longer presumed

[150] H. A. Prichard defends this view in response to the question "Why should we be moral?" nicely in "Does Moral Philosophy Rest on a Mistake?" *Mind*, 21 (1921): 21-37.

to be something that is to be obeyed, but rather reduced to a tool whose instrumental value must first be assessed before we have any reason for obeying it. This is not to deny that these instrumental reasons can serve as a justification for obeying conscience. What I am claiming is that these instrumental reasons fail to account for the kind of justification that seems most important to our experience of conscience: we *take it for granted*, we *presume*, that conscience is something to be obeyed. We generally do not obey conscience because we find it to be useful, nor do we obey it because we want to avoid feelings of guilt. We do not appeal to such reasons because such reasons do not map onto our experiences of feeling obligated to conscience. So though these instrumental reasons may give some reason for obeying conscience, they fail to capture the kind of reason that is central to our relationship with our conscience: conscience presents itself as something to be obeyed in virtue of what it is, not in virtue of what it is useful for.

4.6 Problem for Smith: Conscience as *The* Moral Authority

In many ways, Smith's account fares well when compared with our experience of conscience. We experience conscience as a moral guide, as that which directs us to the proper sentiment and action. For Smith, we view conscience as a moral guide because we recognize it as a better moral judge, and so rank it above us such that we appeal to it as a source of more accurate moral judgments. Conscience also speaks to us in different volumes: sometimes, it is a strong motivating force while at other times it is a mere whisper. For Smith, the difference in volume is accounted for by the degree of respect we give conscience. The weak person who does not give conscience much respect experiences a much quieter (sometimes absent) conscience, while the virtuous person who gives conscience the utmost

respect experiences the pangs of conscience more forcefully or with greater frequency.[151] Finally, even when we feel we ought to obey conscience, we often fail to do so. For Smith, we can and do disobey conscience because conscience is simply one of a range of motivating forces. Sometimes, our self-love might be a stronger motivating force than conscience, especially if we do not give conscience much respect. Other times, we might give greater respect to something other than conscience, like the general rules of morality.

Still, there is one key feature of our experience of conscience that Smith fails to capture: conscience is often seen as *the* moral judge, not merely a better judge. This feature, I believe, is at the core of our experience of conscience as authoritative. When we speak of conscience as having authority over us, we do not simply mean that conscience is a strong motivating force, or that it is a useful moral guide. Rather, conscience makes a demand on us that we have difficulty ignoring. It is not simply someone we respect, like a virtuous person, issuing a command; it is the voice of God or of moral law. It is not simply *a* moral guide; it is *the* moral guide. Though we might, upon reflection, come to question the accuracy of conscience's moral judgments, we are compelled to accept conscience's judgments as accurate *in the moment we experience conscience*. In this way, conscience *imposes* itself on us, and we have difficulty denying its commands, even when we ultimately decide to act against it. Thus, as Bonaventure and Aquinas noted, whenever we act in the belief that we are doing something wrong.

This feature of conscience is altogether absent in Smith's account of conscience's authority. Because Smith grounds authority in the sentiment of respect, conscience does not have this special place as the moral authority except for those who give conscience the

[151] Freud also pointed out that the virtuous person has a much stronger, louder conscience, and as a result, feels guilt and remorse with greater intensity than someone who is far from virtuous (Freud 1930).

highest respect. Yet as even Smith recognized, there are few people who hold conscience in such high regard that they are able to never "forget for one moment the judgment which the impartial spectator would pass upon [their] sentiments and conduct" (*TMS* 3.1.3.25). The common person gives conscience some respect, but not the highest degree, and so is not as strongly influenced by it as he might be (*TMS* 3.1.3.23-24). Furthermore, Smith claims that we general give the highest respect ("sacred regard") to the general rules of morality, not conscience, and that it is these general rules that are elevated to the status of the commands of God:

> This reverence is still further enhanced by an opinion which is first impressed by nature, and afterwards confirmed by reasoning and philosophy, that those important rules of morality are the *commands and laws of the Deity* (*TMS* 3.1.5.3, emphasis added).

So even though most adults give conscience some respect, this respect, and the corresponding authority, is not elevated to the status of the voice of God. Its authority falls below the authority of the general rules of morality. So in moments of conflict between the judgments of conscience and these general rules, we take the general rules to be *the* authority on what is proper, and disregard conscience as errant.

Moreover, under Smith's account, we seem to be able to deny conscience's demands on us more easily than we generally experience. Since conscience's authority is grounded in our sentiment of respect, we can break away from our obligation to conscience by losing respect for it. We can essentially take away conscience's authority by changing our sentiments toward it. This process seems possible when we are observing the conscience of others. But we have a much more difficult time doing so with our own conscience. This is evident, I think, when there is a conflict between our conscience and the general rules of

morality. Suppose I come across a serial child rapist who is so clever that he is able to get away with his crimes every time. I have that person cornered, and am deciding whether or not to kill him. I might hold a general rule, e.g. do not murder, in high regard. Yet in that moment, my conscience judges that it is proper to kill this person, given the kind of harm he will continue to inflict. Setting aside whether my conscience is erring or not, I will have a difficult time convincing myself that conscience is mistaken, even if I decide to follow the general rule. Conscience continues to torment me, even though I try to convince myself that I did what is proper by following the general rule. This torment is not simply feelings of guilt or remorse; this torment is rooted in a person's difficulty if not inability to accept the propriety of what she has done in acting against conscience, is proper. However, if we decide to act in accordance with conscience, even though doing so breaks the general rule, we have a much less difficult time convincing ourselves that what we have done is proper. We might still feel some doubt as to the propriety of what we did, but we do not experience the same kind of torment that we would experience when violating conscience.

My point is that we have difficulty denying the superiority of conscience to even the most widely endorsed rules of conduct. We treat it as *the unique* moral authority, as the voice of God, as the final judge of what is proper, in a way that we do not treat any other moral guide. We can more easily disagree with the judgments of a virtuous person or even a highly regarded moral rule than we can with conscience. Conscience seems to hold a special place as the moral authority, and its place as this authority does not seem to be a matter of degree. There is something unique about our relationship with our own conscience that makes it difficult for us to disassociate ourselves from its judgments.

4.6 An Alternative: The Authority of Conscience as Commitment to Self

There is, I believe, an alternative account of the authority of conscience present in Smith's theory, though not endorsed by Smith himself, that can account for our experience of conscience as *the* moral authority. According to Smith, conscience is our judging ourselves as an impartial spectator. Central to his account is that conscience's judgments are our own. These judgments are formed by the exercise of *our* faculties and principles, not those of any other. They are *our* judgments, not the judgments of others. Hence, we are committed to conscience's judgments in a way that we are not committed to others' judgments. When another person forms a moral judgment, we have yet to agree or disagree with that person's judgment. However, if *we* form a moral judgment, we are, *in forming that judgment*, agreeing with it because we cannot see our own faculties as anything but proper in our exercising those faculties. It is this blind spot that we have to our own faculties that leads to self-deception. And it is this blind spot that leads to our inability to see conscience's judgments as anything but correct.[152]

This is not to say that we cannot recognize the errors of our own faculties. For instance, when we are presented with repeated examples of how others' consciences have erred, we might come to doubt the reliability of our *own* conscience as a moral judge

[152] I have been influenced here by the works on first-person authority by Korsgaard (in ethics) and Moran (in epistemology). However, my claim that we are committed to our own faculties is not the stronger Kantian claim that both Korsgaard and Moran make. For Korsgaard, I commit myself to my judgment of propriety by taking a deliberative stance, i.e. asking myself "What should I do?", and settling the question of what to do by deciding on what to do (Korsgaard 1996). Similarly, for Moran, I commit myself to my beliefs by taking a deliberative stance, i.e. asking myself "What should I believe?", and settling the matter by forming a belief (Moran 2001). I am claiming neither. I am reiterating Smith's claim that we have no other way to judge other faculties than by using our own: "Every faculty in one man is the measure by which he judges of the like faculty in another…. *I neither have, nor can have, any other way of judging about them*" (*TMS* 1.1.3.10). We are limited by our faculties as a matter of empirical fact. This limitation includes the formation of beliefs and moral judgments of others and of ourselves. The commitment I am referring to, then, is not to ourselves in virtue of taking this deliberative stance, but rather to our faculties in virtue of our having used these faculties to form our beliefs and judgments. Thus, I have the same commitment to my passive sentiments arising from my faculties about which I cannot deliberate.

(because we recognize that our faculties are susceptible to the same errors as those of others). Jonathan Bennett presents a range of cases in which we might do so: the conscience of Heinrich Himmler, Jonathan Edwards, and Huck Finn. All three people have conscience that issued (what we now know or believe to be) incorrect moral judgments, i.e. they were all in the grips of a "bad morality".[153] Similarly, we can question our own conscience if we reflect back on our conscience's prior judgments and find that we now disagree with most of them. Under such considerations, we might begin to question the accuracy of our own conscience as it operates now. Moreover, this phenomenon seems quite common, evident in the little credence that some people give to conscience.

Smith also recognized that we disagree with our own faculties, for instance, when we adjust our sentiments because we find them improper in light of disagreement with others (*TMS* 1.1.4.8; 3.1.1.3). We also find our visual faculties inaccurate at times in measuring the size of objects (*TMS* 3.1.3.2). We can also recognize and be "mortified" when we realize that we have, though unintentionally, misled someone because of some error in our judgment, memory, or reasoning (*TMS* 7.4.1.27). If these errors occur frequently, then we might also lose confidence in the faculty, as we find it to be unreliable. For this reason, we sometimes seem to appeal to others rather than our conscience:

> The agreement or disagreement both of the sentiments and judgments of other people with our own, is, in all cases, it must be observed, of more or less importance to us, exactly in proportion as we ourselves are more or less uncertain about the propriety of our own sentiments, about the accuracy of our own judgments (*TMS* 3.1.2.16).

[153] Bennett describes "bad morality" as "a morality whose principle I deeply disapprove of. When I call a morality bad, I cannot prove that mine is better; but when I here call any morality bad, I think you will agree with me that it is bad; and that is all I need" (Bennett 1974, 123-4). Himmler judged it proper to exterminate the Jews. Huck judged slavery of some to be proper. Edwards thought it proper to torture sinners for eternity. As was with Bennett, for my purposes, all I need is that in all these instances we would find these moral judgments (of Himmler, Huck, and Edwards) to be incorrectly formed by their consciences.

Smith admits that we can be more or less certain, more or less confident, in the accuracy of our judgments of ourselves, i.e. conscience's judgments. We can doubt the accuracy and reliability of conscience in the same way we can with all our other faculties. If so, then it seems we do not have such a strong commitment to our own faculties. We can disregard it if we find them to be mistaken or unreliable. So it might seem that Smith may have been correct to say that the authority of conscience depends on the respect we give it. Once we lose respect for conscience, say by frequently disagreeing with its judgments, conscience loses its authority.

I accept the possibility that we do not accept our own faculties to be accurate. Furthermore, the variance in our confidence in our own judgments can account for how conscience seems to speak to us in different volumes. When we are confidence in conscience's judgments, we have a more difficult time denying that we ought to obey conscience and are more willing to obey conscience in the face of public opposition.[154] However, when we do not have great confidence in conscience's judgments, we tend to, as Smith claims, rely on other people or other means for confirmation. Conscience's judgments, in such cases, are given little weight, and we might be more willing to disregard and disobey conscience's judgments as mistaken (*TMS* 3.1.2.24).

However, there is another sense in which we are deeply committed to the accuracy of our faculties. Though we might be able to question the accuracy of conscience in our reflecting on it, we do not do so *in our exercising of them.* Consider the kinds of cases in which we come to doubt our conscience. We doubt our conscience when we face disagreements with others and are now prompted to reconsider our initial judgment. We

---

[154] Smith provides a parallel case of mathematicians who "have the most perfect assurance, both of the truth and of the importance of their discoveries" and so are "very indifferent about the reception which they may meet with from the public" (*TMS* 3.1.2.20).

doubt our conscience when we find that this faculty tends to be unreliable in others or in ourselves, evident in prior (errant) judgments. We doubt our conscience when we accept that our conscience originates from an unreliable source (e.g. if I adopted Freud's account of conscience and believed my parents and society to be vicious people). In all such instances, we are *reflecting on* and *reevaluating* our *prior* moral judgments. We are not committed to our own faculties, but more specifically, we are not committed to our own faculties as they operated in forming those prior judgments. However, we still retain a commitment to our own faculties, but this commitment is to our faculties in *reevaluating* our prior judgments or current inclinations, not to our faculties that gave rise to these prior judgments or current inclinations. We are only forced to regard our faculties as operating properly *as we exercise those faculties*, not in the operations or outputs of these faculties upon which we reflect. Consider Smith's discussion of our initial judgments of the size of objects at a distance:

> As to the eye of the body, objects appear great or small, not so much according to their real dimensions, as according to the nearness or distance of their situation…. In my present situation an immense landscape of lawns, and woods, and distant mountains, seems to do no more than cover the little window which I write by, and to be out of all proportions less than the chamber in which I am sitting (*TMS* 3.1.3.2).

I find my initial judgment of the size of objects to be proper. But then I reevaluate that initial judgment by "transporting myself, at least in fancy, to a different station… and thereby form some judgment", some *new* judgment, of their proportions. We have essentially exercised our faculties again from a different situation, and in doing so, we *replace* our prior judgment. We *replace* because we are committed to the new judgment, not the old; we deem the new judgment proper and the old not. Smith parallels this kind of reevaluation with our appeal to conscience. We as agents initially form a judgment that we believe to be proper, even if it is

distorted by the "selfish and original passions of human nature" (*TMS* 3.1.3.3). We then *reevaluate* that judgment and replace it with the judgment we formulate from the situation of an impartial spectator, i.e. as conscience. We find our judgment as agent to be improper because we are now committed to a contrary judgment we have formed as conscience. Because we override our judgments as agent in our reevaluating them, we view conscience (and not the agent) as *the* moral authority.

Still, even if we have such a commitment to conscience in our exercising of it, we are able to reevaluate *conscience's* judgments by judging as agent. After conscience issues a judgment of what is proper, we as agent need to decide what to do in light of conscience's judgments, just as we do in light of other spectators' judgments. Hence, we are able to ask whether or not we should obey conscience at all.[155] And when we are faced with evidence that conscience is not a reliable source of moral judgments or that it does not judge correctly in this particular situation, we are genuinely questioning whether conscience is a moral authority (much less *the* moral authority), whether its commands are something that we should obey.

I think this picture is mistaken. Though I do not deny that we *can* view our conscience as something other than *the* moral authority, we can only do so when we disassociate ourselves from our own conscience. In other words, to doubt our conscience, we need to view it the same way we view the conscience of others: as something distinct from us, who doubt conscience. To illustrate what I have in mind, consider how we judge

[155] Generally, this kind of situation is presented as a conflict between our knowledge of what is proper and our passions. For both Butler and Smith, we might accept conscience as judging correctly, but struggle about what to do because we are motivated by other passions, namely self-love. Bonaventure and Aquinas saw these cases as cases of *akrasia*, in which we know what is right to do, namely what conscience judges to be proper, and yet do otherwise because of our other desires. But understood as such, there is no doubt about conscience's authority. Conscience is still viewed as *the* moral authority, and the agent as unable (or unwilling) to obey that authority because of the strength of other motives.

ourselves in the standard case. We divide ourselves into spectator and agent, and we as spectator form a moral judgment of us as agent. When we as spectator judge us as agent to have felt or acted improperly, we are claiming that the agent's judgment about what is proper was mistaken. It might seem as if what I am describing is a case in which we as conscience are disassociating ourselves from our faculties as agents. As Smith describes, we as conscience are judging us as agents just as we would judge *others* (*TMS* 3.1.1.6). However, there are two different ways in which we as spectator can view ourselves as agents. The standard case is one in which we view the agent as one with us in this special relationship that is distinct from our relationship with friends, family, colleagues, and strangers. We view the agent's sentiments and actions as *our* sentiments and actions. In this sense, we as spectator *identify* or *associate* ourselves with the agent. Alternatively, we can, with some effort, view the agent as distinct from us, on par with a friend, a family member, a colleague, or a stranger. The agent is someone we used to be, or not who we *genuinely* or *authentically* are (e.g. when under the influence of alcohol or under extremely stressful situations). When we view the agent in this manner, we are *disassociating* ourselves from the agent. The same distinction arises when we as agent are responding to our conscience. We can associate ourselves with our conscience, viewing it as one with us, or we can disassociate ourselves from our conscience, viewing it as distinct from us.

When we consider cases in which we doubt conscience, it seems we are able to do so *because we disassociate ourselves from our conscience*. For instance, when we doubt our conscience because we doubt the conscience of others, we are viewing our conscience the way we view the consciences of other people: as the operations of faculties distinct from our own. Similarly, when we question our conscience based on how our conscience operated in

the past, we are viewing our past conscience as distinct from who we are now. That past conscience was who I used to be, before I took ethics courses and trained my conscience. When we associate with conscience, it becomes much more difficult to deny that conscience is *the* authority. Why is this so? In associating with conscience, we are accepting the conscience's judgments *as our own*, as the operations of *our* faculties. And as I stated earlier, all of our faculties appear to us to be proper. Thus, in associating ourselves with conscience, we are essentially accepting conscience's judgments to be accurate; we are accepting conscience as *the* moral authority.

If I am correct, then it makes no sense to ask whether we should obey our conscience because in our associating ourselves with our conscience, we already recognize conscience as *the* moral authority, as *the* proper moral guide, as that which we ought to obey. Questioning whether we should obey conscience only makes sense once we disassociate ourselves from conscience. However, when we disassociate ourselves from conscience in this way, we are no longer describing the kind of phenomenon that is of interest to us. Once we view conscience on par with the conscience of others, we do not experience conscience as "pangs" from within or the voice of God in us because we have lost the connection to conscience that makes it a part of *us*. In a way, we no longer experience conscience.

What about cases in which we lack confidence in our own judgments? Such instances are not ones of reflection or reevaluation of the prior exercises of our faculties. Rather, we seem to lack confidence in our faculties *in the midst of our exercising them*. We form a moral judgment as conscience, but lack very much confidence that we have judged correctly (*TMS* 3.1.2.16). In those moments, we seem to lack much, if any, commitment to conscience, and are open to conscience's being mistaken.

There is, however, a way to address this worry. When we lack confidence in our judgments of ourselves, we are not questioning the accuracy of conscience in the way that we do when we evaluate our conscience's prior judgments. Instead, we are *unsure about what our conscience's judgment is*. As Smith puts it, we seek out the opinions of others in these cases because we are "more or less uncertain about the propriety of our own sentiments" (*TMS* 3.1.2.16). Our "sentiments" here is our sentiments as agents, which we are attempting to assess. So when Smith claims that we are unsure about "the accuracy of our own judgments", he means that we are unsure about whether or not our sentiments as agents were proper and to what degree. We are not unsure in the same way that we are unsure about our prior judgments. When we are unsure about our prior judgments, we are not questioning what our prior judgments were. Our uncertainty is whether or not that judgment was accurate. However, when we are unsure about the propriety of our own sentiments, we are unsure about what to judge in the first place. In other words, we are unsure what conscience's judgments are, what our faculties determine to be the proper sentiment. Thus, we do not lack commitment to our own conscience, but rather, we simply do not know what our conscience judges so there is no judgment to commit to.

My suggestion, then, is that our commitment to our own faculties is a commitment to the *present exercising* of those faculties. When judging as agents, we are committed to the accuracy or excellence of our faculties as agents, finding our sentiments to be proper even when they are distorted by self-love. When judging as conscience, we are committed to the accuracy or excellence of our faculties as conscience, not to our faculties as agent. We cannot have a divided commitment to self because we can only exercise our faculties from one perspective at a time. We cannot exercise our faculties simultaneously as both agent and as

conscience. Moreover, this account can explain why it is that we experience conscience as *the* moral authority. Though we as conscience and we as agents are equally committed to our own faculties, there is an important difference: we generally appeal to conscience to reevaluate the judgments of the agent, but we generally do not appeal to the initial reactions we experience as agents to reevaluate the judgments of conscience. When we appeal to conscience, we are committed to the accuracy of conscience *over that of the agent*. However, when we are agents, we are not committed to the accuracy of our judgments *over* that of conscience; we simply form and are committed to our judgments. This difference is not a difference in our respect for us as agents and for us as conscience. Rather, the difference is rooted in conscience's being the means by which we evaluate ourselves as agents. For Smith, we cannot evaluate ourselves except through this imaginative exercise of entering into the situation of another, whether it is actual spectators or conscience. This process is similar to Hume's account of how we correct our moral judgments through reflection, through imagining ourselves in the situation of other spectators to evaluate ourselves as agents (see *THN* 3.3.1.15, 3.3.3.2). We evaluate ourselves by turning our own faculties in upon ourselves. However, in doing so, we are committing ourselves to our faculties in this reflective process, not to our faculties as they have operated in us in forming those immediate reactions that constitute our perspectives as agents.[156] Moreover, because there is no other

[156] The reflective process I am describing is different from what Korsgaard calls 'reflective endorsement', which she attributes to Hume: "The reflective endorsement theorist tries a new tact. Morality is grounded in human nature. Obligations and values are projections of our own moral sentiments and dispositions. To say that these sentiments and dispositions are justified is not to say that they track the truth, but rather to say that they are good. We are the better for having them, for they perfect our social nature, and so promote our self-interest and our flourishing" (Korsgaard 1996, 91). Reflective endorsement involves an approval of the system as a whole, e.g. system of appealing to conscience, system of belief-formation. But the kind of reflection I am considering involves a commitment to the operation of our faculties in that given moment, not to the system that gives rise to that particular judgment.

means for judging our own faculties, it will always be conscience to which we are committed whenever we are debating between our judgments as agent and our judgments as conscience.

4.7 Huckleberry Finn and Our Commitment to Self

Let me close my discussion by providing an alternative explanation of Huck's acting against his conscience in Mark Twain's classic *The Adventures of Huckleberry Finn.* After a lengthy journey with Jim, Huck decides to help Jim escape *in spite of his conscience*. According to his conscience, he ought not to have helped Jim escape. In his words:

> That was where it pinched. Conscience says to me: 'What had poor Miss Watson done to you [referring to himself], that you could see her nigger go off right under your eyes and never say one single word? What did that poor old woman do to you, that you could treat her so mean? Why, she tried to learn you your book, she tried to learn you your manners, she tried to be good to you every way she knowed how. *That's* what she done' I got to feeling so mean and so miserable I most wished I was dead (*The Adventures of Huckleberry Finn* Chapter 16).

Huck's conscience tells him it is wrong to let Jim escape. However, Huck ultimately decides to act against his conscience. What is interesting is Huck's thoughts and feelings in *justifying* his actions after he acts against conscience:

> "They went off and I got aboard the raft, feeling bad and low, because I knowed very well I had done wrong, and I see it warn't no use for me to try to learn to do right; a body that don't get *started* right when he's little aint got no show — when the pinch comes there ain't nothing to back him up and keep him to his work, and so he gets beat. Then I thought a minute, and says to myself, hold on; s'pose you'd a done right and give Jim up, would you felt better than what you do now? No, says I, I'd feel bad — I'd feel just the same way I do now. Well, then, says I, what's the use you learning to do right when it's troublesome to do right and ain't no trouble to do wrong, and the wages is just the same? I was stuck. I couldn't answer that. So I reckoned I wouldn't bother no more about it, but after this always do whichever comes handiest at the time (*The Adventures of Huckleberry Finn* Chapter 16).

Huck concludes that in this case, he would feel bad regardless of what he did, and so there was no use in considering what conscience demands. He is essentially dismissing the authority of conscience: conscience is not any kind of moral authority and should have no weight in our decisions of what to do.[157]

Jonathan Bennett interprets Huck in this scene as undergoing a conflict between conscience and feelings of sympathy toward Jim.[158] According to Bennett, Huck's conscience is clear in judging it wrong to help Jim escape. But because Huck sympathizes with Jim, Huck decides to help Jim escape: "sympathy wins over morality" (Bennett 1974, 126). Like Bonaventure and Aquinas, Bennett treats Huck's case as a conflict between conscience, which the agent accepts as correct, and sentiments that motivate the agent to act against conscience:

> Huck doesn't weigh up pros and cons: he simply *fails* to do what he believes to be right — he isn't strong enough, hasn't 'the spunk of a rabbit'. This passage in the novel is notable not just for its finely wrought irony, with Huck's weakness of will leading him to do the right thing, but also for its masterly handling of the difference between general moral principles and particular unreasoned emotional pulls (Bennett 1974, 127).[159]

[157] As Bennett points out, Huck repeats this claim that conscience is not to be obeyed or given influence later in Chapter 33: "It don't make no difference whether you do right or wrong, a person's conscience ain't got no sense, and just goes for him *anyway*. If I had a yaller dog that didn't know no more than a person's conscience does, I would poison him. It takes up more room than all the rest of a person's insides, and yet ain't no good, nohow."

[158] By "sympathy", Bennett is referring to "every sort of fellow-feeling, as when one feels pity over someone's loneliness, or horrified compassion over his pain, or when one feels a shrinking reluctance to act in a way which will bring misfortune to someone else" (Bennett 1974, 124). Sympathy, as Bennett understands it, is not the same as moral judgments, including judgments of what is proper, but rather a particular kind of feeling distinct from moral judgments.

[159] Similarly, Anders Schinkel (2011) interprets Huck as experiencing a conflict between moral language and moral feelings. My response to Bennett (to follow) also applies to Schinkel.

The problem with Bennett's interpretation is that Huck's conscience is not so much issuing moral judgments as eliciting feelings of sympathy: sympathy for Miss Watson. Conscience, in this case, is not an application of some general rule or principle about stealing. Rather, conscience seems to present Huck with a series of consideration, namely what Miss Watson had done for him—considerations which seem to call for the kind of gratitude incompatible with aiding the "theft" of her "property." Conscience is making Huck consider features of the situation that he had not considered before, and eliciting feelings of guilt at the thought of harming Miss Watson. But what Huck undergoes when conscience speaks to him is no different from what Huck undergoes when Jim speaks of his plans to free his wife and children. There is, in other words, no difference in what conscience is doing and what Jim is doing. Conscience is not issuing moral judgments while Jim is himself feeling sympathy. Both Jim and his conscience are eliciting feelings of sympathy and feelings of guilt.

Bennett's account does not seem to fully capture the dilemma that Huck faces. Huck is having trouble deciding whether or not to obey his conscience, but this case is not so much a conflict of what he knows to be right and what he desires to do. Huck feels *equally bad* about helping and not helping Jim. Thus, this is not a case of a sentiment overpowering our knowledge of what is right. Rather, what Huck's case demonstrates is how we frequently identify with, and then disassociate ourselves from conscience. Conscience is clear on what he must do: do not help Jim escape for doing so will harm Miss Watson who does not deserve that harm. But Huck also hesitates to act as conscience demands, not simply because he "feels bad", but because his "feeling bad" indicates to him that he is doing something bad. In other words, he is able to step away from himself as conscience and in doing so come to question whether he should obey conscience. He is engaged in a battle with himself: "I was

fidgeting up and down the raft, *abusing myself to myself*" (emphasis added). This abuse was occurring both ways. When Huck associated himself with conscience, he abused Huck as agent. When Huck disassociated himself from conscience, he abused Huck as conscience. And ultimately, when Huck decides that there is no use giving any credence to conscience, Huck can do so only because he is now severing all ties to his own conscience. His conscience is no longer his, but rather something external to himself that influences him, the same way that the words and actions of others impact him.

Works Cited

Andrew, Edward G. 2001. *Conscience and Its Critics: Protestant Conscience, Enlightenment Reason, and Modern Subjectivity*. Toronto: University of Toronto Press.

—— 1999. "Hobbes on Conscience within the Law and without". *Canadian Journal of Political Science* 32: 203-335

Bennett, Jonathan. 1974. "The Conscience of Huckleberry Finn". *Philosophy* 49: 123-134.

Berns, Lawrence. 1994. "Aristotle and Adam Smith on Justice: Cooperation Between Ancients and Moderns?" *The Review of Metaphysics* 48: 71-90.

Braun, Harald and Vallance, Edward, eds. 2004. *Contexts of Conscience in Early Modern Europe, 1500-1700*. New York: Palgrave Macmillan.

Broad, C. D. 1940. "Conscience and Conscientious Action". *Philosophy* 15: 115-130.

Broadie, Alexander. 2010. "Aristotle, Adam Smith and the Virtue of Propriety". *The Journal of Scottish Philosophy* 8: 79-89.

—— 2006. "Sympathy and the Impartial Spectator" in *The Cambridge Companion to Adam Smith*, ed. Knud Haakonssen. Cambridge: Cambridge University Press.

Brown, Vivienne. 1994. *Adam Smith's Discourse*. London: Routeledge.

Butler, Joseph. 1736/2006. *The Analogy of Religion, Natural and Revealed, to the Constitution and Course of Nature* in *The Works of Bishop Butler*, ed. David E. White. Rochester, NY: University of Rochester Press.

—— 1729/2006. *Fifteen Sermons Preached at the Rolls Chapel* in *The Works of Bishop Butler*, ed. David E. White. Rochester, NY: University of Rochester Press.

Campbell, T. D. 1971. *Adam Smith's Science of Morals*. Totowa, NJ: Rowman and Littlefield.

Carrasco, Maria A. 2008. "Adam Smith on Morality, Justice, and the Political Constitution of Liberty". *The Journal of Scottish Philosophy* 6: 135-156.

—— 2004. "Adam Smith's Reconstruction of Practical Reason". *The Review of Metaphysics*

58: 81-116.

Cohen, Rachel. 1997. “The Common Point of View in Hume’s Ethics”. *Philosophy and Phenomenological Reseach* 54: 827-850.

Conscience, n. OED Online. September 2011. Oxford University Press. 28 August 2014 <http://www.oed.com.ezp1.villanova.edu/view/Entry/39460?rskey=Y73I9p&result=1#eid>.

Darwall, Stephen. 1995. *The British Moralists and the Internal “Ought”, 1640-1740*. Cambridge: Cambridge University Press.

—— 1999. “Sympathetic Liberalism: Recent Work on Adam Smith”. *Philosophy and Public Affairs* 28: 139-164.

Evensky, Jerry. 1987. “The Two Voices of Adam Smith: Moral Philosopher and Social Critic”. *History of Political Economy* 19:.447-468.

—— 2005. *Adam Smith’s Moral Philosophy: A Historical and Contemporary Perspective on Markets, Law, Ethics, and Culture*. Cambridge: Cambridge University Press.

Feldman, Karen. 2006. *Binding Words: Conscience and Rhetoric in Hobbes, Hegel, and Heidegger*. Evanston, IL.: Northwestern University Press.

Fleischacker, Samuel. 1999. *A Third Concept of Liberty: Judgment and Freedom in Kant and Adam Smith*. Princeton, NJ: Princeton University Press.

—— 1991. “Philosophy in Moral Practice: Kant and Adam Smith”. *Kant-Studien* 82: 249-269.

Fleischacker, Samuel & Brown, Vivienne. 2010. “Introduction” in *The Philosophy of Adam Smith: The Adam Smith Review*, vol. 5, eds. Vivienne Brown and Samuel Fleischacker. New York: Routledge.

Forman-Barzilai, Fonna. 2010. *Adam Smith and the Circles of Sympathy*. Cambridge: Cambridge University Press.

—— 2006. “Smith on ‘connexion’, culture and judgment” in *New Voices on Adam Smith*, e.d. Leonidas Montes and Eric Schliesser. New York: Routledge.

Freud, Sigmund. 1930/1989. *Civilization and its Discontents*, trans. James Strachey. New York, NY: W. W. Norton & Co.

Fuss, Peter. 1964. "Conscience". *Ethics* 74: 111-120.

Goldman, Alvin I. 1989. "Interpretation Psychologized". *Mind and Language* 4:161-185.

—— 2006. *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford: Oxford University Press.

Gordon, R. M. 1986. "Folk Psychology as Simulation". *Mind and Language* 1: 158-171.

—— 1995. "Sympathy, Simulation, and the Impartial Spectator". *Ethics* 7: 727-742.

Greene, Robert A. 1991. "Synderesis, the Spark of Conscience, in the English Renaissance". *Journal of the History of Ideas* 52: 195-219.

Griswold, Jr., Charles. 1999. *Adam Smith and the Virtues of Enlightenment*. Cambridge: Cambridge University Press.

Grusec, Joan E. 2006. "The Development of Moral Behavior and Conscience from Socialization Perspectives" in *Handbook of Moral Development*, eds. Melanie Killen and Judith Smetana. Mahwah, NJ: Lawrence Erlbaum Associates.

Haakonssen, Knud. 1790/2002. "Introduction" in *The Theory of Moral Sentiments* by Adam Smith. Cambridge: Cambridge University Press.

—— 1981. *The Science of a Legislator*. Cambridge: Cambridge University Press.

Hanley, Ryan Patrick. 2009. *Adam Smith and the Character of Virtue*. Cambridge: Cambridge University Press.

Heilbroner, Robert L. 1982. "The Socialization of the Individual in Adam Smith". *History of Political Economy* 14: 427-439.

Hill, Lisa. 2001. "The Hidden Theology of Adam Smith". *European Journal of the History of Economic Thought* 8: 1-29.

Hill, Jr., Thomas E. 1998. "Four Conceptions of Conscience" in *Integrity and Conscience*,

eds. Ian Shapiro and Robert Adams. New York: New York University Press.

—— 2002. *Human Welfare and Moral Worth: Kantian Perspectives.* Oxford: Oxford University Press.

Hobbes, Thomas. 1640/2008. *The Elements of Law*, ed. J. C. A. Gaskin. Oxford: Oxford University Press.

—— 1651/1994. *Leviathan*, ed. Edwin Curley. Indianapolis: Hackett.

Hope, Vincent M. 1989. *Virtue by Consensus: The Moral Philosophy of Hutcheson, Hume and Adam Smith*. Oxford: Oxford University Press.

Hume, David. 1751/1998. *An Enquiry Concerning the Principles of Morals*, ed. Tom L. Beauchamp. Oxford: Oxford University Press.

—— 1739/2000. *A Treatise of Human Nature*, ed. D. F. Norton & M. J. Norton. Oxford: Oxford University Press.

Kant, Immanuel. 1788/1997. Critique of Practical Reason, trans. and ed. Mary Gregor. Cambridge: Cambridge University Press.

—— 1785/1998. *Groundwork of the Metaphysics of Morals*, ed. Mary Gregor. Cambridge: Cambridge University Press.

—— 1797/1996. *Metaphysics of Morals*, ed. Mary Gregor. Cambridge: Cambridge University Press.

—— 1793/1998. *Religion within the Boundaries of Mere Reason*, ed. A. Wood and G. D. Giovanni. Cambridge: Cambridge University Press.

Kelly, Kevin T. 1967. *Conscience: Dictator or Guide?* Wimbledon, London: Geoffrey Chapman Ltd.

Kennedy, Gavin. 2008. *Adam Smith: A Moral Philosopher and His Political Economy*. New York, NY: Palgrave Macmillan.

Keppler, Jan Horst. 2010. *Adam Smith and the Economy of the Passions*. New York: Routledge Press.

Khalil, Elias L. 2005. "An Anatomy of Authority: Adam Smith as Political Theorist". *Cambridge Journal of Economics* 29: 57-71.

—— 2009. "Self-deceit and self-serving bias: Adam Smith on 'General Rules'". *Journal of Institutional Economics* 5: 251-258.

Kleer, Richard A. 1995. "Final Causes in Adam Smith's *Theory of Moral Sentiments*". *Journal of the History of Philosophy* 33: 275-300.

Kohlberg, Lawrence. 1973. "The Claim to Moral Adequacy of a Highest Stage of Moral Judgment". *The Journal of Philosophy* 70: 630-646.

—— 1981. *The Philosophy of Moral Development: Moral Stages and the Idea of Justice*. San Francisco: Harper & Row.

Koops, W., Brugman, D., & Ferguson, T. J. 2010. "The Development of Conscience: Concepts and Theoretical and Empirical Approaches" in *The Development and Structure of Conscience*, eds. Koops, Brugman, Ferguson, and Sander. New York: Psychology Press.

Korsgaard, Christine. 1999. "The General Point of View: Love and Moral Approval in Hume's Ethics". *Hume Studies* 25: 3-41.

—— 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.

Langston, Douglas. 2001. *Conscience and Other Virtues*. University Park, PA: The Pennsylvania State University Press.

Lewis, C. S. 1960. *Studies in Words*. Cambridge: Cambridge University Press.

Lyons, William. 2009. "Conscience – An Essay in Moral Psychology". *Philosophy* 84: 477-494.

Macfie, A. L. 1967. *The Individual in Society: Papers on Adam Smith*. London: George Allen & Unwin Ltd.

Martin, Marie, A. 1990. "Utility and Morality: Adam Smith's Critique of Hume". *Hume Studies* 16: 107-120.

Mill, John Stuart. 1863/1998. *Utilitarianism*, ed. Roger Crisp. Oxford: Oxford University

Press.

Montes, Leonidas. 2004. *Adam Smith in Context: A Critical Reassessment of Some Central Components of His Thought*. New York: Palgrave MacMillan.

Moran, Richard. 2001. *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton: Princeton University Press.

Nanay, Bence. 2010. "Adam Smith's Concept of Sympathy and its Contemporary Interpretations" in *The Philosophy of Adam Smith: The Adam Smith Review*, vol. 5, eds. Vivienne Brown and Samuel Fleischacker. New York: Routledge.

O'Brien, Wendell. 1991. "Butler and the Authority of Conscience". *History of Philosophy Quarterly*, 8: 43-57.

Otteson, James. 2002. "Adam Smith's Marketplace of Morals". *Archiv für Geschichte der Philosophie* 84: 190-211.

Oxley, Julinna C. 2011. *The Moral Dimensions of Empathy: Limits and Applications in Ethical Theory and Practice*. New York: Palgrave Macmillan.

Özler, Sule and Gabrinetti, Paul A. 2013. "A Known World: An Analysis of Defenses in Adam Smith's *The Theory of Moral Sentiments*" in *The Adam Smith Review*, vol. 8, ed. Fonna Forman-Barzilai. New York: Routledge.

Paton, H. J. 1979. "Conscience and Kant". *Kant-Studien,* 70: 239-251.

Penelhum, Terence. 1985. *Butler*. London: Routledge & Kegan Paul.

Potts, Timothy. 1980. *Conscience in Medieval Philosophy*. Cambridge: Cambridge University Press.

Prichard, H. A. 1912. "Does Moral Philosophy Rest on a Mistake?" *Mind* 21: 21-37.

Raphael, D. D. 1949. "Bishop Butler's View of Conscience". *Philosophy* 24: 219-238.

—— 1972/3. "Hume and Adam Smith on Justice and Utility". *Proceedings of the Aristotelian Society* New Series 73: 87-103.

—— 1975. "The Impartial Spectator" in *Essay on Adam Smith*, eds. Andrew S. Skinner and

Thomas Wilson. Oxford: Clarendon Press.

—— 2007. *The Impartial Spectator*. Oxford: Oxford University Press.

Rick, Jon. 2007. "Hume's and Smith's Partial Sympathies and Impartial Stances". *The Journal of Scottish Philosophy*, 5: 135-158.

Ross, Ian Simpson. 2010. *The Life of Adam Smith*, 2nd ed. Oxford: Oxford University Press.

Ryle, Gilbert. 1940. "Conscience and Moral Convictions". *Analysis* 7: 31-39.

Sayre-McCord, Geoffrey. 2010. "Sentiments and Spectators: Adam Smith's Theory of Moral Judgment" in *The Philosophy of Adam Smith: The Adam Smith Review*, vol. 5, eds. Vivienne Brown and Samuel Fleischacker. New York: Routledge.

Schinkel, Anders. 2011. "Huck Finn, Moral Language and Moral Education". *Journal of Philosophy of Education* 45: 511-525.

Schliesser, Eric. 2011. "Reading Adam Smith after Darwin: On the Evolution of Propensities, Institutions, and Sentiments". *Journal of Economic Behavior & Organization* 77: 14-22.

—— 2008. "Review of D. D. Raphael (2007) *The Impartial spectator: Adam Smith's Moral Philosophy* and Leonidas Montes (2004) *Adam Smith in Context*." *Ethics* 118: 569-575.

Shaver, Robert. 2006. "Virtues, Utility, and Rules" in *The Cambridge Companion to Adam Smith*, ed. Knud Haakonssen. Cambridge: Cambridge University Press.

Smith, Adam. 1795/1980. *Essays on Philosophical Subjects*. Oxford: Oxford University Press.

—— 1776/2007. *An Inquiry into the Nature and Causes of the Wealth of Nations*. Petersfield, Hampshire: Harriman House Ltd.

—— 1976/1982. *Lectures on Jurisprudence*, eds. R. L. Meek, D. D. Raphael, and P. G. Stein. Indianapolis, IN: Liberty Fund, Inc.

—— 1790/2002. *The Theory of Moral Sentiments*, ed. Knud Haakonssen. Cambridge: Cambridge University Press.

Stout, Lynn. 2011. *Cultivating Conscience: How Good Laws Make Good People*. Princeton: Princeton University Press.

Stueber, Karsten. 2009. “The Ethical Dimension of Folk Psychology?” *Inquiry* 52: 532-547.

—— 2006. *Rediscovering Empathy: Agency, Folk Psychology, and the Human Sciences.* Cambridge, MA: MIT Press.

Sturgeon, Nicholas L. 1976. “Nature and Conscience in Butler’s Ethics”. *The Philosophical Review* 85: 316-356.

Vivenza, Gloria. 2001. *Adam Smith and the Classics*. Oxford: Oxford University Press.

Weber, Dominique. 2010. “Thomas Hobbes’s Doctrine of Conscience and Theories of Synderesis in Renaissance England”. *Hobbes Studies* 23: 54-71.