Perspective Taking and Moral Evaluation: Themes from Adam Smith

by

Warren Alexander Herold

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Philosophy)
in the University of Michigan
2014

Doctoral Committee:

Professor Elizabeth S. Anderson, Chair
Emeritus Professor Stephen Leicester Darwall
Professor Phoebe C. Ellsworth
Professor Allan F. Gibbard
Professor Daniel Jacobson

ACKNOWLEDGMENTS

I am indebted to many people for their help on this project. I have received invaluable guidance from my committee members: Elizabeth Anderson, Stephen Darwall, Phoebe Ellsworth, Allan Gibbard, and Daniel Jacobson. I have also received enormously helpful feedback from Gordon Belot, Mara Bollard, Paul Boswell, Emily Brady, Sarah Buss, Steve Campbell, Nathaniel Coleman, Daniel Drucker, Samuel Fleischacker, Michael Frazer, Dmitri Gallow, Gordon Graham, Jason Konek, Ethan Kross, Louis Loeb, Ishani Maitra, Sven Nyholm, Dan Peterson, Peter Railton, Alex Sarch, Eric Schliesser, Alex Silk, Dan Singer, Walter Sowden, Chandra Sripada, Nils-Hennes Stear, Karsten Stueber, Kendall Walton, Brian Weatherson, David Wiens, and Robin Zheng.

Finally, I would like to thank my wife, Laura Herold. She supported me in countless ways as I worked on this project, and I am extremely grateful to her.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

TMS	Adam Smith, *The Theory of Moral Sentiments*. References to the 6th edition are by Part, Section, Chapter, and Paragraph (e.g., "*TMS* I.ii.3.4"), except for Part III, which has no Sections. References to earlier editions are identified as such and given by page number (e.g., "*TMS*, Ed. 2, 207").
Treatise	David Hume, *A Treatise of Human Nature*. References are by Book, Part, Section, and Paragraph (e.g., "*Treatise* 1.2.3.4").

CHAPTER 1. PERSPECTIVE TAKING AND MORAL THEORY

1. From Imaginative Perspective Taking to Utilitarianism

Moral deliberation is, in part, an act of imagination. When we evaluate the moral permissibility of an action, we take it upon ourselves to consider it not only from our own perspective, but from the perspectives of others; and this requires the imagination. Of course, moral deliberation involves other things as well. Reason and empirical investigation play a role – in determining an action's likely consequences, for instance, and identifying the most efficient means to a predetermined end. But there are some facts that neither reason nor empirical investigation can uncover. Though they can inform us of the likely effects of our actions – *that* such-and-such would likely happen to so-and-so if we were to act in a certain way – they cannot convey a sense of *what it is like* to be so affected.[1] The latter information is essential to moral thought, and it can only be acquired through the imagination of life in another person's shoes. Moral deliberation, we might therefore say, involves not just an act of imagination, but a particular type of imaginative act: that of imagining what it would be like to be someone else.

I take this claim to be obvious and uncontroversial. The idea (or something like it) is common to both everyday thought and abstract ethical theory. It finds expression in such things as the *Golden Rule* ("Do unto others as you would have them do unto you"), the directive to "Imagine how you would feel if someone did that to you", and, in one form or another, the work

[1] For a discussion of this distinction and its importance to moral thought, see Hare (1981, pp. 91-92).

of countless moral philosophers.[2] The idea appears in the work of social and developmental psychologists as well. According to Martin Hoffman, for instance, the experience of being told to "Imagine how you would feel if someone did that to you" – or, more generally, any experience that directs an individual's attention to the perspective of another, highlights the latter's feelings of pleasure or distress, and encourages an appropriate response in the former (e.g., pride, guilt, a desire or feeling of obligation to help) – plays a central role in the moral development of children (M. Hoffman, 2000). It is from such experiences, Hoffman argues, that children learn to consider and incorporate the feelings of others into their own practical deliberations. It is from such experiences that they learn to deliberate *morally*.

The idea seems simple enough: imagine that you were in another person's shoes, and do as you would be done by. Planning to steal someone's money, I stop to consider how I would feel if someone were to steal mine. Finding that I wouldn't like it, and assuming (plausibly) that my prospective victim is like me in this respect, I decide not to proceed. Other cases are slightly more complex, but still not very difficult. About to play my trumpet, but aware of my neighbor's aversion to the sound, I pause to consider the situation from his point of view.[3] Unlike the previous case, though, I do not imagine *myself* in my neighbor's shoes; for, unlike my neighbor, I love the sound of the trumpet and would not mind if my neighbor were to play. Rather than imagining how I would feel if my neighbor were to play, I imagine instead how my neighbor would feel if I were to play. Paraphrasing R.M. Hare, I imagine not myself in my neighbor's situation with *my own* likes and dislikes, but myself in my neighbor's situation with *my*

[2] I am thinking, for instance, of Adam Smith's "impartial spectator" (1976 [1790]), Immanuel Kant's "categorical imperative" (1997 [1788], 2002 [1785]), the utilitarian theories of John Harsanyi (1953, 1977, 1982) and, especially, R.M. Hare (1963, 1981), and the contractualist accounts of people like John Rawls (1971, 2001) and T.M. Scanlon (1982, 1998).

[3] I borrow this example from Hare (1963), who borrowed (and adapted) it from Braithwaite (1955).

neighbor's likes and dislikes (1963, p. 113). I imagine myself not just in my neighbor's shoes, but as if I were just like him.[4]

In his book, *Moral Thinking*, Hare argues that *all* moral deliberation ought to proceed in this way: that whenever we engage in critical moral thought, we take it upon ourselves to imagine being in other people's shoes, with *their* characteristics, and then incorporate their *actual* preferences into our own deliberations.[5] We must then decide what to do, he claims, by weighing our preferences against one another – our own, as well as those we acquire through the process just described – and, in cases of conflict, choosing the course of action that maximizes their strength-weighted satisfaction. About to play my trumpet, I weigh my preference that I play against my preference that, were I in my neighbor's situation with his aversion to the sound of the trumpet, I not play – where the strength of the latter is equal to the strength of my neighbor's preference that I not play. According to Hare, the requirement of imaginative perspective taking leads directly to utilitarianism.

Why does Hare think that this is how moral deliberation proceeds? He argues that his conclusions follows from two features of moral language, combined with what he claims to be a conceptual truth regarding the relationship between imagining what it is like to be someone else, on the one hand, and forming a preference regarding the hypothetical situation in which one is that other person, on the other. Hare's first claim is that moral language is *prescriptive*: moral statements, he says, express the speaker's preferences (1981, pp. 21-22).[6] More specifically, he

[4] George Bernard Shaw makes the same point in his *Maxims for Revolutionists*: "Do not do unto others as you would that they should do unto you. Their tastes may not be the same." Similarly, Karl Popper writes that the "golden rule is a good standard which is further improved by doing unto others, wherever possible, as *they* want to be done by" (1962, p. 438, emphasis in original).

[5] Hare does allow corrections for misinformed or irrational preferences (1981, pp. 101-106). I shall return to this point briefly below.

[6] Hare later writes that "moral judgments are prescriptive, and to have a preference is to accept a prescription" (1981, p. 91).

claims that moral statements express the speaker's *all-things-considered* preferences: the preferences one comes to after weighing all of one's preferences against one another and resolving any conflicts in accordance with the conflicting preferences' relative strengths. According to Hare, the statement "I ought to do X" expresses my all-things-considered preference that I do X.

Hare's second claim is that moral judgments are *universalizable*: that our moral judgments "entail identical judgments about all cases identical in their universal properties" (1981, p. 108). In particular, Hare argues that the universalizability of moral judgments entails that

> if I now say that I ought to do a certain thing to a certain person, I am committed to the view that the very same thing ought to be done to me, were I in exactly his situation, including having the same personal characteristics and in particular the same motivational states.

According to Hare, combining prescriptivity and universalizability yields the following conclusion: that to say now that I ought to do a certain thing to a certain person is to express my all-things-considered preference that the thing be done, where this all-things-considered preference is produced by weighing and aggregating not only my own preferences regarding the situation, but the preferences that I now have for the hypothetical scenario in which I am in the position of, and have the same personal characteristics and motivational states as, the person to whom I claim I ought to do the certain thing.[7] Considering whether or not to play my trumpet, I decide what to do by weighing my preference that I play against my preference that, were I in my neighbor's situation, with his aversion to the sound of the trumpet, I not play. I decide to play if and only if the former is stronger than the latter.

[7] I am assuming, for simplicity, that no other people have any preferences regarding whether or not the "certain thing" is done to the "certain person." If there were other people with relevant preferences, then Hare argues that their preferences would need to be accounted for in the same manner as those of the "certain person" to whom the "certain thing" is done (1981, pp. 110-111). But I will ignore this here.

Many philosophers would, I think, agree that moral judgments are universalizable in something like the way Hare suggests. T.M. Scanlon's "universality of reason judgments" requirement, for instance, is nearly identical to Hare's universalizability requirement, though Scanlon presents his claim as one about practical reasons in general, rather than moral statements in particular (1998, pp. 73-74). Similarly, John Rawls insists that any acceptable ethical principle must satisfy a number of formal constraints, the first being that it must be expressible without the use of "proper names" or "rigged definite descriptions" (1971, p. 113). Rawls's constraint is not unlike Hare's universalizability requirement.[8] Hume's "general point of view" expresses something similar.[9] When evaluating a person's character, Hume claims that we necessarily "over-look our own interest" and "consider not whether the persons affected by the qualities be our acquaintance or strangers, countrymen or foreigners" (*Treatise* 2.3.1.17). Though expressed quite differently, Hume's point is again similar to Hare's: that a person's moral judgments are necessarily invariant with respect to her position relative to others. On its own, universalizability requirement is relatively uncontroversial.

But Hare's utilitarian conclusion does not follow from the universalizability of moral judgments alone, nor even from the combination of universalizability and prescriptivity. Taken together, these two features of moral language can imply, at most, that to say now that I ought to do a certain thing to a certain person is to express my all-thing-considered preference that the thing be done, where this all-things-considered preference is produced by aggregating my own preferences regarding the situation and the preferences that I now have for the hypothetical

[8] Rawls's terminology is different – and a potential source of confusion. He refers to the constraint just mentioned as the "generality condition," and he uses the term "universal" to refer to something different: a principle of justice which holds "for everyone in virtue of their being moral persons" (1971, p. 114). For a discussion of these and other "formal constraints of the concept of right," see Rawls (1971, pp. 112-118). For a critique of Rawls's terminology, see Hare (1981, pp. 63-64).

[9] For a helpful discussion of this aspect of Hume's account, see Bricke (1996, pp. 110-115).

scenario in which I am in the position of, and have the same personal characteristics and motivational states as, the person to whom I claim I ought to do the certain thing. In order to establish his utilitarian conclusion, however, Hare would need to show not only this, but something more: namely, that, when I deliberate correctly, the preferences I now have for the hypothetical scenario in which I am in the position of, and have the same personal characteristics and motivational states as, the person to whom I claim I ought to do the certain thing *must correspond to that person's preferences about whether or not I do it*. In order to establish his conclusion, in other words, Hare would need to show that when I imagine myself in other people's shoes, I *must* (if I am doing it correctly) acquire preferences regarding their circumstances that perfectly mirror their own.

This brings us to Hare's aforementioned claim regarding the relationship between imagining what it is like to be someone else, on the one hand, and forming a preference regarding the hypothetical situation in which one is that other person, on the other. Suppose that I were to imagine what it is like to be someone who is suffering – e.g., to be someone who has just broken his neck. Hare asks:

> Can I properly be said to know what it is like for him (not just to know that his neck is being broken), unless I myself have an equal aversion to having that done to me, were I in his position with his preference? […] Suppose that I said "Yes, I know just how you feel, but I don't mind in the least if somebody now does it to me": should I not show that I did not really know, or even believe, that it was like *that*? Would not my lack of knowledge, or else my insincerity, be exposed if somebody said "All right, if you don't mind, let's try"? (1981, p. 94)

In order to truly and successfully imagine what it is like to be someone else, Hare claims that one must *fully represent* to oneself that person's situation, including the fact that they have whatever personal characteristics and motivational states that they in fact have. And, if done correctly, Hare insists that this process will necessarily lead one to acquire a set of *conditional preferences*

(i.e., preferences for the hypothetical scenario in which one is the other person) that perfectly mirror the other person's actual preferences (1981, p. 99; see also Hare, 1988, p. 216). Returning to the above example, Hare writes:

> If, by some quirk of nature, I were a person who knew that he did not feel pain in that situation, or if I knew that I was going to become such a person by being anaesthetized, then I might indeed sincerely say that I did not mind being subjected to the experience (ignoring for the sake of arguments its consequences). But this would be irrelevant; and so would it be if I knew that I would feel pain, but for some reason would not mind it. For I am to imagine myself in his situation with *his* preferences. Unless I have an equal aversion to myself suffering, forthwith, what he is suffering or going to suffer, I cannot really be knowing, or even believing, that being in his situation will be like *that*. (1981, pp. 94-95)

Though I may imagine what it is like to be another person without thereby acquiring a set of conditional preferences that mirror their actual preferences, I cannot, according to Hare, be said to have fully represented to myself the other person's experiences unless I do acquire such a set of conditional preferences. Unless I acquire such a set of conditional preferences, I am, Hare insists, doing something wrong. I will refer to this as the *conditional reflection principle* (Gibbard, 1988). If Hare is right, then critical moral thought must always be conducted in accordance with this principle. If he is right, then whenever we imagine what it is like to be someone else as part of our moral deliberation, we must acquire a set of conditional preferences regarding that person's situation that perfectly mirrors his or her own.

Adding conditional reflection to prescriptivity and universalizability yields Hare's utilitarian conclusion: namely, that to say now that I ought to do a certain thing to a certain person is to express my all-things-considered preference that the thing be done, where (i) my all-things-considered preference is produced by aggregating my own preferences regarding my situation and the preferences I now have for the hypothetical scenario in which I am in the position of, and have the same personal characteristics and motivational states as, the person to whom I claim I

ought to do the certain thing, and (ii) the latter preferences perfectly mirror those of that person. Given prescriptivity and universalizability, in other words, the addition of condition reflection implies that I must decide what to do by weighing my own preferences against the preferences of the other person. The combination of prescriptivity, universalizability, and conditional reflection leads directly to a form of preference utilitarianism.

2. An Objection

As an empirical claim about what actually happens when we imagine ourselves in other people's shoes, Hare's conditional reflection principle is certainly false. When we imagine ourselves in the position of, and with the same personal characteristics and motivational states as, someone who prefers with strength *s* that *p* should occur, we *may* now acquire a preference of strength *s* that, were we in that person's circumstances and with that person's characteristics and motivational states, *p* should occur; but we may not. We may acquire a preference that is stronger or weaker than *s*. As Bernard Williams points out, this can even happen when we imagine being in a situation in which *we ourselves* would in fact prefer with strength *s* that *p* should occur.

> I indeed know, for instance, that if my house caught fire, I would prefer, with the greatest possible intensity, that my family and I should get out of it. Since I am a moderately rational agent, I take some action now to make sure that we could do that if the situation arose, and that action comes of course from a preference I have now. But there is no sense at all in which the present prudential preference is of the same strength as the preference I would have if the house were actually on fire (driving almost every other consideration from my mind), and it is not rational that it should be. (Williams, 1985, p. 90)

Richard Brandt makes a similar point:

> Suppose, after four hours of no liquid intake on a hot day, I am thirsty and want a drink. We know something about how this works: the dehydration of the cells

> signals to the brain through the blood, with the result that the idea of a drink becomes attractive and I exert myself to get a drink. But suppose I merely consider now, with vivid imagination, that I am forthwith to be deprived of liquid, be dehydrated, and feel as I do when I am. Shall I now want a drink, for this hypothesized near-future situation? Perhaps it is true that I shall want not to be unpleasantly thirsty owing to no liquid intake. But the causal process which brings about the thirst, or wanting a drink, in the real case, is totally different from what it is in the case of the mere belief about being dehydrated. (Brandt, 1988, p. 34)

Even if it is the case that, after four hours without any liquid on a hot day, I would in fact prefer with strength *s* to drink a glass of water, I may not now prefer with strength *s* that, were I in such a situation, I drink a glass of water. This may be so even if I were to imagine as vividly as possible that I have been deprived of liquid for four hours on a hot day. My conditional preferences now may fail to reflect what my actual preferences would be in such a situation.

Indeed, we can push this point even further. When we imagine ourselves in the position of, and with the same personal characteristics and motivational states as, someone who prefers with strength *s* that *p* should occur, we may not acquire *any* preference that, were we such a person, *p* should occur. We may acquire an altogether different preference, perhaps even an opposing one: we may prefer that, were we such a person, *p* should *not* occur. Imagining myself as a person with a strong desire to rape another individual, for example, I do not merely fail to acquire a sufficiently strong preference that, were I such a person, I should rape another individual; I fail to acquire *any* preference of the kind. Imagining myself as such a person, I do not come to prefer that, were I such a person, I should rape another individual. I prefer instead that I should *not* rape anyone – and, for that matter, that I should not have such a desire.

So far, all of this is consistent with Hare's account. For Hare agrees that, as a matter of fact, people often fail to acquire conditional preferences that mirror those of the people whose circumstances, personal characteristics, and motivational states they imagine. But he denies that

this is a problem for his theory. As he sees it, the conditional reflection principle is normative and conceptual, not empirical. Whatever *actually* happens when we imagine ourselves as other people, Hare's claim is that *good* critical moral thought requires us to "fully represent to ourselves" other people's circumstances, personal characteristics, and motivational states; and he insists that, when done correctly, such full representation necessarily leads us to acquire a set of conditional preferences that perfectly mirror those of the people whose circumstances, personal characteristics, and motivational states we represent (1981, p. 99; see also Hare, 1988, p. 216). Brandt's example of the thirsty individual, Hare argues, serves only to demonstrate the "difficulty of fully representing to ourselves absent states of experience" (1988, p. 217).

> No doubt causal (e.g., physiological) factors favour or impede full representation; but the connection between full representation and having the desire is conceptual. (1988, p. 216)

The fact that the conditional reflection principle fails as an empirical hypothesis does not, on its own, show that critical moral thought does not require us to acquire conditional preferences that correspond to the preferences of those whose circumstances, personal characteristics, and motivational states we imagine. If Hare is right, then all it shows is that most people are not good critical moral thinkers: that they often fail to fully represent to themselves other people's circumstances, characteristics, and motivational states.

Let's set empirical objections aside, then, and consider Hare's claim as he intends it to be understood: that *ideal* critical moral thought requires full representation of other people's circumstances, personal characteristics, and motivational states; and such full representation necessarily leads to the acquisition of conditional preferences that mirror other people's actual preferences. Is this plausible? Let's take Hare's definition of "full representation" as given. Let's assume, in other words, if only for the sake of argument, that fully representing to oneself

another person's circumstances, characteristics, and motivational states necessarily leads one to acquire a corresponding set of conditional preferences. Given this understanding of full representation, is it plausible to claim that good critical moral thought requires it? Are we required to add other people's preferences to our own, and then deliberate on the basis of their relative strengths alone?

Hare's claim seems reasonable in some situations. When deciding whether or not to play my trumpet, it seems reasonable to proceed by weighing my own preference that I play against my neighbor's preference that I do not play. As a general principle of moral deliberation, though, Hare's claim is deeply counterintuitive. Return, for instance, to the rape example. When evaluating the moral permissibility of rape, we do not pause to consider how we would feel if we had a desire to rape but were prevented from doing so.[10] We do not, in other words, weigh the preferences of rapists against those of their potential victims when deciding whether they should be allowed to rape, as if the former could ever outweigh the latter and thereby justify an act of rape. Unlike my neighbor's preference for silence, we do not give a rapist's preference to rape any weight at all.

On its own, of course, the fact that we do not give any weight to rapists' preferences in our own deliberations is not an objection to Hare's account. For, as we have already seen, Hare's claim is normative, not empirical; it is about how critical moral thought *ought ideally* to work, not about our everyday deliberations. The suggestion I wish to make at present is not merely that we *do* not give any weight to rapists' preferences, but that our actual practice is appropriate – that we *ought* not to give weight to rapists' preferences. The problem with Hare's account is not that it implies that we do in fact give weight to rapists' preferences in our moral deliberations,

[10] I take this example from Barry (2001, p. 34), though Barry uses it for a different purpose.

when, in reality, we do not. Rather, the problem is that Hare's account implies that we ought to give weight to rapists' preferences, whereas it seems intuitively obvious that we ought not to. Our actual deliberative practices seem, in this respect, to be as they ought to be. In contrast, Hare's proposal seems misguided.

It is not only the existence of horrific preferences, like those of the rapist, moreover, that present a problem for Hare's account. Even the *absence* of a preference can cause trouble. Consider, for instance, the case of a person living in conditions of severe and deeply entrenched inequality. As Amartya Sen notes, people in such circumstances typically "do not go on grieving and grumbling all the time, and may even lack the motivation to desire radical change of circumstances" (1992, p. 6). They respond to their situation not by lamenting or attempting to change it, but by changing themselves: by learning to "take pleasures in small mercies" and "cut down personal desires" to more "realistic" levels (1992, p. 55). As a result, such people can often be reasonably content with their lives; they may even lack the desire for change.

From the fact that such people lack the desire for change, however, it does not follow that no reason for change exists – that no harm has been done to them, that they deserve no better. Imagining ourselves in their shoes, we feel resentment on their behalf, coupled with a strong desire for change. We do so even if – perhaps, as Hume would say,[11] even *more so* if – they exhibit no such feelings. We imagine ourselves not as if we were just like them – with preferences adapted to unjust circumstances – but as if we were slightly different. And, again, our practice seems to be as it ought to be: it is appropriate that we should feel resentment on behalf of those who have been treated unjustly, gratitude on behalf of those who have benefited

[11] For Hume's view, see *Treatise* (2.2.7.5-6).

from uncommon kindness, and so forth – whether or not those for whom we feel such things exhibit any such feelings.

Hare is aware of such objections, of course, and he has a response. He introduces information and rationality requirements that he thinks capable of protecting his account from potential problems due to misinformed or irrational preferences, and he develops a sophisticated two-level account of moral deliberation that he thinks capable of reconciling his theory with precisely the sorts of everyday moral intuitions on which I have relied.[12] *Critical moral thinking*, as he calls it, does indeed require us to give equal weight to the fully-informed and rational preferences of everyone – to the sadistic pleasures of the Marquis de Sade as well as the benevolent pleasures of Mother Teresa (1981, pp. 140-141). But the apparent implausibility of this claim does not, Hare argues, constitute a genuine objection to his account. For he claims that at the *intuitive level* we are right to regard certain pleasures (e.g., those of the rapist) as horrific and unworthy of moral consideration. We are right to do so, he says, not because such pleasures really are unworthy of moral consideration, but because viewing them as such helps to guide our everyday deliberations. The world is a better place because we have and regulate our actions in accordance with such intuitive moral feelings; they are the feelings that we would have if we had been "brought up successfully by thoughtful utilitarians" (1981, p. 141). In the end, though, Hare insists that these intuitive feelings are, strictly speaking, wrong. At the critical level, all fully-informed and rational preferences must be treated equally.

I do not think that Hare's two-level proposal can rescue his account.[13] My reason is simple. What is so counterintuitive about Hare's theory is precisely the fact that it insists that critical

[12] For a detailed description of this two-level account, see Hare (1981, pp. 25-64).
[13] For criticisms, see Mackie (1977, pp. 83-102), Williams (1985, pp. 82-92), and the contributions to Seanor and Fotion (1988).

moral thought requires equal weight to be given to everyone's fully-informed and rational preferences – sadistic and benevolent alike. *That* is the claim that would need to be either modified or supported with additional argument in order to adequately respond to the objection outlined above. But Hare's two-level theory neither modifies this claim nor gives his critics any additional reason to think it correct. That Hare thinks the world a better place because most people have moral intuitions that conflict with his theory of critical moral thought does not change the fact that his theory is deeply counterintuitive. Consequently, people (like myself) who have moral intuitions that conflict with Hare's theory – and who view these intuitions as intuitions *about the right way to engage in critical moral thought* – will have just as much reason to object to his theory as before. Even with the distinction between *critical* and *intuitive* thought in place, Hare's remains as counterintuitive as ever.

3. An Alternative Approach

When we imagine ourselves in other people's shoes for the purposes of moral deliberation, we do not take all of their preferences, feelings, attitudes, values, principles, and so forth, as given and deliberate on the basis of their strength alone; we do not deliberate in accordance with Hare's theory of critical moral thought. We discount some feelings, emphasize others, and introduce those we think lacking (e.g., resentment and a desire for change on behalf of those living in conditions of severe and deeply entrenched inequality). The question is: how do we do this? How do we decide which preferences, feelings, attitudes, values, principles, and so forth, to incorporate into our deliberations, and which to exclude? Put differently, how do we decide which of other people's characteristics to incorporate into our imaginative exercises and which to exclude? How *ought* we to do this?

Numerous answers could be given: we should assimilate other people's tastes into our imaginative simulations, but not their ideals;[14] or we should assimilate their "personal" or self-regarding preferences, but not their "external" or other-regarding preferences;[15] or, as Hare himself suggests, we should assimilate only their rational and well-informed preferences.[16] Unfortunately, though, none of these proposals can solve the problems mentioned earlier. The difficulty in the rape example derives neither from the rapist's endorsement of a deviant ideal nor from a questionable external preference, but rather from the rapist's horrific taste for rape.[17] Excluding ideals, external preferences, irrational preferences, or ill-informed preferences would therefore do nothing to eliminate the problem. Similarly, we feel resentment, rather than contentment, on behalf of people living in conditions of severe and deeply entrenched inequality not because we think their actual feelings distorted by objectionable ideals or external preferences, nor even because we think them uninformed or irrational,[18] but because that is how people in their circumstances *ought* (in at least some sense of the term) to feel. We feel resentment on behalf of such people because resentment is *called for* in their circumstances – because it is a *proper* or *fitting* response to the way they have been treated.

This, at any rate, is Adam Smith's view, as developed in *The Theory of Moral Sentiments*. Like Hare, Smith thinks that imagining oneself in another person's shoes is central to moral

[14] Williams suggests this as a "natural" way of understanding "everyday uses of role-reversal arguments" (1985, p. 86).
[15] I borrow this distinction from Ronald Dworkin (1977), who uses it for a different purpose. Hare indicates that he thinks "external" preferences ought to be taken into account like any other, but he excludes them from his official theory for the sake of simplicity, leaving the issue unresolved (1981, p. 104).
[16] For Hare's view on this, see Hare (1981, pp. 101-106).
[17] I do not mean to say that the rapist's ideals or moral beliefs are irrelevant. The rapist's horrific personal taste for rape may be made possible by his *lack* of a particular moral belief – namely, the belief that rape is morally wrong. If he had such a belief, then perhaps he would not retain his personal taste for rape.
[18] As Sen writes, "in situations of adversity which the victims cannot individually change, *prudential reasoning* would suggest that the victims should concentrate their desires on those limited things that they *can* possibly achieve, rather than fruitlessly pining for what is unattainable" (1992, p. 55, emphasis in original). Their preferences may be perfectly rational.

thought. Unlike Hare, though, Smith flatly denies that engaging in a process of imaginative perspective taking necessarily leads one – or ought to lead one – to feel as the person whose perspective one imagines feels. We may feel differently, even when we perform the task correctly. Correspondence of feeling is, for Smith, neither the objective of imaginative perspective taking nor a necessary condition of its success. According to Smith, when we imagine ourselves in other people's circumstances, we do so neither to experience their feelings nor to incorporate their feelings into our own deliberations, but rather to establish a standard of propriety against which to evaluate their conduct. We imagine ourselves in their shoes and simulate a response in order to identify the feelings that it would be *proper* for them to feel. We feel as they do – we *sympathize* with them, as Smith would say – when and only when we feel as they do; and we approve of their conduct as proper when and only when they feel as they ought to. We decide what to do not by aggregating our and other people's actual preferences, but (very roughly) by considering how people *ought* to react to alternative courses of action. We decide whether a particular action would be morally permissible by asking ourselves whether anyone could properly resent it.

The objective of the present work is to examine and develop Smith's alternative, non-utilitarian account of imaginative perspective taking and moral thought. This way of characterizing Smith's work – as an alternative to utilitarianism – may come as a surprise to some readers, as Smith is often portrayed as an early utilitarian, one whose ideas are not at all unlike those of Hare. In *A Theory of Justice*, for example, John Rawls includes Smith (along with Hume, Bentham, and Mill) in his list of "great utilitarians" and *The Theory of Moral Sentiments*

in his list of "major eighteenth century" utilitarian works (1971, pp. xvii, 20).[19] Similarly, John Harsanyi describes his own decision-theoretic argument for utilitarianism as nothing more than "a modern restatement of Adam Smith's theory of an impartially sympathetic observer" (1982, p. 46). Even some Smith scholars have endorsed a utilitarian reading of Smith's account.[20]

But I think that this view is mistaken. In what follows, I argue that Smith's account leads in a different direction.[21] I begin, in Chapters 2 and 3, by examining Smith's account of *sympathy* – the core of his moral theory. In Chapter 2, I identify and discuss two distinctive features of Smith's account of sympathy: (i) his claim that we attempt to sympathize with other people by *imagining* and *simulating* a response to their circumstances, and (ii) his portrayal of this mechanism as fundamentally *evaluative* – one aimed not at reproducing other people's feelings, but at determining the proper response to other people's circumstances. After describing these elements of Smith's account, I defend Smith's position against common objections and suggest a few ways in which his proposal could be clarified or improved. The chapter's objectives are modest: it aims only to outline the primary features of Smith's account of sympathy, thereby setting the stage for the subsequent discussion.

Chapter 3 continues the work of Chapter 2 by examining Smith's psychological proposal in more detail. I ask how the imaginative process at the heart of Smith's account is supposed to work. Does Smith think that we attempt to sympathize with other people by imagining *ourselves* in their shoes? Or, like Hare, does he think that we are required to assimilate their personal characteristics and simulate a response to their circumstances *as if we were just like them*? Or

[19] Interestingly, Rawls gets the title of Smith's book wrong, referring to it as "*A Theory of the Moral Sentiments*", rather than "*The Theory of Moral Sentiments*". This has led one Smith scholar, D.D. Raphael, to conclude "that Rawls did not play close attention to Smith's *Moral Sentiments*" (2007, p. 46).
[20] See, for example, Campbell (1971) and Campbell and Ross (1981).
[21] I am by no means alone in denying that Smith is a utilitarian. Others who reject a utilitarian reading of Smith's work include Darwall (1999, 2004, 2006), Fleischacker (1999, 2004a), Harman (2000), Raphael (2007), Rothschild (2001), and Sayre-McCord (2010).

does he have something else in mind? After discussing the strengths and weaknesses of several possible readings, I present my own interpretation of Smith's account. We attempt to sympathize with (and evaluate) others, I claim, by assimilating their characteristics *to the best of our ability*, and then *attempting to* simulate a response to their circumstances as if we were just like them. Our sympathetic responses are constrained, however, by several limitations on our capacity to actively imagine feelings that are not currently our own. And this, I claim, is a good thing. The limitations on our imaginative abilities provide us with the distance needed to effectively evaluate other people's conduct. More specifically, I argue that our imaginative simulations are regulated by three distinct and beneficial constraints: (i) our "divided" sympathy (which leads us to take into account not only the perspectives of the individuals under evaluation, but of all affected parties), (ii) the moderating effect of distance (which, by reducing our susceptibility to "hot" emotional outbursts, prompts us to evaluate other people more calmly), and (iii) our basic incapacity to imagine "seeing" certain things as instantiating (or not instantiating) certain evaluative properties.

In Chapters 4 through 6, I shift my attention from Smith's account of sympathy to his account of self-evaluation. Smith famously argues that we evaluate our own feelings and behavior by reference to the judgments of an *impartial spectator*. Some interpreters have argued that Smith's psychological proposal commits him to thinking that we self-evaluate by simply internalizing the judgments typically made of us by our informed and impartial peers. I consider this interpretation of Smith's proposal in Chapter 4 and identify four immediate problems: (i) the reading incorrectly implies that Smith thinks it impossible for people to criticize their own society's evaluative standards, (ii) it fails to capture an important egalitarian ideal built into Smith's moral psychology, (iii) it ignores Smith's claims regarding the psychological effects of

self-evaluating via the impartial spectator model, and (iv) it leaves no room for Smith's distinctive notion of sympathy to play a role in the self-evaluative process. I argue, moreover, that the passages typically relied upon to support the aforementioned interpretation actually point in a different direction: to the claim that we self-evaluate not by internalizing the judgments made of us by *others*, but by imagining *ourselves* the spectators of our own conduct.

In Chapter 5, I explore and develop this last claim in more detail. I present empirical evidence to support Smith's assertion that by adopting a "self-distanced" perspective (i.e., viewing ourselves from a spectator's point of view), we can moderate our emotional, behavioral, and physiological reactivity, thereby enabling us to detach from our emotions and engage in a more critical and thoughtful self-evaluative process. I defend Smith's claim that we are motivated to self-evaluate (via self-distancing) by feelings of accountability, triggered by the presence of those to whom we are accountable. And I argue that the presence of such people does more than just prompt us to self-distance. In addition to triggering us *to* self-distance, the presence of observers can, in the right circumstances, *positively affect the way* we self-distance. More specifically, I show that interacting with a diverse group of people can enhance our self-evaluative judgments by exposing us to new information and improving the way we process it, motivating us to consider a wider range of perspectives, reducing our susceptibility to cognitive bias, and enhancing the quality, creativity, and originality of our evaluative thought.

I conclude my discussion in Chapter 6. Building on the notion of accountability introduced in Chapter 5, I suggest that Smith's account of sympathy and moral evaluation has more in common with the decidedly non-utilitarian contractualist accounts of people like John Rawls than with Hare's utilitarian moral theory.

CHAPTER 2. ADAM SMITH'S ACCOUNT OF SYMPATHY

1. Hume and Smith on Sympathy

To sympathize with someone, according to Hume, is "to receive by communication their inclinations and sentiments, however different from, or even contrary to our own," and then experience those inclinations and sentiments ourselves (*Treatise* 2.1.11.2). Hume believes the tendency to sympathize with one's associates – to experience their passions, share their desires, assume their opinions, and so forth – to be a natural and universal human trait, found not only in children but in "men of the greatest judgment and understanding." Even the "proudest and most surly" of men, he writes, "take a tincture from their countrymen and acquaintance." Placed in the company of others, they cannot help but be affected by their companions' mental states. Pleasure and pain, happiness and sadness, love and hatred, gratitude and resentment, esteem and contempt; all such passions, according to Hume, may be transmitted from one person to another by the non-verbal mechanism of sympathetic communication.

Nowadays the word "sympathy" is often used to refer to something different: a feeling that "(i) responds to some apparent obstacle to an individual's welfare, (ii) has that individual himself as object, and (iii) involves concern for him" (Darwall, 2002, p. 51; see also Darwall, 2006, pp. 43-48). To sympathize in this sense is to experience a particular emotion for a person in a particular type of scenario – to feel concern for a friend whose parent has recently passed away, for instance. But this is not what Hume has in mind. To sympathize with someone in Hume's sense is not to feel concern for that person; nor, for that matter, is it even something that we do

because we are concerned for someone.[22] As he uses the term, sympathy is not itself an emotion (e.g., concern); it is the experience of another person's emotion, or perhaps an emotion suited to another person's circumstances, whatever it might be. Moreover, for Hume, our sympathetic responses are not limited to negative contexts, as they are in the modern sense of the term. It is possible, he thinks, to sympathize with the good as well as the bad – with someone's joy over the birth of their child as well as their grief over the death of their parent. And Hume's notion of sympathy does not have the individual as its object. To sympathize with someone, according to Hume, is to feel *with* that person, not to feel *for* that person.

Smith introduced his account of sympathy roughly twenty years after Hume published his, and Smith's theory is similar to Hume's in several respects. Like Hume, for instance, Smith uses the word "sympathy" not in its modern sense, but to mean something like empathy or fellow-feeling: roughly, feeling what another person feels. And, again like Hume, Smith describes our tendency to sympathize with others as a universal human trait. Even the "greatest ruffian, the most hardened violator of the laws of society," he writes, "is not altogether without it" (*TMS* I.i.1.1). Moreover, Smith follows Hume's lead in depicting sympathy as a natural, and sometimes unavoidable, psychological phenomenon. When we "see a stroke aimed and just ready to fall upon the leg or arm of another person," he writes, "we naturally shrink and draw back our own leg or our own arm"; and when the stroke falls, "we feel it in some measure, and are hurt by it as well as the sufferer" (*TMS* I.i.1.3). We do not always wish to experience the

[22] Hume's view is closer to the opposite: insofar as we are concerned for people other than our friends, family members, and immediate associates, he believes that we are concerned for them *because* we sympathize with them. He writes, for instance, that the "good of society, where our own interest is not concern'd, or that of our friends, pleases only by sympathy" (*Treatise* 3.3.1.9); we have no "concern for society but from sympathy; and consequently 'tis that principle, which takes us so far out of ourselves, as to give us the same pleasure or uneasiness in character which are useful or pernicious to society, as if they had a tendency to our own advantage or loss" (*Treatise* 3.3.1.11). Similarly, he notes that "the public good is indifferent to us, except so far as sympathy interests us in it" (*Treatise* 3.3.6.1), and that "the happiness of strangers affects us by sympathy alone" (*Treatise* 3.3.6.2). For a related discussion, see Vitz (2004).

feelings of others, but our sympathetic experiences are not always under our control. We sometimes experience the feelings of others not because we wish to, but because we must: because our psychological makeup renders the experience inescapable.

Smith also follows Hume's lead in placing sympathy at the center of his moral theory. Hume believes that all moral distinctions arise from the tendency of a person's character to affect either their own interests or the interests of others, and he claims that our sympathy with these effects is what is ultimately responsible for our sentiments of approbation or disapprobation. The virtues, Hume writes, "must derive all their merit from our sympathy with those, who reap any advantage from them" (*Treatise* 3.3.6.1). Smith's account of merit and demerit is similar. After identifying the "sentiment or affection of the heart from which any action proceeds" as that "upon which its whole virtue or vice must ultimately depend,"[23] Smith identifies "the beneficial or hurtful effects which the affection proposes or tends to produce" as the determinant of the action's merit or demerit (*TMS* I.i.3.5). We judge a person's action to have merit or demerit, he claims, whenever we find that we can sympathize with the gratitude or resentment that it provokes in people.

Hume's influence on Smith is clear and unmistakable. But there are several important differences between their accounts. To begin, Smith and Hume propose different psychological mechanisms to explain our sympathetic experience. Very roughly, Hume claims that we sympathize with others by inferring their mental state from their behavior and then experiencing the inferred mental state ourselves. Smith concedes the initial plausibility of such an account. He notes, for instance, that sympathy does often appear "to arise merely from the view of a certain emotion in another person" (*TMS* I.i.1.6). "A smiling face," he writes, "is, to every body that

[23] Hume makes a similar claim (*Treatise* 3.2.1.2), as does Francis Hutcheson (2004 [1725], p. 101), who taught both Hume and Smith and had a clear influence on their work.

sees it, a cheerful object," just as a "sorrowful countenance is "a melancholy one."[24] But Smith nevertheless rejects Hume's proposal for two reasons. First, although he concedes that strong expressions of grief and joy often produce concordant feelings in spectators, he denies that all passions follow this pattern. The "furious behavior of an angry man," he writes, "is more likely to exasperate us against himself than against his enemies" (*TMS* I.i.1.7). Unless we are acquainted with the circumstances that provoked the man's anger, we will never be able to "bring his case home to ourselves" or "conceive any thing like the passions which it excites." Second, even when we do experience other people's feelings directly, Smith insists that until we are informed of the circumstances that excited them, our sympathy will remain "extremely imperfect" (*TMS* I.i.1.9). "General lamentations," he writes,

> which express nothing but the anguish of the sufferer, 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. The first question which we ask is, What has befallen you? (*TMS* I.i.1.9)

Though we may feel something, our responses to others will remain weak and provisional until we understand the circumstances from which their feelings arose.

It is with these objections in mind that Smith proposes his alternative psychological mechanism. Rather than relying on an inferential move from the observable effects of a person's mental state to the mental state itself, Smith claims that we attempt to sympathize with others by *imagining* being in their circumstances and then *simulating* a response. Discussing an example of a person being tortured upon the rack:

[24] The language employed by Smith to discuss this issue is reminiscent of Hume's. Smith writes, for example, that 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" (*TMS* I.i.1.6). This echoes Hume's description of passions "so contagious, that they pass with the greatest facility from one person to another, and produce correspondent movements in all human breast" (*Treatise* 3.3.3.5; see also 2.1.11.2, 3.3.1.7, and 3.3.2.2).

> By the imagination we place ourselves in his 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* I.i.1.2)

Our sympathetic suffering is triggered, according to Smith, not by the sufferer's behavior, but by the consideration "of the situation which excites it" (*TMS* I.i.1.10). We experience his pain – or something like it – by imagining that we were in his shoes.

That is the first major difference between Smith's account and Hume's. The second is even more substantial. Hume portrays sympathy as a non-evaluative phenomenon. To sympathize, in Hume's sense, is simply to feel what another person feels, or perhaps what someone in their situation typically feels, unaffected by any consideration of what they *ought* to feel. In contrast, Smith's account is evaluative. When, by the use of the imagination, we simulate a response to another person's circumstances, we do so, Smith claims, neither to identify nor to experience their *actual* feelings, but to identify the feelings that would be *proper*. We engage in imaginative simulation in order to establish a standard of propriety against which to compare people's conduct. We *sympathize*, according to Smith, when and only when our simulated feelings match the other person's actual feelings. Though Smith and Hume would thus agree that to sympathize with another is to feel what they feel, their explanations of *why* we feel what others feel are entirely different. For Hume, when we sympathize with someone, we feel what they feel *because they feel it*, whereas, for Smith, we do so *because it is the proper thing to feel*. We sympathize with people, according to Smith, when we feel what they ought to feel *and they do too*.

One final difference (implicit in what has been said) between Hume's and Smith's accounts needs to be mentioned. It has to do with Smith's notion of *propriety* – and the form of approval, and sense of "ought", to which it is connected. As mentioned above, Hume claims that the merit

of virtuous motives derives from our sympathy with those who "reap any advantage from them" (*Treatis*e 3.3.6.1). More generally, he writes that all "moral distinctions" arise, at least in great measure, from "the tendency of qualities and characters to the interests of society," and that it is "our concern for that interest, which makes us approve or disapprove of them" (*Treatise* 3.3.1.11). And as he believes that "we have no such extensive concern for society but from sympathy," he concludes that it is our sympathy with a motive's beneficial or harmful effects that determines our approval or disapproval. In short, we approve of a motive when it tends to have beneficial consequences, and we disapprove of it when it typically proves harmful.

Smith's picture is more complex. When we evaluate a person's sentiment or affection (or the action to which it leads), Smith denies that we consider *only* "the end which it proposes" or the "effect which it tends to produce" (*TMS* I.i.3.5). In addition to its end or effect, we consider also its "relation to the cause which excites it." We consider not only whether a particular action is, or aims to be, beneficial or harmful, in other words, but also whether it is a *fitting response* to the circumstances in which it was performed. An action's "merit or demerit" (the "qualities by which it is entitled to reward, or is deserving of punishment") resides in the former, Smith concedes; but its "propriety or impropriety" (its "suitableness or unsuitableness" to the "cause or object which excites it") resides in the latter (*TMS* I.i.3.6-7). And it is the latter, not the former, that Smith thinks we evaluate through sympathy. When he writes that to approve of another person's feelings is "*the same thing* as to observe that we entirely sympathize with them," his claim is about propriety, not merit. To find that we sympathize with a person's feelings is, according to Smith, to judge them *proper*.

I examine the details of Smith's psychological proposal in the next chapter. Before doing so, though, I must address two possible objections to his account. In Section 2 of this chapter, I

consider and respond to the claim that Smith's sympathy-based account of evaluative judgment fails because it necessarily precludes disapproval. In Sections 3-5, I consider Smith's claim that to approve of something (as proper) is to find that we sympathize with it. I review objections to Smith's claim, attempt to clarify his position, and then explain his reason for insisting on such a tight connection between sympathy and approval.

2. Two Senses of Knowing

As already noted, the primary role of imaginative mental simulation in Smith's account is evaluative: we use it to establish a standard of emotional and behavioral propriety. At one point, however, Smith appears to give simulation a non-evaluative role as well. "As we have no immediate experience of what other men feel," he writes, "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* I.i.1.2). The passage suggests that Smith thinks we use simulation to determine *both* the proper *and* the actual feelings of others – indeed, that it is *only* by simulating a response to other people's circumstances that we can learn about their actual mental states. If true, though, this would present an obvious problem for Smith's account. If we were to use the same psychological mechanism to determine both the actual and the proper feelings of others, then we would be incapable of disapproving of anyone's feelings or behavior. If our determination of the actual and the proper were the same, we would approve of each and every person's conduct in each and every case.[25]

[25] Robert Gordon criticizes Smith for precisely this reason, arguing that his account runs into difficulty because it fails to distinguish between two forms of imaginative simulation: in Gordon's words, between "imagining being in X's situation and making the further adjustments required to imagine being X in X's situation" (1995b, p. 741). We engage in the latter, Gordon thinks, to learn about X's actual mental state. In order to determine which feelings are proper, however, he thinks that we must "hold back […] from identification with the other person" and simulate (something like) our own response.

This can't be what Smith intends to say. Unfortunately, though, he provides no clear indication as to how he thinks the problem ought to be resolved. Indeed, he doesn't even acknowledge that there *is* a problem. This is clearly a flaw in Smith's presentation, but I do not think that it presents a serious problem for his account. In this section, I attempt explain how the problem can be resolved.

The passage quoted above notwithstanding, there is evidence in Smith's writing that he thinks that there are other methods available by which we can come to know how other people actually feel. Consider, for instance, the case of an individual who has benefited from the generosity of another. Smith believes that we determine the proper response to such a situation by imagining ourselves as the beneficiary of some great act of generosity. But he hints at a different mechanism for determining the beneficiary's actual feelings. We "are shocked beyond all measure," he writes, "if *by their conduct they appear* to have little sense of the obligations conferred upon them" *(TMS* II.i.5.3, emphasis added). Though Smith leaves the details unspecified, he clearly thinks it possible to determine a person's actual feelings from his observable behavior alone.[26]

Upon reflection, this seems obvious. Of course we can know that someone who smiles and laughs feels happy, that someone who cries feels sad, that someone who shouts feels angry, and

[26] Smith may be thinking of something broadly Humean: that we infer a person's mental state from her observable behavior by relying on our experience of the correlation between feelings and behavior. But it is also possible that he has something different in mind. Rather than inferring a person's mental state from her observable behavior, for instance, Smith may think that we interpret certain behavioral patterns as *expressions of* certain mental states. There is some reason to think that this is his view. He writes, for instance, of the grief and joy which can be "strongly expressed in the look and gestures of any one" *(TMS* I.i.1.6); of the general lamentations which "express nothing but the anguish of the sufferer" *(TMS* I.i.1.9); and, later, of those passions which, though reasonable to feel, are nonetheless "indecent to express very strongly, even upon those occasions, in which it is acknowledged that we cannot avoid feeling them in the highest degree" *(TMS* I.2.Intro.2). This is similar to Peter Goldie's view, according to which certain forms of behavior are best understood not as signs of some underlying emotional state, but as partly constitutive of that state (see Goldie, 2000, pp. 184-185).

so forth.[27] Whatever the correct explanation of our capacity to interpret other people's behavior, it is obvious that we have such a capacity. And I suspect that it is this fact – the near self-evidence of the claim – that accounts for Smith's failure to make the point explicit. He takes for granted that we can learn about other people's feelings from their behavior and therefore sees no problem in claiming that we can compare such feelings to those we think proper in the relevant circumstances. Smith's objection to Hume's account is not that we cannot *know* other people's feelings based on their behavior alone, but that our *sympathetic feelings* – and, by extension, our moral sentiments – cannot be a function of such behavior alone.

Why, then, does Smith write that 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"? There is an ambiguity in Smith's language that must be cleared up before we can answer this question. To "form an idea of the manner in which they are affected" could mean either of two things: it could mean "know what they feel" or "know what it is like to feel as they do." These are not the same thing. In order to know what another person feels, one would need only to know the answer to the question "What is the other person feeling?" (Goldie, 2000, p. 33). But one can know what another person is feeling without knowing what it is like to feel that way. My mother once told me, for example, that, until she was about eight years old, she did not know what it felt like to be hungry, as her own mother never let her go long between meals. This did not prevent her from recognizing the signs of

[27] The interpretive project is more complex than I make it out to be. People can laugh in aggravation, cry in joy, shout in jest, and so forth. It takes a keen observer to distinguish between these different possibilities. But I will set these complications aside.

hunger in others, though. My mother knew *that* other people were hungry when they behaved in certain ways; she just didn't know *what it felt like* to be hungry.[28]

If, in the passage under consideration, Smith meant that we cannot know *what other people feel* in any other way than by conceiving what we should feel in their situation, then his account would indeed face the problem mentioned in the paragraphs above. But I do not think that this is what Smith means. His point, as I understand it, is that because we do not have direct access to other people's feelings, we can gain no understanding of *what it is like to feel as they do* without consulting our own experiences, real or imaginary. Observing the man on the rack, it is "the impressions of our own senses only, not those of his, which our imaginations copy" *(TMS* I.i.1.2). Our own impressions "never did, and never can, carry us beyond our own person, and it is by the imagination only that we can form any conception of what are his sensations." We can determine *that* the man on the rack is feeling pain from his behavior alone, but we can never understand *what it is like to feel as he does* unless we consult our own experience.

By allowing our determination of other people's feelings to vary independently of our mental simulations, this interpretation avoids the problem identified above. It still faces a problem, however. The interpretation still claims that we can never know what it is like to feel as another person feels without conceiving what we ourselves should feel in the like situation. But surely even this is wrong. Consider the example of an angry man. I have claimed that Smith thinks that I can know that an angry man is angry based on his behavior alone. As long as I also know what it is like to feel angry, I could presumably come to understand what it is like to feel as he does by simply combining my knowledge of what he is feeling with my independent knowledge of what

[28] Goldie tells a similar story about his father, who claimed never to have experienced a headache and thus not to know what one felt like (2000, p. 34). Christopher Peacocke uses the example of seasickness to make the same point (1985, pp. 33-34).

it is like to feel that way. But if this correct, then I could come to understand what it is like to feel as he does without considering his circumstances at all. Granted, my understanding of what it is like to feel as the angry man feels wouldn't necessarily compel me *to feel* angry. If Smith is right, then in order to share the man's anger, we would need to become acquainted with his circumstances. But it is consistent with this to say that I can know what it is like to have such a feeling without being thus acquainted.

These considerations suggest the following modified version of Smith's account. We can determine *that* people feel certain ways from their behavior alone (and perhaps in other ways as well). As we have no direct access to their feelings, however, we can gain no understanding of *what it is like* to feel as they do unless we consult our own (real or imaginary) experiences. The consulted experiences need not be in any way related to the circumstances faced by the people whose feelings we are trying to understand. We do not even need to know their circumstances; and, if we do know their circumstances, we may be able to understand what it is like to feel as they do even if, imagining ourselves in their shoes, we feel very differently. But knowledge of their circumstances is essential if we are to have any chance of *sympathizing* with them – of feeling along with them. It is here that Smith's notion of imaginative mental simulation comes into play. Although we may know what a person feels based on his behavior alone, and we may know what it is like to feel as he does based on our own experience of such feelings, we will not share (or approve of) a person's feelings unless we both know his circumstances and find that, imagining ourselves in their situation, we feel as he does.

3. Sympathy and Approval

"To approve of the passions of another" as "suitable to their objects," Smith writes, is *the same thing* as to observe that we entirely sympathize with them" *(TMS* I.i.3.1, emphasis added). And not to approve of them as such "is *the same thing* as to observe that we do not entirely sympathize with them." Why does Smith insist on such a tight connection between sympathy and approval? And is he right to do so?

Several philosophers have argued that he is not.[29] For instance, Théodore Jouffroy argues that the facts of everyday experience contradict Smith's claim:

> I share a thousand emotions, without morally approving or disapproving them; I condemn many emotions which I share […] and I even approve emotions which I not only do not participate in, but which are absolutely displeasing to me. (1841, p. 146)

Páll Árdal makes a similar observation, noting that one may both "approve a person's anger without in any sense sharing it" and "become angered by the same thing as someone else and at the same time disapprove of both my own anger and his" (1966, pp. 141-142). James Farrer goes so far as to conclude that Smith's account is, for the very reason given by Jouffroy and Árdal, fundamentally flawed. "It is difficult to read Adam Smith's account of the identification of sympathy and approbation," Farrer writes,

> without feeling that throughout his argument there is an unconscious play upon words, and that an equivocal use of the word 'sympathy' lends all its speciousness to the theory he expounds. The first meaning of the word sympathy is fellow-

[29] In addition to the objections reviewed in the text, see the discussions in Campbell (1971, pp. 89-93) and Raphael (2007, pp. 17-20).

> feeling, or the participation of another person's emotion […] the second meaning contains the idea of approval or praise. (1881, pp. 196-197)[30]

I am not persuaded by these objections. In this and the subsequent two sections, I attempt to explain why.

How does Smith defend his claim? He begins with some suggestive examples. The man who resents my injuries as I do, he writes, "necessarily approves of my resentment"; the man who grieves as I do "cannot but admit the reasonableness of my sorrow"; the man who admires a poem as I do "must surely allow the justness of my admiration"; and the man who "laughs at the same joke, and laughs along with me, cannot well deny the propriety of my laughter" *(TMS* I.i.3.1). In each case, Smith's suggestion is that the concurrence of sentiments is inescapably linked to approval, and, conversely, that a conflict of sentiments necessarily entails at least some degree of disapproval. If my resentment or grief extends beyond that level with which my friend can sympathize, if my admiration differs from his, if I "laugh loud and heartily when he only smiles," then I will necessarily earn my friend's (at least partial) disapproval. As my friend's sentiments "are the standards and measures by which he judges of mine," he will necessarily disapprove of my sentiments in proportion to the discordance between them.

But these examples do not constitute the whole of Smith's argument. Their purpose is to clear a path for Smith's primary argument, not to establish his conclusion. Smith's primary argument relies on an analogy between sentiment and belief. He argues that the relation between *approving of* a sentiment and *feeling* that sentiment is the same as that between *approving of* an opinion and *adopting* that opinion:

[30] Thomas Brown offers a similar criticism: he writes that, in Smith's account, sympathy "is generally employed […] to signify a mere participation of the feelings of others; but it is also frequently used as a signification of approbation" (1846, p. 165).

> To approve of another man's opinions is to adopt those opinions, and to adopt them is to approve of them. If the same arguments which convince you convince me likewise, I necessarily approve of your conviction: and if they do not, I necessarily disapprove of it: neither can I possibly conceive that I should do the one without the other. To approve or disapprove, therefore, of the opinions of others is acknowledged, by every body, to mean no more than to observe their agreement or disagreement with our own. But this is equally the case with regard to our approbation or disapprobation of the sentiments or passions of others. *(TMS* I.i.3.2)

Smith's claims are not entirely consistent. The passage begins with the assertion that to approve of another man's opinions is *to adopt* them. Two sentences later, however, Smith says something different: that to approve of other people's opinions is *to observe their agreement* with our own, and that the same relation holds in the case of passions. To make matters worse, Smith states elsewhere that the emotion "in which [a spectator's] sentiment of approbation properly consists" is one "which *arises from* his observing the perfect coincidence between this sympathetic passion in himself, and the original passion in the person principally concerned" *(TMS* I.iii.1.9.fn, emphasis added). The suggestion in this last passage is that the observation of agreement *causes* approval, not that the two are *identical*.

The first of these claims is almost certainly not Smith's real view. Smith suggests a relation of identity between approval and adoption only the one time and, as just noted, contradicts it a mere two sentences later, as well as elsewhere in his work. I will therefore set that view aside. Deciding between the second and third proposals is more difficult, as each enjoys at least some textual support.[31] Campbell (1971) and Raphael (2007) both argue that the third proposal is the most plausible. But the difference between the second and third proposals will not matter for our purposes and I will therefore ignore the distinction here. The important claim (which Smith

[31] For a helpful discussion of this point, see Raphael (2007, pp. 17-19).

undoubtedly endorses) is that there is a one-to-one correspondence between approval and the observation of the relevant type of agreement. That is all I will assume.[32]

But why does Smith believe this to be the case? What does he mean when he claims that we approve of the opinions of others only when we observe them to be in agreement with our own? What sort of approval is he talking about? And what does his discussion of epistemic approval tell us about the relationship between *approval of* and *sympathy with* passions?

4. Evaluating Beliefs

Begin with belief. It is part of the nature of belief, I will assume, that it aims to be true: our epistemic aim in forming and revising beliefs is to believe what is true and avoid believing what is false. We are not always guided by epistemic considerations alone, of course. Practical considerations can sometimes provide us with reasons to believe what is false or to disbelieve what is true. One might have reason, for instance, to believe one's chances of surviving a potentially deadly illness to be higher than they are if doing so would improve one's prospects. In general, one may have a non-epistemic reason to believe a false proposition or disbelieve a true one whenever doing so would have beneficial consequences. Despite this, however, it would clearly be a mistake to analyze belief in terms of non-epistemic considerations alone – as if, lacking any standards of its own, belief could *only* be justified instrumentally. What one believes

[32] Campbell (1971) commits a related interpretive error. He claims that Smith uses the term "sympathy" to denote not "fellow-feeling", but an "*awareness*, on the part of the person sympathizing, that he shares the feelings of another" (1971, p. 94, emphasis in original). There are two reasons to reject this reading. First, Smith never defines or uses the term "sympathy" in this way; he uses the word "to denote our fellow-feeling with any passion" *(TMS* I.i.1.5), not to denote our *awareness of* such a fellow-feeling. Second, if correct, Campbell's interpretation would render Smith's claim regarding the relation between sympathy and approval implausible. If, as Campbell suggests, to sympathize is to be aware of a correspondence between one's feelings and those of one's sympathetic target, then to approve of the passions of another as suitable to their objects would be *to observe that we are aware of* a correspondence between our feelings and those of our sympathetic target – a kind of second-order awareness. But this is implausible, and it is clearly not what Smith has in mind. Sympathy, for Smith, is fellow-feeling, not an awareness of one's fellow-feeling.

no doubt has consequences for the way one's life goes; and in some cases what one ought, all things considered, to believe may depend, in part, on such consequences. But to concede this is by no means to conclude that belief lacks standards of its own. A belief can be correct or incorrect independently of the practical consequences of holding it.

Practical considerations aside, then, truth is the standard of correctness for belief: a belief is correct if and only if it is true (Gibbard, 2005b). This claim might be thought to imply another one: that to evaluate a belief is simply to evaluate its truth. There is a sense in which this is true. Often, when we evaluate another person's belief, what we are doing is attempting to determine whether or not their belief is correct. This view of epistemic evaluation suggests one possible interpretation of Smith's claim regarding the relation between approval and agreement. To evaluate a belief, the reading goes, is to evaluate its truth. But as we have no other way of evaluating the truth of a person's belief than by comparing it to our own – to ask whether *p* is true is just to ask whether *one believes* it to be true – it would follow that to evaluate another person's belief is simply to compare it to our own. I approve of my friend's belief that the Empire State Building is taller than the Statue of Liberty if and only if I myself believe the former to be taller than the latter, and I disapprove if and only if I do not. From the claim that to evaluate a belief is to evaluate its truth, we are led directly to Smith's claim that to approve or disapprove of the opinions of others is "to observe their agreement or disagreement with our own."

In stating this, of course, I do not mean to suggest that Smith thinks that our beliefs can *make* other people's beliefs correct or incorrect.[33] My friend's belief that the Empire State Building is taller than the Statue of Liberty is made correct not by its agreement with my own, but by the

[33] Of course, one person's belief *could* make another person's belief correct or incorrect if the latter person's belief were about what the former person believes. But I will ignore this case here.

fact that the Empire State Building *is* taller than the Statue of Liberty. When I approve of my friend's belief, I do so not because I share it, but because I take there to be adequate reason to hold it.[34] Regardless of whatever ultimately grounds my evaluative conclusion, though, my own beliefs will necessarily be implicated in the evaluative process. I take this to be Smith's point. The act of determining whether my friend ought to believe as he does *compels* me to consult my own beliefs on the matter. "I judge of your sight by might," Smith writes, "of your ear by my ear, of your reason by my reason" (*TMS* I.i.3.10). I *must* judge in this way, he thinks, for "I neither have, nor can have, any other way of judging about them." Though the truth of the proposition believed by my friend is independent of my thoughts on the matter, my determination of its truth is not. My assessment of my friend's belief is necessarily the product of my own.

That is one way that we evaluate belief, at any rate. But it is not the only way. To approve of a person's belief can be to affirm its truth. There is another sense, however, in which to approve of a belief is to do something different: to acknowledge the belief's *justification from the believer's point of view*.[35] Gibbard (2005b) distinguishes between the "objective" and

[34] "If the same arguments which convince you convince me likewise," Smith writes, "I necessarily approve of your conviction: and if they do not, I necessarily disapprove of it: neither can I possibly conceive that I should do the one without the other" *(TMS* I.i.3.2). As Raphael puts the point, Smith's "ground for approval" of another person's opinion "is that there is (what he takes to be) sound argument for the opinion, not the mere fact that he himself shares the opinion" (2007, p. 20).

[35] Campbell suggests a third possibility: granting that "approve usually means 'agree with' or 'think to be correct'," he claims that "approval" and "disapproval" could also "mean holding 'pro' or 'con' attitudes" (1971, p. 92). And he takes this to be an argument against Smith's account: "it is often the case that we agree or disagree with the opinions of others without approving or disapproving of them [in the 'pro' or 'con' sense]. Indeed we may disapprove of someone's expressing an opinion with which we agree in the sense that we think it to be true, since the opinion may not be favourable to our interests. Smith is, therefore, mistaken in making the logical point that, where adopting an opinion and approving of it is concerned, 'I cannot possibly conceive that I should do the one without the other'." Campbell, however, gives no reason to think that this is what Smith has in mind when he writes of "approval" and "disapproval", and there is good reason to think that it is not. Smith explicitly denies that our approval or disapproval of another person's opinion is ever a function of the consequences of their holding it. We "approve of another man's judgment," Smith writes, "not as something useful, but as right, as accurate, as agreeable to truth and reality: and it is evident we attribute those qualities to it for no other reason but because we find that it agrees with our own […] The idea of the utility of all qualities of this kind, is plainly an after-thought, and not what first recommends them to our approbation" *(TMS* I.i.4.4; see also IV.2.3-7). I shall return to this point in Section 5.

"subjective" senses of the "primitive, non-moral ought," which he thinks applies, among other things, to belief:

> You flip a coin and hide the result from both of us. If in fact the coin landed heads, then in the objective sense, I ought to believe that it landed heads […] In the subjective sense, though, I ought neither to believe that it landed heads nor believe that it landed tails. I ought to give equal credence to its having landed heads and to its having landed tails. The coin in fact landed heads, imagine, and so the correct belief for me to have is that the coin landed heads. Such a full belief would be silly, but it would be correct. (2005b, p. 340)

Using Gibbard's terminology, we could say that to approve of a person's belief in the objective sense is to affirm its truth, whereas to approve of a person's belief in the subjective sense is to judge it suitable to the person's epistemic position. I approve of my friend's belief in the objective sense if and only if I believe it to be correct. I approve of it in the subjective sense if and only if I believe it to be what it ought to be, given the information available to my friend at the time. And, importantly, because the information available to my friend may differ from the information available to me, the objective and subjective senses of epistemic approval may come apart. I might approve of my friend's belief in the subjective sense without sharing it, and I might disapprove of my friend for having an unjustified belief that I share.

It is the subjective sense of epistemic evaluation that Smith must have in mind if his argument is to do the work it is intended to do. Recall that Smith's aim in introducing the idea of epistemic evaluation is to draw an analogy between it and the evaluation of the "suitableness or unsuitableness" of a person's sentiments or passions to the cause or object which excites them. The latter is relational: its object is the relation between a particular mental state (a sentiment or passion) and the circumstances in which that mental state has arisen. There is no sense in which a mental state's propriety can be evaluated independently of the circumstances in which it arises, and it is only in the subjective form of epistemic evaluation that we find a relational element

analogous to Smith's fundamentally relational notion of emotional and behavioral propriety. If it is to succeed, then, Smith's analogy between epistemic evaluation, on the one hand, and the evaluation of emotional and behavioral propriety, on the other, must be understood in terms of subjective, not objective, epistemic evaluation. Smith's claim must be that we evaluate the propriety of people's sentiments or passions in something like the way we evaluate the *justification* – not the *truth* – of their beliefs.

If this is correct, then to approve of another person's belief is not necessarily to adopt it; it is to find that one *would* adopt it *if* one were in the other person's epistemic position. This fits nicely with some of Smith's claims. Consider, for example, his claim that "[i]f the same arguments which convince you convince me likewise, I necessarily approve of your conviction: and if they do not, I necessarily disapprove of it." This passage points not to the *truth* of the belief as the determinant of one's approval or disapproval, but rather to *the arguments* on which the belief is based. Because you and I may have access to different information – we may have been exposed to different arguments – I may approve of your belief without sharing it or disapprove of it even as I share it. I may approve of your reasoning but reject your conclusion, or endorse your conclusions but disapprove of your reasoning.

Unfortunately, Smith makes other statements that appear to contradict this reading. To "approve of another man's opinions is to adopt those opinions," he writes, "and to adopt them is to approve of them." And elsewhere, in an effort to distinguish his view from Hume's, Smith notes that

> we approve of another man's judgment, not as something useful, but as *right*, as *accurate*, as *agreeable to truth and reality*: and it is evident we attribute those qualities to it for no other reason but because we find that it agrees with our own. *(TMS* I.i.4.4, emphasis added)

If, as I have argued, Smith is thinking of epistemic evaluation in the subjective sense, then why would he claim that to approve or disapprove of the opinions of others is nothing more than "to observe their agreement or disagreement with our own"? Why would he insist that we approve of another man's judgment not as something useful, but as "agreeable to truth and reality"?

The answer, I think, is that Smith does not consider the possibility of significant variation between different people's epistemic positions. If there were no variation in people's epistemic positions, then the difference between agreeing with someone, on the one hand, and finding that one *would* agree with them *if* one were in their shoes, one the other, would disappear – as would the difference between thinking the person's judgment justified and thinking it "agreeable to truth and reality." Smith's claims regarding the relation between epistemic approval and *actual* agreement could be understood as shorthand for the claim that approval entails *conditional* agreement.

As it turns out, there is good reason to think that this is what is going on. Shortly after presenting his analogy, Smith distinguishes between two types of situations in which we evaluate the opinions or sentiments of others. The first is one in which both the evaluator and the individual under evaluation stand in exactly the same position relative to the "objects which excite" the mental state in question *(TMS* I.i.4.1). In the second, the two stand in different positions. It is only in the first that Smith claims that to approve of someone's passion or opinion is to find that it corresponds perfectly to one's own:

> The beauty of a plain, the greatness of a mountain, the ornaments of a building, the expression of a picture, the composition of a discourse, the conduct of a third person, the proportions of different quantities and numbers, the various appearances which the great machine of the universe is perpetually exhibiting, with the secret wheels and springs which produce them; all the general subjects of science and taste, are what we and our companions regard as having no peculiar relation to either of us. We both look at them from the same point of view, and we

> have no occasion for sympathy, or for that imaginary change of situations from which it arises, in order to produce, with regard to these, the most perfect harmony of sentiments and affections. *(TMS* I.i.4.2)

Because I stand in the same relation to the Empire State Building and Statue of Liberty as my friend, there is no need, when evaluating my friend's belief regarding their relative heights, for me to consider the opinion that I *would* have *if* I were to occupy his position. Because my friend's epistemic position is identical to my own, the opinion I *would* have *if* I were to occupy his position is just the opinion I *do* have.

The second sort of situation is different. When our evaluations involve objects "which affect in a particular manner either ourselves or the person whose sentiments we judge of," Smith thinks it inappropriate to proceed by comparing the other person's mental state to our own *(TMS* I.i.4.5). In such cases, he writes that it is necessary for the evaluator

> 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* I.i.4.6)

Because we do not initially view all individual misfortunes, injuries, and so forth, "from the same station, as we do a picture, or a poem, or a system of philosophy," we must shift our perspective before reaching any evaluative conclusions.

To summarize: We approve (in the relevant sense) of another person's opinion not when we think it correct, but when we think it justified, warranted, or well-suited to the circumstances in which it has been formed. And we determine whether it is well-suited to its circumstances, according to Smith, by asking ourselves whether we would hold the belief if we were in that person's shoes. To approve of another person's opinions is not, as Smith literally (and somewhat

carelessly) says, "to observe their agreement" with our own, but rather to observe their agreement with what ours *would* be – though the two will of course turn out to be the same whenever our epistemic positions are identical, as is often the case.

5. Evaluating Passions

Having discussed belief, let us turn now to passions. In the case of (subjective) epistemic evaluation, the objective is to assess a belief's suitability to the circumstances in which it has been formed. In the case of sentiments or passions, Smith thinks that the objective is to assess their suitability to the object or circumstances which excited them. As noted in Section 1, this is one of the features of Smith's account that distinguishes it from Hume's. For Hume claims that our assessments of passions are a function of their typical consequences alone: that we approve or disapprove of people's passions insofar as they tend, on average, to produce pleasurable or painful states in people. According to Smith, however, Hume's analysis omits a crucial feature of our evaluative judgments. When we blame a man for excessive love, grief, or resentment, for instance, Hume's account commits him to saying that we do so *only* because such emotional displays tend, on average, to be detrimental to the interests of society. But Smith denies that our focus is exclusively, or even primarily, on "the ruinous effects" that such emotions tend to produce." The primary determinant of our adverse judgments is not the effects of such emotions, "but the little occasion which was given for them."

> The merit of his favourite, we say, is not so great, his misfortune is not so dreadful, his provocation is not so extraordinary, as to justify so violent a passion. We should have indulged we say; perhaps, have approved of the violence of his emotion, had the cause been in any respect proportioned to it. *(TMS* I.i.3.8)

We disapprove of excessive emotions not because they are harmful, but because they are unjustified, unwarranted, or poorly suited to their circumstances.

The parallel between the evaluation of belief and the evaluation of passion is clear. As noted in the previous section, we are not always guided by epistemic considerations alone when forming and revising our beliefs; practical considerations can sometimes provide us with non-epistemic reasons to hold epistemically unjustified beliefs. As Smith points out, though, it does not follow from this that *all* epistemic evaluation is carried out in such consequentialist terms. "Originally," Smith writes, "we approve of another man's judgment, not as something useful, but as right, as accurate, as agreeable to truth and reality" *(TMS* I.i.4.4). Smith's claim is that the same holds for our judgments concerning sentiments and passions:

> Taste, in the same manner, is originally approved of, not as useful, but as just, as delicate, and as precisely suited to its object. The idea of the utility of all qualities of this kind, is plainly an after-thought, and not what first recommends them to our approbation. *(TMS* I.i.4.4)

Smith objects to Hume's account, among other reasons, because it ignores this. There are standards *internal to* emotional and behavioral propriety, just as there are standards *internal to* epistemic justification.

Already we are in a position to see one problem with Jouffroy's objection to Smith's account. Jouffroy's, recall, writes:

> I share a thousand emotions, without morally approving or disapproving them; I condemn many emotions which I share […] and I even approve emotions which I not only do not participate in, but which are absolutely displeasing to me.

Focus on the last claim: that we sometimes approve of emotions which are *displeasing* to us. This is undoubtedly true, but it is also irrelevant. For it is Hume, not Smith, who claims that we

approve of those passions that bring us pleasure and disapprove of those that bring us pain (and that the relevant pleasures and pains are the result of our sympathy with the passions' typical effects).[36] Smith shares Hume's sentimentalist metaethic, but he rejects Hume's claim regarding the relation between approval and pleasure, disapproval and pain. This part of Jouffroy's criticism is entirely misplaced.[37]

What about Jouffroy's claim that we can share emotions without approving or disapproving of them, condemn emotions that we share, and approve of emotions that we do not share? Unlike the previous criticism, this objection is at least on point. But it is still mistaken. Smith anticipates and responds to just this sort of argument, taking care to distinguish his view from the sort of view to which Jouffroy's criticism would apply.[38] He notes, for example, that we may approve of a joke and "think the laughter of the company quite just and proper, though we ourselves do not laugh, because […] we are in a grave humour, or happen to have our attention engaged with other objects" *(TMS* I.i.3.3). Though in our present mood we cannot join the laughter, we nonetheless approve of the company's laughter because "we are sensible that upon most occasions we should very heartily join in it." Similarly, Smith notes that upon meeting a stranger in the street who has just received news of the death of his father, we will undoubtedly approve of his grief, even if we fail to feel any part of his sorrow:

> Both he and his father, perhaps, are entirely unknown to us, or we happen to be employed about other things, and do not take time to picture out in our imagination the different circumstances of distress which must occur to him. We have learned, however, from experience that such a misfortune naturally excites

[36] As Hume puts it, "moral distinctions depend entirely on certain peculiar sentiments of pain and pleasure, and that whatever mental quality in ourselves or others gives us a satisfaction, by the survey or reflection, is of course virtuous; as every thing of this nature, that gives uneasiness, is vicious" (*Treatise* 3.3.1.3).

[37] Campbell makes the same error when he objects to Smith's account by noting that "we may disapprove of someone's expressing an opinion with which we agree in the sense that we think it to be true, since that opinion may not be favourable to our interests" (1971, p. 92). See footnote 35 for a discussion of this point.

[38] As noted above, Árdal (1966) makes the same criticism, as do both Brown (1846) and Farrer (1881). Árdal is the only one who acknowledges Smith's anticipation of and response to the objection (see p. 142).

> such a degree of sorrow, and we know that if we took time to consider his situation, fully and in all its parts, we should, without doubt, most sincerely sympathize with him. It is upon the consciousness of this conditional sympathy, that our approbation of his sorrow is founded, even in those cases in which that sympathy does not actually take place. *(TMS* I.i.3.4)

And this is precisely what one would expect Smith to say, based on his discussion of epistemic approval. Just as we can approve (in the subjective sense) of another person's belief without sharing it, Smith thinks we can approve of another person's sentiment without feeling it. And, similarly, just as we can disapprove of another person's belief that we share – say, if we think their belief to be based on insufficient evidence – Smith would undoubtedly say that we can disapprove of another person's sentiment as unsuitable to the circumstances which provoked it even if we feel that same sentiment in our own circumstances. The sentiment in question might be inappropriate in the other person's circumstances, but appropriate in our own. Smith's claim is not, as Jouffroy appears to think, that we approve of those sentiments and passions that we share and disapprove of those that we do not share. It is that we approve of those sentiments and passions of others that we *would* share *if* we were to imagine ourselves in their shoes, and disapprove of those that we *would not* share *even if* we were to imagine ourselves in their shoes.[39] Consequently, Jouffroy-style counterexamples cannot touch Smith's account.

Smith's account of sympathetic evaluation is thus analogous to subjective epistemic evaluation insofar as it (i) proceeds via the consideration of another person's point of view and (ii) targets the relation between the mental state in question and the circumstances in which it arose (as opposed to that mental state's consequences). It is analogous to epistemic evaluation in an additional respect as well. As we have seen, Smith claims that when evaluating another person's belief, we have no choice but to compare it to our (conditional) belief. We have no

[39] Gibbard (2005a) draws this same distinction, calling Smith's analysis "dispositional" rather than "expressivist".

choice, he claims, because we have no other way of evaluating another person's belief. And this point, he thinks, is general:

> 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, 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* I.i.3.10, emphasis added)

Just as we *must* evaluate other people's beliefs by comparing them to the beliefs that we ourselves would have in their circumstances, Smith claims that we must evaluate other people's passions by comparing them to the passions we ourselves feel when we imagine ourselves in their shoes. We *must* evaluate their passions in this way, he insists, as we "neither have, nor can have, any other way of judging about them." Our own passions are necessarily implicated in the process of evaluating other people's passions, just as our own beliefs are implicated in the process of evaluating other people's beliefs.

There is, however, one important difference between Smith's accounts of epistemic and non-epistemic evaluation worth mentioning – one respect in which they are *not* analogous. When we evaluate another person's belief, I have claimed that Smith thinks we proceed by asking ourselves what we would believe if we were in their shoes. We approve if we find that we would believe what they believe, and we disapprove if we find that we would not. If this is correct, and if the evaluation of belief and passion were perfectly analogous, then Smith would need to say that we evaluate other people's passions by asking ourselves how we would feel if we were in their shoes. Perhaps surprisingly, though, this is not what Smith says. Smith's claim is not that we judge a person's feeling proper if we find that we would feel the same way if we were in his

shoes; it is that we judge it proper if we find that we would feel the same way *if we were to imagine that we were in his shoes*.[40]

As we shall see in the next chapter, this small difference will have important implications for Smith's account.

[40] This distinction is often missed. Frierson (2006b) and Sayre-McCord (2010) correctly emphasize it.

CHAPTER 3. SYMPATHY AND THE IMAGINATION

1. Empathetic Simulation vs. In-His-Shoes Simulation

In the previous chapter, I reviewed the general features of Smith's account of sympathy. In particular, I emphasized Smith's claim that we attempt to sympathize with other people by imagining and simulating a response to their circumstances, and I explained that Smith portrays the process of imaginative mental simulation as evaluative: we imagine and simulate a response to other people's circumstances, he claims, neither to identify nor to experience their *actual* feelings, but to identify the feelings that would be *proper* in their situation. Having described the general framework of Smith's account, let us turn now to its details. In particular, let us ask: When we imagine ourselves in another person's shoes, what is it that Smith thinks we imagine? Do we merely imagine being in their shoes? Or do we imagine being *just like* them in their shoes? Or something else?[41]

Several passages in *The Theory of Moral Sentiments* support the second option. As we have seen, for instance, when attempting to sympathize with a man who is suffering, Smith writes that we "place ourselves in his situation, […] enter as it were into his body, and become in some measure *the same person* with him" (*TMS* I.i.1.2, emphasis added). Even more explicitly, he insists that the "imaginary change of situations" from which sympathy arises "is not supposed to

[41] The answer might seem obvious: according to Smith, we imagine ourselves neither as ourselves nor as another particular person, but "impartially, *as any one of us*" (Darwall, 1999, p. 142, emphasis in original). Ultimately, I will argue that this is correct. But it will take some work to explain what it means to imagine a person's circumstances in this. This chapter could be seen as the first step toward a complete explication of Smith's "impartial spectator" model of moral evaluation. I discuss the "impartial spectator" model in detail in Chapters 4 and 5 and explain how impartiality regulates our sympathy with (and assessments of) other people in Chapter 6.

happen to me in my own person and character, but *in that of the person with whom I sympathize*" (*TMS* VII.iii.1.4, emphasis added).

> When I condole with you for the loss of your only son, in order to enter into your grief I do not consider what I, a person of such a character and profession, should suffer, if I had a son, and if that son was unfortunately to die: but I consider what I should suffer *if I was really you*, and I not only change circumstances with you, *but I change persons and characters*. (*TMS* VII.iii.1.4, emphasis added)

These passages suggest that Smith thinks we attempt to sympathize with others by imagining not just being in their shoes, but being *just like* them.

This type of imaginative process – involving a complete imaginative identification with one's target – is thought by some to be essential to what is today called "empathy".[42] Peter Goldie, for example, writes that to empathize with someone requires more than just imaging oneself in their circumstances. It requires "bringing to bear in the imaginative process a *characterization*" of that person, including not only a complete psychological profile (i.e., character traits and emotional dispositions), but a list of all relevant non-psychological attributes (e.g., their height, that they work for minimum wage) (2000, p. 198, emphasis in original). I will call this sort of process *empathetic simulation*.

Could this be what Smith has in mind? Does he think that we attempt to sympathize with others by assimilating their personal characteristics and then simulating a response to their circumstances as if we were just like them? The interpretation faces an obvious objection. Recall that Smith's simulation mechanism is supposed to play an evaluative role in his account: we simulate responses to other people's circumstances, he claims, neither to identify nor to experience their actual feelings, but to establish a standard of propriety against which to compare

[42] See, for example, Goldie (2000), Goldman (1995a), Goldman (1995b), Goldman (2006), Gordon (1995a), Gordon (1995b), and M. Hoffman (2000).

their conduct. In order to succeed in this role, it would need to be possible for our simulated responses to differ from people's actual responses.[43] The problem is that the process of empathetic simulation seems to rule out such differences. If, when imagining myself in the shoes of another, my simulated response differs from his, then it would seem that I must not really be imagining myself as if I were just like him. For if I *were* imagining myself as if I were just like him, I would presumably simulate his exact response.[44] I would *necessarily* approve.[45]

Smith, of course, thinks (and is right to think) that we can, and often do, judge other people's feelings and behavior improper, and he denies that our simulated responses to other people's circumstances always correspond to reality.

> We sometimes feel for another, a passion of which he himself seems altogether incapable; because, when we put ourselves in his case, that passion arises in our breast from the imagination, though it does not in his from the reality. (*TMS* I.i.1.10)

We are embarrassed by the "impudence and rudeness" of a person who behaves inappropriately, for example, "though he himself appears to have no sense of the impropriety of his own behavior." We feel anguish for the "poor wretch" who has lost his reason, even as he laughs and sings, "altogether insensible of his own misery" (*TMS* I.i.1.11). The problem with empathetic simulation is that it seems incapable of accounting for such experiences. It seems incapable, that

[43] Charles Griswold makes a similar point: "Smith's insistence on the priority of entering into another person's situation, rather than simply of entering into another person's feelings, is important […] If we were unable to see the situation except from the standpoint of the person affected or identified completely with the agent's emotions, no independent evaluation would be possible" (1999, p. 87). See also Frierson (2006a) and Nanay (2010).

[44] To keep things simple, I have assumed that there is only one way that a person with a given set of characteristics could respond to a given set of circumstances. Dropping this assumption would not change anything, however. The relevant question would just become: is the person's actual response included in the set of responses that I can simulate when imagining myself as if I were just like him? If the above argument works in the simplified case, then it would work in the more complex case as well.

[45] Of course, if the person disapproves of his own conduct, then, by imagining myself as if I were just like him, I would presumably be able to simulate not only his response to his circumstances, but also his disapproval of his response. One form of disapproval would thus be possible. But I would be incapable of offering any *independent* disapproval. Moreover, the account would remain open to the objection presented on the next page.

is, of explaining how our simulated responses to other people's circumstances could ever differ from their actual responses, or how we could ever disapprove. Empathetic simulation requires complete imaginative identification; but evaluation, it would seem, "requires holding back" (Gordon, 1995b, p. 740).

The empathetic interpretation faces another objection as well. As noted in the previous chapter, Smith believes that our evaluative judgments are grounded in *our own* affective tendencies. When we evaluate a person's affection, he writes, "it is scarce possible that we should make use of any other rule or canon but the correspondent affection in ourselves" (*TMS* I.i.3.9).

> 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, 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* I.i.3.10)

Strictly speaking, Smith's claim is implausible.[46] If I knew my eyesight to be poor, for instance, I would not judge another person's eyesight by measuring it against my own; I would apply some other evaluative method – a test performed by a licensed optometrist, say. But I would like to set such cases aside. The important point is that Smith thinks that our judgments are typically grounded in (a subset of) *our own* characteristics. The trouble with empathetic simulation is that it appears to contradict this claim. By forcing us to use *all and only other people's* characteristics as inputs in our simulation, the mechanism appears to exclude our own characteristics from the evaluative process.

In light of these problems, it is tempting to consider the opposite extreme: that Smith thinks we evaluate other people's conduct by imagining *ourselves* in their circumstances, complete with

[46] For helpful criticisms, see Campbell (1971) and Raphael (2007).

our own profile of psychological and non-psychological characteristics. Borrowing again from Goldie (2000), I shall call this *in-his-shoes simulation*. The reading fits nicely with Smith's claims regarding the autonomy of our evaluative judgments, as well as his explanations of cases involving disapproval. "We blush for the impudence and rudeness of another, though he himself appears to have no sense of the impropriety of his own behavior," Smith writes, "because we cannot help feeling with what confusion *we ourselves* should be covered, had we behaved in so absurd a manner" (*TMS* I.i.1.10, emphasis added). And, similarly, Smith claims that the spectator's anguish on behalf of the individual who has lost his reason arises "from the consideration of what *he himself* would feel if he was reduced to the same unhappy situation, and, what perhaps is impossible, was at the same time able to regard it *with his present reason and judgment*" (*TMS* I.i.1.11, emphasis added). In each case, Smith's claim seems to be that whenever we disapprove of another person's conduct, we do so because it differs from what *we* would do in their shoes.

Unfortunately, this interpretation faces problems as well. To begin, though consistent with the passages cited in the previous paragraph, it contradicts Smith's other claims – e.g., that when I condole with you for the loss of your son, I "not only change circumstances with you, but I change persons and characters" (*TMS* VII.iii.1.4). The in-his-shoes reading of Smith's account can do no more to solve the interpretive puzzle posed by Smith's seemingly contradictory remarks than the empathetic reading. Moreover, the interpretation is independently objectionable insofar as it attributes to Smith a rather unappealing account of evaluative judgment. Though not infinitely flexible, our judgments of propriety are surely flexible enough to allow people with different personal characteristics to respond differently to the same circumstances – at least in some cases. There may be a range of proper responses to a given set of circumstances, each one

fitting for a different sort of person. (Recall Hare's trumpet example from Chapter 1.) In-his-shoes simulation is objectionable because it denies this, and considerations of charity speak against attributing to Smith an objectionable view (provided some less objectionable alternative can be found).

We need an alternative to empathetic and in-his-shoes simulation, and the discussion thus far suggests four conditions that any successful proposal would need to satisfy. First, the mechanism must be sensitive to differences in personal characteristics capable of justifying different emotional and behavioral responses to the same circumstances. Second, it must provide one with the independence needed to evaluate people critically. Third, it must ground one's judgments in one's own affective tendencies. And fourth, it must reconcile Smith's seemingly contradictory claims: that we attempt to sympathize with others by *changing persons and characters* with them, but sometimes disapprove when we find that *we* would have conducted ourselves differently. In short, a successful simulation mechanism must provide a *critical* perspective that is simultaneously *respectful* of other people's points of view and *endorsable* from one's own. The challenge is to construct a simulation mechanism that successfully balances these competing demands – and that does so in a way that is consistent with Smith's claims.

2. Some Alternative Proposals

How might these demands be balanced, and Smith's claims reconciled? According to Páll Árdal, we have but one choice: "we must understand the imaginative substitution of place differently in different cases" (1966, p. 136). Sympathizing with a person's sorrow, for example, Árdal claims that we must imagine both being in that person's shoes and having that person's "sensitivities". When we disapprove of a person's inappropriate behavior, though, Árdal insists that we must

imagine ourselves not with that person's sensitivities, but with our own. We must alternate between empathetic and in-his-shoes simulation, in other words, as the circumstances dictate.

Robert Gordon suggests something similar. When we imagine ourselves in another person's shoes for the purpose of offering advice, Gordon writes that "it is always important" to

> hold back in certain ways from identification with the other person – that is, from making the further adjustments required to imagine being not just in that person's situation but *that person* in that person's situation. Otherwise we lose the very advantages that make our advice worthwhile: the special know-how or the independent judgment. (Gordon, 1995b, p. 740, emphasis in original)[47]

When formulating practical advice for another, we imagine ourselves neither as ourselves nor as the other, but as something in between. We incorporate *some but not all* of the other person's characteristics into our imaginative simulation; we rely on our own characteristics to provide the remainder.

Though reasonable, these proposals are incomplete, as they leave unexplained how we are to determine how to simulate in any given case. How do we know whether to simulate empathetically or in-his-shoes? How do we know which characteristics to "hold back"? These are precisely the questions with which we began our investigation; they are the questions we are trying to answer. But neither proposal makes any attempt to answer them. Instead, they assume that we already have the answers – that we already know how to simulate in each and every case.

How might we know this? What would we need to know in order to know how to simulate in any given case? According to Annette Baier, Smith's account "requires a judgment about propriety to mediate any real sympathy" (2008, p. 78). That is to say, she claims that we would "need to decide whether or not another deserves our sympathy, whether her shoes are clean

[47] Gordon's comments in this passage are not about Smith. But he connects his claim to Smith's account in the subsequent paragraph.

enough for us to step into, before letting her have any sympathy." If Baier is right, then Smith's account is in trouble. For if we had some independent capacity to make reliable judgments of propriety, then Smith's account of sympathy – the centerpiece of his entire moral theory – would be superfluous; and his claim that "to approve of the passions of another" as proper "is the same thing as to observe that we entirely sympathize with them" would be false. If our capacity to sympathize (or not) with other people required a prior capacity to make evaluative judgments, then Smith's account would fail.

In order to show that Smith's account does not fail, I must show that Baier's claim is false: that instead of relying on a prior judgment of propriety to guide our sympathetic experiences, Smith's account can answer all relevant questions endogenously. In the next two sections, I attempt to do just that.

3. Empathetic Simulation Redux: Divided Sympathy and Moderation

In Section 1, I raised a possible objection to the empathetic interpretation of Smith's account. I suggested that by insisting that Smith thinks we sympathize with others by imagining being just like them, the interpretation precludes the possibility of disagreement and disapproval. For if I were to imagine myself as if I were just like another person, the objection goes, I would *necessarily* reproduce that person's conduct in each and every case. I would be incapable of disapproval.

This objection relies on an assumption: that, given the right inputs, we *can* simulate anything. Or, put differently: that whenever we fail to simulate a particular response to a particular situation, it is necessarily the case that we *decided* to fail. This assumption is apparent in Árdal's commentary – for instance when, while discussing our disapproval of Smith's obliviously rude

man, he writes that "we could not imagine that we were [that] person, with all his psychological characteristics, for then we should presumably be as insensitive to the situation as he is" (1966, p. 136). It is implicit in Gordon's claim that evaluation "requires holding back" (1995b, p. 740) and made explicit in Baier's insistence that Smith's account requires us "to *decide* whether or not another deserves our sympathy […] before letting her have any" (2008, p. 78, emphasis added). It is the primary reason for rejecting the empathetic interpretation of Smith's account.

But the assumption is almost certainly false, and Smith clearly rejects it. We always *want* to sympathize with others, Smith believes, and would never *choose* not to do so; there are just some cases in which we *cannot*. He writes, for instance, that we are "pleased when we are *able* to sympathize" with others, and "hurt when we are *unable* to do so"; that "it is always disagreeable to feel that we *cannot* sympathize" with someone; and that "it hurts us to find that we *cannot* share his uneasiness" (*TMS* I.i.2.6, emphasis added). Discussing the link between sympathy and approval, Smith writes that we "approve or disapprove of the conduct of another man according as we feel that, when we bring his case home to ourselves, we either *can* or *cannot* entirely sympathize with the sentiments and motives which directed" (*TMS* III.1.2, emphasis added). And he refers to our incapacity to sympathize in his examples of disapproval as well. We blush for the rudeness of another, Smith writes, "because we *cannot* help feeling with what confusion we ourselves should be covered, had we behaved in so absurd a manner" (*TMS* I.i.1.10, emphasis added). Similarly:

> A person becomes contemptible who tamely sits still, and submits to insults, without attempting either to repel or to revenge them. We *cannot* enter into his indifference and insensibility. (*TMS* I.ii.3.3, emphasis added)

And, along the same lines, he writes that when we hear a person lamenting "misfortunes" that "*can* produce no such violent effect upon us, we are shocked at his grief […] and, because we

cannot enter into it, call it […] weakness" (*TMS* I.i.2.6, emphasis added). We are "disobliged" even with a person's joy when we "*cannot* go along with it" and "call it levity and folly."

Once we realize that Smith does not think that we *can* sympathize with just anything, a new strategy presents itself. Rather than searching for some external rule to guide our simulations (i.e., something to tell us when to simulate empathetically, when to simulate in-his-shoes, which of our characteristics to hold back, or something along these lines), we might begin instead with empathetic simulation and then look for one or more *internal* mechanisms capable of blocking our sympathy. In place of an exogenous rule, in other words, we might search for an endogenous process that would lead us to sympathize with – and approve of – some things, but not others.

What might block our sympathy? One answer is obvious: our sympathy will be blocked if we fail to make the adjustments needed to simulate empathetically – e.g., if we fail to imagine a person's circumstances accurately or in sufficient detail, or fail to "bring to bear" their complete set of personal characteristics. I shall refer to such things as *cognitive sympathetic failures.* They are not uncommon. Individuals often fail to accurately account for other people's characteristics when attempting to understand their mental states. They rely on crude stereotypes (Duncan, 1976; Hugenberg & Bodenhausen, 2003; Sagar & Schofield, 1980) or project their own knowledge or personal characteristics onto their targets (Goldman, 2006, pp. 40-42). The latter mistake is particularly common. People often assume that other people share their attitudes, even when they do not (Krueger, 2000; Ross, Greene, & House, 1977); they (especially young children) act as if other people have access to the same information to which they themselves have access, even when they know this to be false (Camerer, Loewenstein, & Weber, 1989; Gopnik & Astington, 1988); and they allow their determination of what other people feel to be influenced by what they themselves feel (Van Boven & Loewenstein, 2003). In addition to

distorting our understanding of other people's minds, such errors could very well hinder our attempts to sympathize (Mackenzie, 2006). But cognitive sympathetic failure cannot be our answer. Though cognitive errors can block our sympathy, they could never ground a legitimate judgment of disapproval. When we fail to sympathize as a result of cognitive error, we do not disapprove; we fail to correctly judge (see *TMS* I.i.3.4 and I.i.4.6).

Assume, then, that we commit no cognitive error. What else might block our sympathy? Smith mentions two possibilities. First, he notes that whenever we evaluate a person whose conduct affects another, our attention is split between the two: we imagine and simulate responses to *both* sets of circumstances, and each simulation affects the other. Consider, for instance, the case of a person who feels hatred or resentment toward another:

> With regard to all such passions, our sympathy is divided between the person who feels them, and the person who is the object of them. The interests of these two are directly opposite. What our sympathy with the person who feels them would prompt us to wish for, our fellow-feeling with the other would lead us to fear. (*TMS* I.ii.3.1)

Our sympathetic fear on behalf of the one "damps our resentment" for whatever the other has suffered. As a result, we do not necessarily feel what resentful people feel, even if we imagine ourselves with all their characteristics. Our divided sympathy opens the door to discordance and disapproval.

Smith's discussion of divided sympathy is important, but it cannot be the whole of his story. For if we could only judge something improper in virtue of its harmful effects on others, then Smith's account would be nearly indistinguishable from Hume's – which, as we have seen, holds that we approve or disapprove of motives based on their beneficial or harmful effects. But Smith thinks that his account differs substantially from Hume's. He criticizes philosophers "of late

years" (by which he clearly means Hume) for focusing "chiefly" on the tendency of affections to produce beneficial or harmful effects, and giving "little attention to the relation which they stand in to the cause which excites them" (*TMS* I.i.3.8). He explains that the "difference" between their accounts and his is that they make "utility, and not sympathy, or the correspondent affection of the spectator, the natural and original measure" of propriety (*TMS* VII.ii.3.21). *They* may measure propriety in terms of utility, but Smith clearly states that *he* does not. There must be more to his account than the phenomenon of divided sympathy.

This brings us to Smith's second claim. Regardless of our effort, Smith insists that our vicarious feelings will nevertheless "be very apt to fall short" of what is felt by the person with whom we are attempting to sympathize. "Mankind," he writes, "though naturally sympathetic, never conceive, for what has befallen another, that degree of passion which naturally animates the person principally concerned" (*TMS* I.i.4.7). We may experience something "analogous", but we will generally fall short of "conceiving any thing that approaches to the same degree of violence." Because I am not presently in danger, I will most likely fail to evoke the sort of fear that I would experience if I were truly in danger. Because I am not presently in the midst of a heated disagreement, I will most likely fail to evoke the degree of anger that I would experience if I were. I may be able to conjure up feelings *like* those that I would actually experience, but they will be moderated.

Like his discussion of divided sympathy, Smith's discussion of our moderated sympathetic responses is important, as it reveals an additional source of potential discordance and disapproval. Viewing things from a spectator's perspective, we are less likely to be caught in the grip of some "hot" emotion (see Metcalfe & Mischel, 1999), and thus better positioned to engage in a calm and thoughtful deliberative process. We will be better judges of propriety. Once again,

though, Smith's account of our moderated responses cannot be the whole of his story. For although Smith thinks that we often disapprove of people for feeling *too much*, there are also cases in which he thinks that we disapprove of people for feeling *too little*:

> The man who should feel no more for the death of his own father, or son, than for those of any other man's father or son, would appear neither a good son nor a good father. Such unnatural indifference, far from exciting our applause, would incur our highest disapprobation. (*TMS* III.3.13)

Indeed, Smith even thinks that we sometimes disapprove of people for *failing to feel something at all*:

> A person becomes contemptible who tamely sits still, and submits to insults, without attempting either to repel or to revenge. We cannot enter into his difference and insensibility. (*TMS* I.ii.3.3)

Neither our divided sympathy nor our tendency to experience moderated emotional responses from a spectator's perspective can explain our disapproval in such cases. Nor can they explain our disapproval of a person who feels *the wrong kind* of emotion in a particular situation (e.g., one who laughs when he ought to cry). Smith's account needs something more. Unfortunately, he offers no clear explanation of what this additional "something" might be. In the next section, I attempt to provide the missing details on Smith's behalf.

4. Empathetic Simulation Redux: Active Imagination and Its Limits

Distinguish between two types of imaginative activity: *propositional imagination* (or imagining *that p*) and *active imagination* (or imagining *Φ-ing*).[48] To imagine *that* something is the case, I will claim, is just to suppose it to be the case (Goldman, 2006, p. 47). It is easy to do. Taking an

[48] On this distinction, see Goldie (2000), Goldman (1995a), Goldman (2006), Moran (1994), Walton (1990), Walton (1994), Walton (1997), and Wollheim (1984).

example from Goldman (2006), suppose, for instance, that I wished to consider how history would have unfolded if the United States had lacked the atomic bomb in 1945. The first step would be to imagine that the United States had lacked the bomb. I could then engage in a (purely cognitive) process of counterfactual historical aimed at determining what would have followed. Such counterfactual reasoning would be difficult, of course; there is no guarantee that I would reach the correct conclusion. But the imaginative act that precedes it – imagining (or supposing) that the United States lacked the bomb – would be easy.

Active imagination is different. To imagine *Φ-ing* is not merely to suppose that one *Φ*'s. To imagine being angry, for instance, requires more than mere supposition; it requires *enacting* the feeling of anger (Goldman, 2006, p. 47). This can be difficult. It is trivially easy to imagine *that* I am angry or afraid or sad; it can be more difficult, though, to imagine *being* angry, afraid, or sad when I am not any of these things – to enact the feelings themselves. More to the point, it can be difficult to imaginatively enact a particular feeling *in response to* or *directed at* a particular situation or object – or to *avoid* feeling something when imagining a particular situation or object. Consider, for example, the following joke,[49] taken from Kendall Walton (1994):

> "Knock, knock."
> "Who's there?"
> "Robin."
> "Robin who?"
> "Robbin' you! Stick 'em up!"

Like Walton, I don't find this joke to be particularly funny. But, again like Walton, I have no trouble imagining finding it funny – or, at least, imagining being amused by it, if not quite

[49] Smith himself uses humor as an example to illustrate his account of sympathy (see *TMS* I.i.3.1).

finding it funny (if there is a distinction to be made between the two).[50] The knock-knock joke is for this reason quite unlike a non-joke statement like "A maple leaf fell from a tree," which, in addition to finding neither funny nor amusing, I am incapable of even imagining finding funny or amusing – though I would have no difficulty imagining that I find it funny or amusing, strange as that would be.

When we attempt to sympathize with someone, we do not proceed by imagining *that* we are just like them and in their circumstances, and then reasoning our way to a response. If that were Smith's view, then his account would indeed preclude all discordance and disapproval. Rather than imagining *that* particular things are true of us, Smith's claim, as I read it, is that we attempt to sympathize with people by imagining *doing* the things they do, *experiencing* the things they experience, *feeling* the things they feel, and so forth (Walton, 1997, p. 38). We attempt to sympathize with others by means of an *active* form of imaginative engagement. And although we will succeed in some cases, there are others in which we will find our sympathy blocked. Though I can sympathize with a person who, unlike me, finds the knock-knock joke funny, I am incapable of sympathizing with someone who finds the non-joke about the maple leaf funny.

Why is this? What is it that enables me to imagine finding the knock-knock joke funny (though I don't actually find it funny), but prevents me from doing the same for the maple leaf comment? I know what it is like to feel amused, after all; and I can imagine someone reciting the maple leaf comment as if it were a joke. Why can't I just combine the two? Why can't I imagine being amused by the maple leaf comment by simply adding imaginary feelings of amusement to an imaginary telling of the joke? According to Walton, the difficulty lies in the combination. In order to imagine being amused by the maple leaf comment, it would not be enough for me to

[50] On this last point, see Walton (1994, p. 50).

simply add to the comment unrelated feelings of amusement. In order to imagine being amused by it, I would need to imagine *a way* in which *it* could be amusing, and to then imagine being amused by it *in that way* (Walton, 1994; Weatherson, 2004). The reason that I can imagine being amused by the knock-knock joke, but not the maple leaf comment, is that whereas I can imagine a way in which the former could be found amusing, I can do no such thing for the latter.

This is the missing element in Smith's account. Though our imaginations are flexible, they are not infinitely so. Our capacity to sympathize with others is constrained by more than the effects of distance and divided attention; it is constrained by our *actual* desires, beliefs, attitudes, principles, and other personal characteristics. Though the number of things that I can imagine finding amusing, for instance, far exceeds the number of things that I actually find amusing, my capacity to imagine finding things amusing is nonetheless limited by my actual sense of humor. It must be limited in this way. For, as we have just seen, in order to imagine finding something amusing, I need to imagine a way in which it could be amusing; but I can do this in no other way than by exploring the boundaries of my own capacity to be amused. It follows that my capacity to sympathize with another person's amusement will ultimately depend, at least in part, on *my own* sense of amusement. And the same is true for other evaluative properties as well. In order to sympathize with a person's resentment, for instance, I would, at the very least, need to be able to imagine a way in which the object of their resentment could be resented. My judgments regarding the propriety of resentment will be grounded in my own disposition to resent. Try as we may to "quarantine"[51] our personal characteristics – i.e., to run our simulations as if we were

[51] On this point, see Goldman (2006, p. 29).

just like another person – some of our own characteristics will inevitably find their way in, placing constraints on what we can and cannot actively imagine.[52]

5. Conclusions

As we learned in Chapter 1, Hare argues that critical moral thought requires us to "fully represent to ourselves" other people's situations, including the fact that they have the particular desires, preferences, etc., that they have. And Hare insists that, when done correctly, such "full representation" necessarily leads to the acquisition of a corresponding set of conditional desires, preferences, and so forth (1981, p. 99; 1988, p. 216). Indeed, Hare builds this into his definition of full representation, insisting that the acquisition of an appropriate set of condition attitudes is "a logically necessary condition of full representation" (1988, p. 216). Of course, even if critical moral thought requires us to do this, it does not follow that we will do it. As Hare himself notes, full representation can be difficult, perhaps even impossible in some cases. But he nevertheless insists that critical thought requires it. It is what we would do if we had "superhuman powers of thought, superhuman knowledge, and no human weaknesses" (1981, p. 44).

My interpretation of Smith's account of imaginative simulation is similar to Hare's in some respects. I have argued, for instance, that Smith thinks that when we imagine ourselves in other people's shoes, we take it upon ourselves to assimilate their characteristics and simulate a response to their circumstances as if we were just like them. And I have emphasized Smith's claim that doing so can be difficult. So far, Smith and Hare are in agreement. But whereas Hare insists that the difficulty of full representation is "one of the main obstacles to good moral thinking" (1988, p. 217), Smith views our occasional failure to mirror other people's mental

[52] My interpretation of Smith's account overlaps somewhat with Julia Driver's more general discussion of the role of "imaginative resistance" in sentimentalist moral theory (Driver, 2008).

states as an essential and highly beneficial feature of our moral psychology. For, according to Smith, when we imagine ourselves in other people's shoes, we do so not to identify or experience their actual feelings, nor to incorporate their actual feelings into our moral deliberations, but rather to evaluate the propriety of their conduct. Far from presenting an obstacle to good moral thinking, I have argued that it is our limited ability to achieve Hare's goal that makes such an evaluative process possible.

We evaluate the conduct of others, according to Smith, by assimilating their characteristics and attempting to simulate a response to their circumstances as if we were just like them. By proceeding in this way – trying our best to see things as other people see them – we maximize our sensitivity to different points of view. At the same time, though, we retain our critical distance in virtue of several features of the imaginative process. Our divided sympathy leads us to imagine not only the perspective of the person under evaluation, but those of all affected parties; and the moderating effect of distance renders us less susceptible to the effects of "hot" emotions, thereby facilitating a more calm and thoughtful evaluative process. By viewing other people's circumstances and conduct from a spectator's perspective, Smith thus thinks that we are naturally triggered to see more broadly and feel with more circumspection. I have argued, moreover, that Smith is committed to an additional claim: that our critical perspective is further enhanced by our inability to imagine certain things as instantiating (or not instantiating) certain evaluative properties. According to Smith, our occasional failure to mirror other people's mental states may, at least in some cases, be the result not of any error in our moral reasoning, but of our sensitivity to evaluative features of the world.[53]

[53] Karsten Stueber makes a similar claim (Stueber, 2011, p. 178).

My interpretation satisfies each of the conditions outlined in Section 1. It yields an evaluative perspective that is simultaneously critical and respectful of other people's points of view – the former because it allows for the possibility of discordance and disapproval via the mechanisms just reviewed, and the latter in virtue of its reliance on the framework of empathetic simulation. Moreover, it both elucidates Smith's claim that our judgments of propriety are grounded in our own affective tendencies, and reconciles his assertion that we attempt to sympathize by changing persons and characters with his seemingly contradictory claim that we sometimes disapprove of other people when we find that we ourselves would have felt or acted differently. According to my reading, we change persons and characters insofar as we simulate empathetically, but, as argued in Section 4, our imaginative simulations are nevertheless constrained by our actual desires, beliefs, attitudes, principles, and so forth. We attempt to simulate as if we were just like the other person, but some of our own characteristics inevitably find their way in.

CHAPTER 4. SELF-EVALUATION AND CULTURAL RELATIVISM

1. The Impartial Spectator

Up to this point, I have focused on Smith's account of our judgments concerning *other people's* feelings and behavior. In doing so, I have followed Smith himself, who devotes the first two parts of *TMS* almost exclusively to our evaluations of others. In Part III, however, Smith turns his attention to our judgments concerning *our own* feelings and behavior. In this chapter and the next, I shall do the same.

The two forms of evaluation present different, but related, challenges. When evaluating other people, we are often quick to condemn. We tend to interpret their conduct uncharitably, to attribute any defect in their behavior to a corresponding defect in their character, rather than acknowledging the features of their circumstances that might explain, excuse, or even justify what they have done.[54] The primary challenge is one of empathy and understanding: of fully appreciating the circumstances and perspectives of those we judge. In contrast, when evaluating ourselves, we are often quick to acquit. Our judgments typically suffer from too much empathy and understanding, rather than too little; and we rarely encounter trouble finding ways to excuse or justify our conduct, even when it is far from justifiable. Though *other people* who rely on social assistance programs are lazy, *we* have merely been unfortunate; though rule-breaking in

[54] The tendency of people to overstate the role played by other people's personal characteristics, relative to environmental influences, when explaining their actions is referred to as the *Fundamental Attribution Error* or *Correspondence Bias* (see Aronson, 2011, p. 164).

others is a clear sign of malice, *our own* violations are always justified by great need.[55] On the surface, the challenges posed by the two forms of evaluation – to be more indulgent in the one case, more critical in the other – are opposites. In each case, though, the underlying difficulty is the same. We fail to evaluate other people fairly when we fail to appreciate their point of view, and we fail to evaluate ourselves critically when we become lost in our own. In each case, successful evaluation requires the inhabitation of a perspective *other than* our own.

Smith is aware of this. From the beginning, he emphasizes the importance of acquiring adequate distance from ourselves whenever we attempt to evaluate ourselves critically. "We can never survey our own sentiments and motives," he writes, "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" (*TMS* III.1.2). His solution:

> 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. (*TMS* III.1.2)

This, in rough outline, is Smith's famous theory of the *impartial spectator*. We engage in critical self-evaluation, Smith thinks, by viewing ourselves not from our own point of view, but from the point of view of an observer – an impartial spectator. We judge our feelings and behavior to be proper insofar as we conclude that an impartial spectator could sympathize with them, and we judge them to be improper insofar as we do not.

[55] The tendency of individuals to attribute their own actions to situational factors while at the same time attributing the actions of others to features of their personalities is referred to as the *Actor-Observer Bias* (see Aronson, 2011, p. 168).

A spectator theory solves the problem of critical distance, but it raises several questions of its own. What is it to view one's conduct from the perspective of a fair and impartial spectator? For that matter, what *is* a fair and impartial spectator? And what reason do we have to view ourselves from such a point of view?

Ideal observer theory provides one way of filling in the details. According to ideal observer theory, the perspective we take on things when engaging in moral deliberation can be modeled as that of a *suitably idealized* observer. In Roderick Firth's archetypal theory, for instance, we approve of something as morally right if and only if it would gain the approval of an observer who is omniscient, omnipercipient, disinterested, dispassionate, and perfectly consistent (Firth, 1952). Is Smith's theory like Firth's? Is the "impartial spectator" an "ideal observer"? Many people have interpreted him in this way. Firth himself mentions Smith as one of the "classical moralists" who, he thinks, "may have" proposed and defended an ideal observer theory similar to his own (1952, p. 318), and John Rawls begins his review of ideal observer theory with Smith and includes Smith's work (along with Firth's) in his list of paradigmatic accounts (1971, p. 161). John Harsanyi claims that his own "equiprobability model" of moral evaluation (which, as he notes, is structurally equivalent to an ideal observer theory in which the observer is both strictly impartial and maximally empathetic) is merely "a modern restatement of Adam Smith's theory of an impartially sympathetic observer" (1982, p. 46). And, though less explicit than Harsanyi, Brian Barry appears to read Smith in roughly the same way (see Barry, 1995, p. 59). Among non-specialists, at least, Smith is quite often viewed as an early pioneer of modern ideal observer theory.

But this interpretation is certainly mistaken.[56] Granted, Smith's impartial spectator is informed and disinterested, but that is all. There is nothing in Smith's work to suggest the sort of idealized observer found in the work of Firth or Harsanyi. Smith's spectator is neither omniscient nor dispassionate, like Firth's observer, nor is it what Harsanyi calls "impartially sympathetic", by which he means something like "bound to experience each person's feelings with perfect fidelity." Indeed, Smith's impartial spectator *couldn't possibly* be "impartially sympathetic" in this sense. For, as we have seen, Smith's account of sympathy is evaluative: to sympathize, he claims, *is* to approve.[57] It is therefore essential that the impartial spectator be capable of *not* sympathizing with at least some of the people some of the time, lest everything be approved of in each and every case. If Smith's account is to succeed, in other words, it is essential that the impartial spectator *not* be ideal in this sense. The ideal observer interpretation of Smith's account is a non-starter.

A more common interpretation among Smith scholars is that Smith's account of critical self-evaluation is really an account of the internalization, on the part of individuals, of existing social norms. According to James Farrer, for example, Smith's notion of self-condemnation is just "the condemnation of public opinion, with which we identify ourselves by long force of habit" (1881, p. 185). All references to Smith's "abstract spectator," Farrer claims, are in fact references to the "verdict of public opinion" (p. 184). T.D. Campbell offers a similar view. He describes the judgments of Smith's impartial spectator as manifestations of "the normal reactions of the ordinary person when he is in the position of observing other people's behaviour" (1971, p. 127). A person learns to evaluate himself critically, according to Campbell's reading, by "internalizing the judgments which he knows, or imagines, that other people pass on his conduct" (p. 147).

[56] For helpful criticisms of the ideal observer interpretation of Smith, see Campbell (1971) and Raphael (2007).
[57] Actually, Smith is not entirely consistent on this point. It is clear, though, that Smith thinks sympathy a necessary and sufficient condition of approval. See Chapter 2 for a discussion of this point.

Smith's impartial spectator is, on this view, nothing but a convenient theoretical representation of "the average or normal or ordinary man" (p. 134). Samuel Fleischacker offers a more complex reading, with two distinct and, as he sees it, conflicting themes (to which we shall return shortly). But one branch of his reading is consistent with Farrer's and Campbell's. Smith's "impartial spectator," Fleischacker writes, "is constructed out of modes of judgment that seem essentially relative to a particular culture. We observe what our friends and neighbors say about how people should behave and we try to win their approval" (2011a, p. 28). According to this interpretation, Smith's moral psychology leads directly to a form of *cultural relativism*. I shall call this the *relativistic reading*.

Advocates of this reading do not claim that Smith conceives of our self-evaluative judgments as reflections of the *actual* assessments made of us by others. This would clearly be a mistake, as Smith repeatedly emphasizes the inadequacy of approval based on "ignorance or mistake," noting that if "we are conscious that we do not deserve to be so favourably thought of, […] our satisfaction is far from being complete" (*TMS* III.2.4). And just as he denies that mistaken approbation can (fully) please us, he insists that when we conduct ourselves with propriety, we can reflect on our behavior with approval, relatively unaffected by whether others approve of us as well. It "often gives real comfort," he writes, "to reflect, that tho' no praise should actually be bestowed upon us, our conduct, however, has been such as to deserve it" (*TMS* III.2.5). We are pleased "not only with praise, but with having done what is praise-worthy," and by the fact that "we have rendered ourselves the natural objects of approbation, though no approbation should ever actually be bestowed upon us." Conversely, when we behave poorly, "we are mortified to reflect that we have justly incurred the blame of those we live with, though the sentiment should never actually be exerted against us." We regard ourselves, Smith writes, "not so much

according to the light in which [others] actually regard [us], as according to that, in which they would regard [us] if they were better informed."[58]

The passages in the previous paragraph all appear in the first edition of *TMS*. In the second edition, Smith develops the thought even further, drawing attention to the relevance not only of our spectators' knowledge, but of the effects that our conduct might have on them. "When we first come into the world," he writes, our "natural desire to please" motivates us "to consider what behaviour is likely to be agreeable to every person we converse with" (*TMS*, Ed. 2, 207). We initially aim to gain the "good-will and approbation of every body." Quickly, though, we learn that this is impossible: "by pleasing one man, we almost certainly disoblige another" and "by humouring an individual, we may often irritate a whole people." To "defend ourselves from such partial judgments," we "set up in our own minds a judge between ourselves and those we live with."

> We conceive ourselves as acting in the presence of a person quite candid and equitable, of one who has no particular relation either to ourselves, or to those whose interests are affected by our conduct, who is neither father, nor brother, nor friend either to them or to us, but is merely a man in general, an impartial spectator who considers our conduct with the same indifference with which we regard that of other people. (*TMS*, Ed. 2, 207-208)

Advocates of the relativistic reading are sensitive to these features of Smith's view. Their claim is not that Smith thinks that we regard ourselves according to the light in which others regard us, but that we regard ourselves according to the light in which others *would* regard us *if* they were informed and impartial. We evaluate ourselves by considering "not what the world approves or disapproves of, but what appears to [the] impartial spectator, the natural and proper object of approbation or disapprobation" (*TMS*, Ed. 2, 209).

[58] See also Smith's discussion of Mandeville (*TMS*, Ed. 1, 474-486) and his reply to Sir Gilbert Elliot's early objection to his theory (Mossner & Ross, 1977, pp. 48-50).

The relativist reading of Smith's view is far more plausible than the ideal observer reading, but there are nevertheless four reasons to resist it. First, by depicting Smithian self-evaluation as a product of the passive internalization of dominant social norms, the interpretation precludes criticism of the norms themselves. It implies that Smith thinks people capable only of applying their society's evaluative standards, not of questioning, revising, or rejecting them – let alone devising their own.[59] But this is implausible. To begin, it is obvious that we *can* criticize our own society's dominant social norms; we do it all the time, as does Smith.[60] Moreover, by equating self-evaluation with the passive internalization of social norms, the interpretation implies that Smith conceives of the process as something that *happens to* individuals, rather than something that they *do*. But self-evaluation is an active process, involving not just the passive application of fixed evaluative standards, but an active engagement with the standards themselves. The view entailed by the relativistic interpretation is implausible because it leaves no room for such active engagement to occur.[61]

[59] Campbell's interpretation is a case in point. He writes that Smith's impartial spectator is "the product of a particular society" (p. 144), and that "talk of *the* impartial spectator is simply a short-hand way of referring to the normal reaction of a member of a particular social group, or of a whole society, when he is in the position of observing the conduct of his fellows" (p. 145, emphasis in original). An agent's authority to dismiss an ignorant assessment, Campbell claims, is "no more nor less than an appeal to the attitudes which the real spectators would adopt did they have the information which is available to the agent" (p. 155). Though he does not entirely endorse the interpretation, Fleischacker makes similar claims. He writes that the "impartial judge within us cannot defend us against [ignorant or impartial judgments] unless it uses the same *standards* of judgment that they do" (2011a, p. 28, emphasis in original). Smith's account, he says, "only allows us to respond to distortions by turning to an idealized version of our friends and neighbors who uses the same standards of moral judgment as they do […] There is little in Smith's construction of the idealized spectator to correct for the surrounding society's standards of judgment; the idealized figure takes over those standards and corrects merely for their partial or ill-informed *use*."

[60] Consider, for instance, Smith's critique of the prevailing attitudes regarding poverty during his lifetime. At a time when many people believed that people poor people were poor in virtue of their own inferiority – indeed, that the hierarchical, deeply inegalitarian, social structure prevailing at the time was a direct reflection of God's will – Smith denied the existence of significant personal differences between the rich and the poor, thus paving the way for highly unconventional arguments on behalf of the least well-off members of society. See Fleischacker (2004b, pp. 62-68) and Baugh (1983).

[61] Of course, that the view entailed by the relativistic reading is implausible does not necessarily mean that the reading is incorrect. Smith could have simply held an implausible view. All else equal, though, it would be better to attribute a plausible view to him than an implausible one, and the implausibility of the view suggested by the relativistic reading thus constitutes a major strike against it. Moreover, Smith's description of individuals whose self-evaluative judgments depend entirely on actual public opinion as "slaves of the world" (*TMS*, Ed. 2, 209)

Second, the relativistic reading fails to capture an important egalitarian strain embedded in Smith's work. As Stephen Darwall observes, by claiming that we evaluate other people's conduct by attempting to imagine their circumstances from *their* perspectives (see Chapters 2 and 3), Smith implies that our judgments of propriety necessarily involve "an implicit identification with, and thus respect for, [each person] as having an independent point of view" (2004, p. 132).[62] Right from the beginning, then, Smith's account presupposes "a moral community among independent equal persons." Moreover, Darwall argues that Smith's theory of justice goes even further, extending to each individual not just membership in a community of independent moral equals, but "implicit standing as one among *mutually accountable* equals" (p. 133, emphasis added). When we are intentionally harmed by another person, what "chiefly enrages" us, Smith notes, is not that we have been harmed, but that the prospect of harming us has received so little weight in our adversary's deliberations (*TMS* II.iii.1.5). We are enraged by

> the little account which he seems to make of us, the unreasonable preference which he gives to himself above us, and that absurd self-love, by which he seems to imagine, that other people may be sacrificed at any time, to his conveniency or his humour. (*TMS* II.iii.1.5)

In short, we are enraged when someone fails to respect us as equals, and we respond by holding him accountable – by doing whatever we can to "make him sensible of what he owes us, and of the wrong that he has done to us" (*TMS* II.iii.1.5). The framework of equality, independence, and mutual accountability, implicit in Smith's moral psychology, grounds a "normative doctrine of equal dignity" (Darwall, 2004, p. 132) that pervades his work (including his economics[63]) and provides the foundation for much of his critical social commentary. The relativistic reading fails

---

suggests that he too would have objected to an overly passive depiction of our capacity for critical self-evaluation. Though some people may evaluate themselves passively, the ideal moral agent does not.

[62] See also the discussion in Darwall (1999).

[63] For a philosophical examination of Smith's economics, see Fleischacker (2004a).

to capture this. By depicting our self-evaluative judgments as products of internalized social norms, the interpretation emphasizes only the importance of respect for existing social standards, not for people as equal, independent, and mutually accountable members of a moral community. It exhibits no sign of the egalitarian normative principles that drive Smith's work.[64]

The third problem is related to the second. In addition to embedding a form of egalitarianism into his moral psychology, Smith thinks that the process of self-evaluation can *cause* us to have certain egalitarian thoughts or attitudes. Though moderately selfish by nature, Smith argues that by viewing ourselves from the perspective of the impartial spectator, we will naturally come to see other people as equally important. Indeed, by the time of the second edition of *TMS*, Smith had come to think of this as the *only* way to "make any proper comparison between our own interests and those of other people" (*TMS* III.3.1").[65]

> [T]o 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 […] than the greatest concern of another with whom we have no particular connexion. His interests, as long as they are surveyed from this station, can never be put into the balance with our own, can never restrain us from doing whatever may tend to promote our own, how ruinous soever to him. 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 with our own eyes, nor yet with his, but from the place, and with the eyes of a third person, who has no particular connection with either, and who judges with impartiality between us. (*TMS* III.3.3)

By viewing ourselves in this manner, we "learn the real littleness of ourselves, and of whatever relates to ourselves"; we begin to see ourselves as "but one of the multitude, in no respect better than any other" (*TMS* III.3.4). It is the impartial spectator, Smith writes, "who shows us the propriety of generosity and the deformity of injustice; the propriety of resigning the greatest

[64] On this problem, see also Fleischacker (2011a).
[65] See also, Fleischacker (2011a, pp. 32-35).

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." The third problem with the relativistic reading is that it is incapable of explaining these remarks – of accounting for the psychological effects of engaging in the self-evaluative process.[66]

Finally, the relativistic reading fails to leave room in the self-evaluative process for Smith's distinctive notion of sympathy. This is a serious problem, as Smith makes clear that he thinks sympathy plays a central role. The "principle by which we naturally either approve or disapprove of our own conduct" is, he claims, "*altogether the same* with that by which we exercise the like judgments concerning the conduct of other people" (*TMS* III.1.2, emphasis added).

> We either approve or disapprove of the conduct of another man according as we feel that, when we bring his case home to ourselves, we either can or cannot entirely sympathize with the sentiments and motives which directed it. And, in the same manner, we either approve or disapprove of our own conduct, according as we feel that, when we place ourselves in the situation of another man, and view it, as it were, with his eyes and from his station, we either can or cannot entirely enter into and sympathize with the sentiments and motives which influenced it. (*TMS* III.1.2)

We approve of ourselves "by sympathy with the approbation" of our "supposed equitable" judges, and we disapprove of ourselves by sympathy with their disapprobation (*TMS* III.1.2). By insisting that Smith thinks we self-evaluate by internalizing the judgments that we know from experience other people would, if informed and impartial, make of us, the relativistic reading implicitly attributes to Smith something like Hume's account of sympathy, according to which we sympathize with others by, roughly speaking, feeling whatever they feel. But, as we know, Smith's account is not like Hume's. Smith's account is evaluative: to sympathize with the

[66] Of course, the psychological mechanism embedded within the relativistic reading *could* produce such results, but it would not *necessarily* do so. The outcome would depend on the characteristics of the society in which one lived. The problem is not that the relativistic reading entails that self-evaluation via the impartial spectator model cannot produce egalitarian sentiments, but that it fails to support Smith's allegation of a direct causal link between his moral psychology and his egalitarian ideals.

approbation or disapprobation of our spectators is, in his view, to judge their approbation or disapprobation proper. By ignoring – or, more accurately, mischaracterizing – the role of Smith's distinctive notion of sympathy, the relativistic reading omits the central (and, I think, most interesting) feature of Smith's work.

To summarize, Smith's account of evaluative judgment presupposes a normative conception of society as a community of independent and mutually accountable equals. This framework of equality, independence, and accountability grounds a normative doctrine of equal dignity, implicit in his moral psychology, which regulates our judgments. When we engage in critical self-evaluation, we do so, according to Smith, by stepping back not only from ourselves, but from our society. We evaluate ourselves and our society in light of the normative conception of social cooperation to which we are implicitly committed. The role of the impartial spectator model is to fill in the psychological details of such a story: to provide an empirically defensible sympathy-based developmental account of how, over time, we can come to freely endorse a normative conception of society as a community of independent and mutually accountable equals, and then both learn and acquire the appropriate motivation to evaluate ourselves (and others) in a way that reflects this conception. The problem with the relativistic reading of Smith's account is that it fails to identify a psychological mechanism capable of filling this role.

Of course, for all I have said, the relativist reading could still be right. Smith may have simply failed to provide an account of sympathy that is capable of doing what he needs it to do. This is Fleischacker's view. Though Fleischacker agrees that Smith endorses egalitarian (even cosmopolitan[67]) principles, he argues that Smith's "system provides us with no good explanation of how anyone can come to have [the corresponding] sentiments" (2011a, p. 32). He denies that

[67] See, for example, *TMS* III.3.3-7, VI.ii.3, and Fleischacker (2011a, p. 32).

Smith's moral psychology is capable of providing people with the tools needed to criticize their own societies on the basis of egalitarian ideals:

> Simply knowing that all human beings are equal is […] not enough: we need also to know what sort of treatment befits these equal beings. And for that, Smith will presumably tell us to ask what sort of treatment we expect to be approved by the impartial spectator. But if the impartial spectator, again, operates on the fundamental standards upheld in the society around it, we can hardly expect it to judge those standards themselves as improper. So the impartial spectator is unlikely to tell us, since it is unlikely to be aware, that a kind of treatment prescribed for all human beings in our society, should be condemned. (Fleischacker, 2011a, p. 36)

Fleischacker agrees, in other words, that the relativistic reading fails to identify a psychological mechanism capable of supporting Smith's egalitarian ideals, but he denies that that this provides any reason to reject it. As Fleischacker sees it, Smith's failure to provide a psychological mechanism capable of supporting his own egalitarian ideals is just an unfortunate feature of his theory. Smith's moral psychology simply fails to deliver.

I disagree, and in the next chapter I present and defend an alternative (non-relativistic) interpretation of Smith's impartial spectator model that, I claim, goes a long way (if not quite all the way) toward bridging the gap between his moral psychology and his egalitarian principles. In the remainder of this chapter I take a closer look at the relativistic reading. In Sections 2 and 3, I examine two different sorts of passages in Smith's work that appear to support such a reading. In each case, I argue that, properly understood, the passages provide no support for the relativistic reading at all – indeed, that they point away from such an interpretation of Smith's view. In the end, I conclude not only that there are good reasons to resist the relativistic reading of Smith's account (as already discussed), but that there are no good reasons to endorse it. Like the ideal observer reading, the relativistic reading is a non-starter.

2. Smith's "Secret Reference" to the Judgments of Others

In his initial formulation of the impartial spectator model, Smith writes that our judgments concerning ourselves "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 judgments of others*" (*TMS* III.1.2, emphasis added). He claims, in other words, that our own self-evaluative judgments necessarily refer to other people's judgments. This claim is not unique. Indeed, in the first five editions of *TMS*, the chapter containing the quote was entitled "In what manner our own judgments refer to what ought to be the judgments of others".[68] And references to the sentiments or judgments of others appear elsewhere in Smith's work as well. He later writes that virtue, for instance, is meritorious not "because it is the object of its own love or of its own gratitude, but because it excites those sentiments *in other men*" (*TMS* III.1.7, emphasis added). Smith thus builds the sentiments of others directly into his concept of virtue: to possess the capacity to excite the love or gratitude of others is, for Smith, what virtue *is*, and analogously for vice. When we judge ourselves virtuous or vicious, Smith thinks that we do so not because our character or action causes *us* to love or hate ourselves (as Hume argued[69]), but in virtue of the relation between our character or action and *other people's* sentiments.

Of course, Smith does not think that our judgments necessarily refer to other people's *actual* sentiments or judgments. We already know that Smith allows for corrections of ignorance and partiality (see Chapter 3). Indeed, he even leaves room for such corrections in the passage just

[68] In the sixth edition, Smith changed the title of the chapter to "Of the Principle of Self-Approbation and Self-Disapprobation".

[69] Hume writes, for example: "since every quality in ourselves or others, which gives pleasure, always causes pride or love; as every one that produces uneasiness, excites humility or hatred: It follows, that these two particulars are to be consider'd as equivalents, with regard to our mental qualities, *virtue* and the power of producing love or pride, *vice* and the power of producing humility or hatred. In every case, therefore, we must judge of the one by the other; and may pronounce any *quality* of the mind virtuous, which causes love or pride; and any one vicious, which causes hatred or humility" (*Treatise* 3.3.1.3).

quoted, acknowledging that our judgments can refer not only to the *actual* judgments of others, but to what their sentiments *would* be "upon a certain condition," by which he undoubtedly means being informed and impartial. The question is: does Smith think that the referent of our self-evaluative judgments can extend any further? Can our self-evaluative judgments refer to anything *other than* the actual or hypothetical (i.e., informed and impartial) judgments of our peers? If not, then Smith's "secret reference" claim would provide strong evidence for a relativistic reading of his account. For if our self-evaluative judgments could *only* refer either to what are or to what would be the judgments of our peers, then our judgments would indeed be restricted, as Fleischacker says, to our peers' "reactive attitudes" or "modes of moral judgments" (2011a, p. 28). Our judgments would be constrained by the evaluative standards of those with whom we live.

According to Campbell (1971), this is exactly what Smith thinks: that our self-evaluative judgments necessarily refer to what are or to what would be the judgments of our peers, if only our peers were informed and impartial; and nothing more. The problem with this interpretation, though, is that it conflicts with what Smith actually writes. Smith simply does not state that our self-evaluative judgments must refer to the actual or hypothetical judgments of our peers. Rather, he writes that our judgments can refer to any of *three* things: (i) to the judgments of others, (ii) to what upon a certain condition (full information and partiality) would be the judgments of others, or (iii) to *what we imagine ought to be* the judgments of others. If Campbell's interpretation is to succeed, he must explain away Smith's third disjunct. He must explain, in other words, why Smith's claim that our self-evaluative judgments can refer to *what we imagine ought to be* the judgments of others does not free us from the constraints of our society's dominant evaluative standards.

To his credit, Campbell does attempt to offer such an explanation. He writes that when Smith says that "we sometimes judge ourselves by 'what ought' to be the judgments of others," all he really means is that we sometimes judge ourselves by the attitudes that our "spectators would adopt did they have the information which is available to the agent" (1971, p. 155). In other words, Campbell's claim is that, properly understood, Smith's third disjunct is identical to his second, and the seemingly three-part formulation of the impartial spectator model is really just the two-part formulation that I have already acknowledged would support a relativistic interpretation of Smith's account.

But Campbell's claim is implausible for two reasons. The first reason is obvious: if, when Smith wrote that our self-evaluations sometimes refer to "what we imagine ought to be" other people's judgments, all he meant to say was that our self-evaluations can sometimes refer to "what upon a certain condition would be" other people's judgments, there would have been no reason for him to include the third disjunct. The first two would have said everything that needed to be said. The fact that Smith, by all accounts an extraordinarily careful writer, nevertheless decided to include the third disjunct – and to retain it in *all six* editions of *TMS* – suggests (albeit, inconclusively) that he intended it to mean something other than that which was expressed by the second. At the very least, we ought to assume, at least provisionally, that Smith did not intend the disjunct to be redundant. This could turn out to be false, of course; Smith could have formulated his impartial spectator model redundantly. But this would need to be shown. The burden of proof is on Campbell's side.

Second, and more importantly, in addition to being a simpler and more natural reading, a literal three-part interpretation of Smith's formulation is actually necessitated by his view of the role played by sympathy in the process of critical self-evaluation – a point Smith makes only a

few sentences before the "secret reference" passage, in the very same paragraph, as well as elsewhere in *TMS*. Smith, recall, writes that we

> either approve or disapprove of the conduct of another man according as we feel that, when we bring his case home to ourselves, we either can or cannot entirely sympathize with the sentiments and motives which directed it. And, in the same manner, we either approve or disapprove of our own conduct, according as we feel that, when we place ourselves in the situation of another man, and view it, as it were, with his eyes and from his station, we either can or cannot entirely enter into and sympathize with the sentiments and motives which influenced it. (*TMS* III.1.2)

We evaluate ourselves and others, Smith thinks, by engaging in an evaluative process of imaginative mental simulation. Our simulated responses to other people's circumstances establish a standard of propriety which we then use to determine the "suitableness or unsuitableness" of their sentiments to the "cause or object which excites" them (*TMS* I.i.3.6). We determine how other people *ought* to feel by imagining being in their circumstances and simulating a response; and we judge that their sentiments are as they *ought* to be if and only if we find that we can sympathize with them – i.e., if and only if their actual sentiments correspond to the results of our simulation. When Smith thus writes that we approve of ourselves "by sympathy with the approbation" of our "supposed equitable" judges (*TMS* III.1.2), he is in fact stating that we evaluate ourselves by determining how our spectators *ought* to judge us. He *needs* the third disjunct in order to make his account consistent.

Our spectators' judgments may be as we think they ought to be, or perhaps their judgments would be as we think they ought to be if only they were fully informed and impartial. If the former, then our self-evaluative judgments will refer to our spectators' actual judgments; if the latter, they will refer to our spectators' hypothetical (informed and impartial) judgments. But it may be that neither is the case, and Smith's three-part formulation allows for this possibility by

enabling us to (i) reject even our informed and impartial spectators' judgments and (ii) evaluate ourselves by reference to what we imagine their judgments ought to be – or, what comes to the same thing for Smith, by reference to the sentiments that we have when we imagine ourselves in their circumstances. What Smith means when he claims that our self-evaluative judgments bear a "secret reference" to the sentiments of others is only this: that they necessarily refer to what our spectators *ought* to think of us. On its own, the "secret reference" passage provides no support at all for a relativistic reading of Smith's account of critical self-evaluation.

3. From Praise to Self-Evaluation?

Nothing pleases us more, Smith writes, "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* I.i.2.1). We long for the sympathy of others and are dismayed when we do not receive it. Unfortunately, though, the sympathy of others is not something that is easily attained. The feelings of even the most sensitive spectators, Smith observes, nearly always "fall short of the violence of what is felt" by the object of their sympathetic efforts. Though "naturally sympathetic," people rarely experience that "degree of passion which naturally animates the person principally concerned" (*TMS* I.i.4.7). If we are to have any hope of securing the sympathy of others, then, Smith claims that we must develop our capacity for emotional regulation: we must learn to dampen our feelings "to that pitch, in which [our] spectators are capable of going along with." And this requires perspective-taking:

> as 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* I.i.4.8)

In order to secure our spectators' sympathy, we must assume their circumstances and feel as we imagine they feel when they assume ours. We must view ourselves from another person's point of view and regulate our conduct accordingly.

Strictly speaking, the above comments are about our capacity for self-control, not our capacity for critical self-evaluation. But they suggest a developmental story that would, if endorsed by Smith, support a relativistic reading of his account. The story goes like this. We have a natural desire for the sympathy and approval of those with whom we live, and this desire leads us to (i) place ourselves in our spectators' shoes and (ii) attempt to feel as we imagine they feel when they place themselves in ours. Driven by our desire for sympathy, in other words, we attempt to match our feelings to those of our spectators. Unfortunately, though, we quickly learn that our spectators' feelings are not all the same. They are determined in large part by their beliefs about us and relations to us, which vary from spectator to spectator. We cannot match all of them simultaneously. So we do the next best thing: correcting for ignorance and partiality, we regulate our feelings by what we think our spectators *would* feel *if* they were fully informed and impartial. Driven by our desire for the sympathy and approval of others, in other words, we internalize the sympathetic tendencies of our peers and evaluate our feelings and behavior accordingly.

This story depicts our capacity for self-evaluation as an offshoot of a developmentally prior capacity for sympathy-based emotional self-regulation, and it depicts our particular self-evaluative judgments as internalizations of refined public opinion – unaffected by our spectators' ignorant and/or partial judgments, but nevertheless grounded in, and ultimately constrained by, our desire for their sympathy and approval. If Smith in fact endorsed this sort of developmental story – if, as Fleischacker claims, he viewed the self-evaluative process as a mere "outgrowth of

each person's search for harmony with the feelings of the people around him or her" (2011a, p. 27) – then he would indeed by committed to a form of relativism.

Several philosophers have read Smith in this way. After claiming that Smith grounds all evaluative judgment in our desire for the sympathy of others, for instance, Gibbard concludes that "if Smith's pragmatic story supports his detached observer theory, it supports the theory *in a relativized form*," as it implies that the "proper feelings for a person […] are those of a detached observer *who belongs to that person's own culture*" (1990, p. 280, emphasis added). If this developmental story is Smith's, then he must think that we learn to self-evaluate by calibrating our judgments by the evaluative standards of those with whom we live. The relativistic reading must be right.

But is this really Smith's view? Would he endorse the developmental story sketched above? Early editions of *TMS* contain some evidence that he would. As we have seen, for instance, Smith claims in the second edition that we are driven by a "natural desire to please" to "consider what behaviour is likely to be agreeable to every person we converse with" and then regulate our conduct accordingly (*TMS*, Ed. 2, 207). And though he emphasizes our authority to reject ignorant or partial assessments of our conduct, Smith initially (i.e., in the first five editions) grounds our authority to do so in the independent evaluative authority of those whose opinion we overrule. Though the "tribunal within the breast" is, as he puts it, the "supreme arbiter of all our actions,"

> though it can reverse the decisions of all mankind with regard to our character and conduct […] yet, if we enquire into the origin of its institution, its jurisdiction we shall find is in a great measure *derived from the authority of that very tribunal, whose decisions it so often and so justly reverses*. (*TMS*, Ed. 2, 206-207, emphasis added)

In short, these passages suggest that Smith's view in the early editions of *TMS* was indeed that we are driven to self-evaluate by our desire to please, and that the authority of our own self-evaluative judgments derives entirely from that of our peers. This is very much in line with Campbell's reading of Smith's view – in particular, his claim that, for Smith, an agent's authority to dismiss an ignorant assessment is "no more nor less than an appeal to the attitudes which the real spectators would adopt did they have the information which is available to the agent" (1971, p. 55). The passages provide clear support for the relativistic reading.

Other passages in *TMS* appear to express similar ideas. Consider, for instance, Smith's denial of the possibility of asocial self-evaluation. Smith writes that if a human being were ever to "grow up" in "some solitary place, without communication with his own species," he could "no more think of his own character, of the propriety or demerit of his own sentiments and conduct, of the beauty or deformity of his own mind, than of the beauty or deformity of his own face" (*TMS* III.1.3).

> All these are objects which he cannot easily see, which naturally he does not look at, and with regard to which he is provided with no mirror which can present them to his view. Bring him into society, and he is immediately provided with the mirror which he wanted before. It is placed in the countenance and behaviour of those he lives with, which always mark when they enter into, and when they disapprove of his sentiments; *and it is here that he first views the propriety and impropriety of his own passions, the beauty and deformity of his own mind.* (*TMS* III.1.3, emphasis added)

Deprived of all social interaction, Smith believes that a person would restrict his attention to "the objects of his passions," that his passions themselves would "scarce ever be the objects of his thoughts." Bring such a person into society, however, and his perspective will change. Finding that other people approve of some of his actions, and are disgusted by others, he will become "elevated in the one case, and cast down in the other." His desires and aversions, joys and

sorrows, will suddenly "become the causes of new desires and new aversions, new joys and new sorrows." His characteristics and conduct, to which he had previously given so little thought, will become the object of a new class of self-regarding sentiments. In short, his exposure to the criticism of others will enable him to self-evaluate. At least at first glance, this reads like a clear endorsement of the sort of development story sketched above, as it suggests that Smith thinks we can only engage in self-evaluation by (i) observing our peers' assessments, (ii) correcting them for errors of ignorance or partiality, and then (iii) internalizing the results.

That Smith endorses such a story is, I think, the strongest argument that can be made in favor of a relativistic reading of his view; and, as the passages just quoted suggest, there is reason to think that he may have initially had something like it in mind. Even in the early editions of *TMS*, however, the picture is not as clear as the above passages make it out to be; there are indications of a different view as well. Consider, for instance, the paragraphs immediately following Smith's rejection of the possibility of asocial critical self-evaluation. In these paragraphs, Smith sketches a developmental account of our capacity for self-evaluation that begins not with our desire for *other people's sympathy*, but with *our assessments* of them. Our "first moral criticisms," he writes, "are exercised upon the characters and conduct of other people" (*TMS* III.1.5). It is only afterward – only upon finding that other people are "equally frank" in their assessments of us as we are in our assessments of them – that we turn our gaze inward. We "become anxious to know how far we deserve their censure or applause, and whether to them we must necessarily appear those agreeable or disagreeable creatures which they represent us." We thus take it upon ourselves 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. This is the only looking-glass

> by which we can, in some measure, with the eyes of others, scrutinize the propriety of our own conduct. (*TMS* III.1.5)

Smith does not write that, after observing our peers' frankness, we become anxious to know their conclusions, or even to know what they would conclude if they were informed and impartial; nor does he claim that our anxiety leads us to measure our conduct against our society's standards. Rather, his claim is that we respond *by examining our conduct for ourselves*. We imagine *ourselves* the "spectators of our own behavior" and "endeavour to imagine" how it would appear *to us* from such a perspective. What matters, Smith appears to be saying in this paragraph, is not whether *our spectators* would approve of our conduct if they were informed and impartial, but whether *we* would approve of our conduct if *we* were in our informed and impartial spectators' shoes.[70]

The evidence in the first five editions of *TMS* is thus mixed – perhaps because Smith himself was unsure of his own position. By the time of the sixth and final edition of *TMS*, however, it is clear that Smith had come to reject the development story sketched above.[71] The evidence for this claim is substantial. To begin, note that although Smith continues to acknowledge that we are "pleased" when other people approve of our conduct, and "hurt when they disapprove" of it (*TMS* III.2.31),[72] he no longer claims that it is our "natural desire to please" which leads us to "consider what behaviour is likely to be agreeable to every person we converse with." And, similarly, while he continues to acknowledge that nature has "endowed" us with a "desire of being approved of," he places far greater weight in the sixth edition on a different desire: that "of

[70] D.D. Raphael writes: "The 'supposed impartial spectator', as Smith often called him, is not the actual bystander who may express approval or disapproval of my conduct. He 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" (1975, p. 90).
[71] The sixth edition of *TMS* was the last to be published during Smith's lifetime, though additional editions were published posthumously.
[72] He writes, in addition, that we "respect the sentiments and judgments" of our "brethren" (*TMS* III.2.31) and "feel pleasure in their favourable, and pain in their unfavourable regard" (*TMS* III.2.6).

being *what ought to be* approved of; *or of being what he himself approves of in other men*" (*TMS* III.2.31, III.2.6-7, emphasis added). What we see in the sixth edition, in other words, is both a marked decline in the idea that our self-evaluative judgments derive from the sentiments of others, and an increased emphasis on the claim that we evaluate ourselves by considering how *we* would view ourselves, if only *we* were informed and impartial.

Other revisions make the case even more clearly. In what appears to be a direct repudiation of the sort of developmental story sketched above, Smith writes in the sixth edition that the "love of praise-worthiness is by no means derived altogether from the love of praise" (*TMS* III.2.2).

> 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*.III.2.2)

Instead of growing from our love for praise, he claims that our desire to be praise-worthy stems from the "love and admiration which we naturally conceive for those whose character and conduct we approve of" (*TMS* III.2.3). We want to be like those whom we love. And, in a striking reversal of his earlier position, Smith explicitly rejects the claim (present in early editions, but not in the sixth) that the "jurisdiction" of our self-evaluative judgments derives from that of our peers. After acknowledging that we may at any time appeal from the verdict of public opinion to the "much higher tribunal" of our own conscience, he writes that, far from being derivatively related, the principles upon which these two tribunals are founded are in fact "different and distinct" (*TMS* III.2.31-32).

> The jurisdiction of the man without, is founded altogether in the desire of actual praise, and in the aversion to actual blame. The jurisdiction of the man within, is founded altogether in the desire of praise-worthiness, and in the aversion to blame-worthiness; in the desire of possessing those qualities, and performing those actions, *which we love and admire in other people*; and in the dread of

> possessing those qualities, and performing those actions, *which we hate and despise in other people*. (*TMS* III.2.32, emphasis added)

In short, rather than evaluating our praise-worthiness based on how *other people* would, if informed and impartial, judge us, Smith's claim in the sixth edition seems to be that we self-evaluate via the impartial application of our own standards of praise-worthiness to ourselves.

The developmental story sketched above reverses Smith's (mature) position regarding the relation between the value of *actual* approval or praise and the value of *being worthy of* approval or praise. According to the developmental story, the desire for actual approval or praise comes first, and the desire to be worthy of such things follows derivatively. The latter is understood as the desire to act in a way that *would* procure our actual spectators' approval or praise, if only they were informed and impartial. But this is precisely the opposite of Smith's mature view. "So far is the love of praise-worthiness from being derived altogether from that of praise," he writes, "that the love of praise seems, at least in a great measure, to be derived from that of praise-worthiness" (*TMS* III.2.3). We desire, first and foremost, not to be praised, but to be praise-worthy; and Smith defines the latter not as that which *our spectators* would praise if they were informed and impartial, but that which *we* praise in others when *we* are informed and impartial.

Why, then, does Smith think that we desire the actual approval and praise of our peers? And what does he mean when he says that the "love of praise" is "derived from that of praise-worthiness"? Despite all that I have said, I do not wish to deny that Smith's account of critical self-evaluation is fundamentally social. He clearly thinks that our spectators' assessments of us influence our own self-evaluative judgments in deep and important ways – even that the presence of critical spectators is a necessary precondition of our own capacity to make any sort of self-evaluative judgments at all. I shall return to this in Chapter 5. Here, I will limit myself to one

comment. When Smith claims that the love of praise derives from that of praise-worthiness, his point is just this: that although our spectators' assessments of us do not *determine* the propriety or impropriety of our conduct, they do provide *evidence* one way or another. The "approbation" of our peers, Smith writes, "confirms our own self-approbation," and their praise "strengthens our own sense of our praise-worthiness" (*TMS* III.2.3). We value praise not in itself, but *as confirmation of our independently valuable praise-worthiness*. Ultimately, though, what matters, according to Smith, is not whether others praise or approve of us, or even whether they *would* praise or approve of us *if* they were informed and impartial, but whether *we* judge praise or approval proper. And if we are confident in our own conclusions, then all actual praise or approval loses its value.[73] For the confident individual, self-approbation "stands in need of no confirmation from the approbation of other men" (*TMS* III.2.8). Such self-approbation is, for Smith, the "principal object" about which one "ought to be anxious. The love of it, is the love of virtue."

[73] I shall qualify this statement in Chapter 6.

CHAPTER 5. SELF-DISTANCING AND SELF-EVALUATION

1. A Second Look at the Impartial Spectator

I argued in the previous chapter that it would be a mistake to read Smith's moral psychology as supporting a form of *cultural relativism*. That is to say, I denied that Smith's account implies that we evaluate our own conduct by passively internalizing the judgments typically made of us by our informed and impartial peers. In this chapter I develop an alternative reading. Rather than internalizing the judgments of others, I begin with the claim that we evaluate our own conduct by imagining *ourselves* the spectators of our own conduct. Impartiality "regulates" this process; but it does so, I argue, by "disciplining the *way*" we view ourselves from a spectator's perspective, "not by providing its own perspective," as implied by the relativistic reading (Darwall, 1999, p. 142, emphasis in original). We are kept in line, and forced to view ourselves impartially, by our respect for, and natural desire to sympathize with, the judgments of our peers. We are motivated to self-evaluate by feelings of accountability. Attempting to see ourselves as others see us, we are profoundly influenced by their views. Ultimately, though, our self-evaluations are our own: reflections of the feelings *we* have when we imagine ourselves the spectators of our own conduct.

Many passages in *TMS* – particularly in the late editions – support the type of reading I have in mind. After writing, for instance, that we self-evaluate by examining "our own conduct as we imagine any other fair and impartial spectator would examine it," Smith clarifies his claim:

> 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* III.1.2, emphasis added)[74]

This passage contains no hint of the claim that we internalize the judgments made of us by others. Rather than securing other people's approval, it suggests that our ultimate, if not always our immediate, aim is to secure our own approval: to conduct ourselves in a manner that *we* would approve of if *we* were to "look upon ourselves with the same eyes with which we look at others" (*TMS*, Ed. 1, 257). The passage suggests that we self-evaluate by *becoming our own impartial observers*.[75]

Other passages also point to an active role for the self in the process of self-evaluation. Smith writes, for instance, that man desires not only to be approved of, but to be "what *he himself* approves of in other men" (*TMS* III.2.7, emphasis added); he defines the "desire of praise-worthiness" as "the desire of possessing those qualities, and performing those actions, which *we* love and admire in other people" (*TMS* III.2.32, emphasis added); and he insists that our desire to be praise-worthy stems not from our desire to be praised, but from the "love and admiration which *we* naturally conceive for those whose character and conduct *we* approve of" (*TMS* III.2.3, emphasis added). Our "love and admiration" for such characters "dispose us to wish to become ourselves the proper objects of such agreeable sentiments," just as the "hatred and contempt which we naturally conceive for others, dispose us […] to dread the very thought of resembling them in any respect" (*TMS* III.2.9). It is not "the thought of being hated and despised" that scares

[74] Similar claims appear in earlier editions. In the second edition, for instance, Smith writes: "If, when we place *ourselves*, in the situation of [an impartial spectator], our own actions appear *to us*, under an agreeable aspect, if *we* feel that such a spectator cannot avoid entering into all the motives which influenced us, whatever may be the judgments of the world, we must still be pleased with our own behavior, and regard ourselves, in spite of the censure of our companions, as the just and proper objects of approbation" (*TMS*, Ed. 2, 208, emphasis added).

[75] D.D. Raphael writes: "The 'supposed impartial spectator', as Smith often called him, is not the actual bystander who may express approval or disapproval of my conduct. He 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" (1975, p. 90).

us, but "that of being hateful and despicable." Similarly, after claiming that our observations of other people's conduct "leads us to form to ourselves certain general rules concerning what is fit and proper either to be done or to be avoided," Smith concludes that such rules are "ultimately founded upon experience of what, in particular instances, *our* moral faculties, *our* natural sense of merit and propriety, approve, or disapprove of" in others (*TMS* III.4.7-8, emphasis added). In each passage, Smith grounds our self-evaluations in *our* assessments of others, not *their* assessments of us.

This way of interpreting Smith's account has several benefits over the reading considered in the previous chapter. Most obviously, by taking seriously his claim that we self-evaluate by imagining and simulating a response to our spectators' circumstances, the interpretation gives a central role to Smith's distinctive (i.e., evaluative, simulationist) notion of sympathy, thereby both accommodating his claim that the "principle" by which we evaluate our own conduct is "altogether the same" as that by which we evaluate the conduct of others and making sense of his assertion (discussed at length in Chapter 4) that our self-evaluative judgments necessarily refer to "what, we imagine, ought to be the judgments of others" (*TMS* III.1.2).[76] Moreover, by grounding our self-evaluative judgments in the mechanism of imaginative mental simulation, the interpretation accurately depicts self-evaluation as an *active* evaluative process – something that people *do*, rather than something that *happens to* them. Finally, by allowing for the possibility that judgments reached through imaginative simulation may in fact differ from those of even our

[76] Recall that Smith's simulation-based account of sympathy is fundamentally evaluative. We determine how other people *ought* to feel, he claims, by imagining being in their circumstances and simulating a response. To say that we engage in critical self-evaluation by imagining and simulating a response to our spectators' circumstances is thus, for Smith, equivalent to saying that we engage in critical self-evaluation by determining how our spectators *ought* to feel about our conduct. That our self-evaluative judgments bear a secret reference to "what, we imagine, ought to be the judgments of others" is entailed by my interpretation.

most informed and impartial spectators, the interpretation opens the door to disagreement with, and criticism of, our own society's dominant evaluative standards.

In this chapter, I develop the above reading and attempt to make it plausible, both as an interpretation of Smith's view and as an account of self-evaluation in its own right. It is the latter that interests me most. Though I believe that my reading captures Smith's view more accurately than either of the alternatives discussed in the previous chapter, my aim in this chapter is only partly exegetical. More than anything, my aim is to highlight what I take to be the fundamental features of Smith's account, and to then develop and augment those features in light of recent research in empirical social psychology. I treat Smith not as an object of historical study, but as a colleague struggling to answer the same questions that I myself find so puzzling. My hope is that by "thinking with Smith" about these problems (as opposed to merely thinking *about* Smith's account), I will find myself "nudged" in the direction of a solution (Fleischacker, 2011a, p. 40). My primary objective in this chapter is thus to think *with* Smith about how to develop a plausible and empirically defensible account of critical self-evaluation that is capable of both expressing and supporting Smith's own egalitarian conception of society – without thinking too much about which aspects of my ultimate proposal are *Smith's*, and which are merely *Smithian*.

The chapter proceeds as follows. I begin by asking whether the psychological mechanism of "self-distancing" (imagining oneself from a spectator's perspective) can, on its own, provide us with any critical perspective on our own feelings or behavior. I argue in Section 2 that it can. More specifically, I present evidence that by adopting the perspective of our own spectator, we can moderate our own emotional, behavioral, and psychological reactivity, thereby enabling us to detach from our emotions and engage in a more critical and thoughtful self-evaluative process. I begin to examine the social components of this process in Section 3, where I argue that we are

motivated to self-distance by feelings of accountability, triggered by the presence of those to whom we are accountable. I conclude that the presence of other people is indeed a necessary precondition of our capacity to self-evaluate, just as Smith claims (see *TMS* III.1.3-5), but that this is consistent with my claim that we self-evaluate by imagining ourselves the spectators of our own conduct. The presence of other people is necessary, I argue, not because we need other people to tell us how to self-evaluate (as implied by the relativistic reading), but because it is they who, by holding us accountable for our conduct, provide us with the motivation to consider ourselves from a spectator's point of view. I continue to explore the role of social interaction in the self-evaluative process in Section 4, where I argue that, in addition to triggering us to self-distance, the presence of actual observers can, in the right social circumstances, *improving the way* we self-distance by exposing us to new information and improving the way we process it, motivating us to consider a wider range of perspectives, reducing our susceptibility to various forms of cognitive bias, and generally enhancing the quality, creativity, and originality of our evaluative thought.

I develop and defend an active, non-relativistic, sympathy-based account of critical self-evaluation that simultaneously embodies Smith's commitment to egalitarian normative principles and provides an empirically defensible psychological explanation of how individuals can, over time, come to both endorse his conception of society as a moral community of independent and mutually accountable equals, and use that ideal as the basis on which to criticize both themselves and the society in which they live.

2. Self-Immersed vs. Self-Distanced Self-Reflection

According to Smith, we evaluate the propriety of other people's conduct by simulating a response to their circumstances and comparing the results to reality. We judge their feelings and behavior proper insofar as we find that we can sympathize with them, and we judge their feelings and behavior improper insofar as we cannot. As I mentioned at the beginning of Chapter 4, when we evaluate other people, our challenge is one of empathy and understanding: of fully appreciating the circumstances and perspectives of those we judge. We are critical by default. In contrast, when we evaluate ourselves, our judgments often suffer from too much empathy and understanding. Our primary challenge when self-evaluating is thus not to find a way to empathize with or understand ourselves, but to find a way to distance ourselves from our own circumstances and feelings – to take up something like the critical perspective that we so naturally assume whenever we evaluate the conduct of others.

Judging ourselves precisely as we judge others won't work, however. For, as just noted, we judge others by imagining and simulating a response to their circumstances. This may work when we are not in the circumstances of the person under evaluation (as argued in Chapter 3), but it would have no chance of succeeding when we *are* – as is necessarily the case when we engage in real-time self-evaluation. By *directly* simulating a response to our own present circumstances, we would simply replicate whatever it is that we actually feel. Lacking critical distance, we would approve of ourselves in each and every case. It is for precisely this reason that Smith denies that we can ever "survey our own sentiments" without first removing ourselves "from our own natural station" and endeavoring "to view them as at a certain distance" (*TMS* III.1.2). In real time, our passions always appear to "justify themselves" and "seem reasonable and proportioned to their objects" (*TMS* III.4.3). In order to effectively self-evaluate, we must

first distance ourselves from our circumstances and sentiments. We must view ourselves from a different point of view.

Let us turn, then, to my proposal. I have claimed that Smith thinks we self-evaluate by imagining ourselves the spectators of our own conduct – or, more precisely, by imagining and simulating a response to the circumstances of our spectators. Could this approach provide us with the critical distance needed for effective self-evaluation? Is it any different from the process rejected in the previous paragraph?

It may seem doubtful that my proposal would provide us with any critical distance at all – at least in the case of *real-time* self-evaluation. I have suggested that we self-evaluate by imagining and simulating a response to the circumstances of our spectators. By assumption, though, our spectators are at the same time evaluating us by simulating a response to our circumstances. It follows, then, that if we were to simulate a response to our spectators' circumstances, we would be simulating a response to the circumstances of people *who are simulating a response to our own*. My proposal thus reduces to the claim that we self-evaluate in the present by *simulating a simulation of life in our own shoes*: by imagining, from our own present circumstances, what it would be like to simulate a response to those very same circumstances from an external point of view.

In the case of real-time self-evaluation, our circumstances and personal characteristics are identical to those of the person we assess, and I have argued that self-evaluation via *direct* simulation would be ineffective for this very reason: I cannot possibly hope to effectively self-evaluate by simply imagining and simulating a response to my own present circumstances. But if this is correct, then shouldn't my proposed mechanism be just as ineffective? I too have claimed that Smith thinks we self-evaluate by simulating a response to our own present circumstances.

The only difference is that I have claimed that we do so *indirectly*: that our simulation of life in our own shoes is embedded within a simulation of life in the shoes of another. But given that we *are* in our own shoes – that we are *not really* the spectators of our own conduct – should this make a difference? Is there any reason to think that we could acquire a critical perspective on ourselves by *pretending* that our circumstances are not as they are? Can simply *imagining* ourselves the spectators of our own conduct actually change the way we see ourselves?

Smith thinks it can. As "nature teaches the spectators to assume the circumstances of the person principally concerned," he writes, "so she teaches this last in some measure to assume those of the spectators" (*TMS* I.i.4.8).

> 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. 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*. (*TMS* I.i.4.8, emphasis added)

Because our simulated spectatorial passions will typically be "much weaker" than the originals, the process of indirect simulation will "[abate] the violence" of our initial feelings (*TMS* I.i.4.8). In short, Smith thinks that imagining ourselves the spectators of our own conduct will provide us with a degree of critical distance by *moderating our emotional reactivity.*

As it turns out, recent empirical research suggests that Smith is exactly right. In a series of experiments, social psychologist Ethan Kross and his colleagues have studied the effects of engaging in self-reflection from two different perspectives: a *self-immersed* perspective, in which past "events and emotions are experienced in the first person," and a *self-distanced* perspective, in which past events are experienced from the perspective of a third-person or distant observer

(Kross, Ayduk, & Mischel, 2005, p. 710). Building on the "hot/cool" framework developed by Metcalfe and Mischel (1999), Kross and his colleagues hypothesized that adopting a self-immersed perspective would lead people to focus on narrowly *recounting* the concrete details of their past experiences (e.g., what they felt at the time), whereas adopting a self-distanced perspective would prompt them to *reconstrue* their experiences in a broader context. Moreover, they predicted that by enabling individuals to "contemplate emotional experiences without activating intense levels of affect," the latter perspective would allow them to process and evaluate their experiences in a "cool, reflective mode" (Kross et al., 2005, p. 710). In short, the authors predicted that self-distanced self-reflection would lead to precisely the *moderated emotional responses* hypothesized by Smith.

To test their hypotheses, the researchers divided subjects into two groups. They instructed the members of one group to recall an experience in which they felt overwhelming anger and hostility from a self-immersed perspective ("go back to the time and place of the experience and relive the situation as if it were happening to you all over again"). They instructed the members of the second group to do the same, only from a self-distanced perspective ("take a few steps back and move away from your experience […] watch the conflict unfold as if it were happening all over again to the distant you"). They then compared the responses of the subjects in the two groups.

The results are striking.[77] As predicted, Kross et al. (2005) found that self-reflection from a self-distanced perspective led to lower levels of emotional reactivity than self-reflection from a self-immersed perspective. In other words, they found that imaginatively viewing oneself from the perspective of an observer yields more moderated emotional responses than does viewing

[77] I will mention only the highlights. For comprehensive reviews, see Kross (2009) and Kross and Ayduk (2011).

oneself from a first-person perspective – less intense and more reflective, cool rather than hot. And there is reason to think that the effect may be at least somewhat general; to date, it has been observed in cases of both anger (Ayduk & Kross, 2008; Kross et al., 2005) and depression (Kross & Ayduk, 2008). Moreover, in addition to moderating emotional and behavioral reactivity (e.g., aggressive thoughts and behavior, sad feelings), the perspective adopted by an individual has been found to have lasting physiological consequences. Compared to subjects who adopted a self-immersed perspective, Ayduk and Kross (2008) found that subjects who reflected from a self-distanced perspective exhibited smaller increases in blood pressure, both while actively analyzing their feelings and during a recovery period following the experiment. The consequences of perspective were found to be both deep and lasting.

Granted, these experiments are limited in an important respect: they examine self-reflection with respect only to *past* experiences. As Smith notes, though, we do not only evaluate our past conduct; we also do so in the present, "when we are about to act" (*TMS* III.4.2). And it is real-time self-evaluation that presents the challenge for the view that I am considering, as it is only in the present that we necessarily share the circumstances and characteristics of the person we judge. Moreover, it is in the present that our capacity for critical self-evaluation is most important, both because it is when we are about to act that our self-evaluative judgments are most likely to affect our behavior and because it is then that we are most likely to be led astray by partiality and self-love. Unfortunately, as Smith himself points out, it is also in the present, when we are about to act, that our ability to self-distance is most likely to fail.

> When we are about to act, the eagerness of passion will seldom allow us to consider what we are doing, with the candour of an indifferent person. 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* III.4.3).

Caught up in the moment, our passions will appear to "justify themselves" and "seem reasonable and proportioned to their objects, as long as we continue to feel them" (*TMS* III.4.3). It is when we are about to act, not afterwards, that we are most susceptible to self-deceit – the "source of half the disorders of human life" (*TMS* III.4.6).

Smith does suggest a way of getting around this problem – at least in part. Despite the difficulty of real-time self-evaluation, he notes that we can protect ourselves from most of the dangers of self-deceit by following a set of general rules, derived from our past observations of others. When, in the position of a spectator, we see other people behaving inappropriately, we can "resolve never to be guilty of the like" and "lay down to ourselves a general rule, that all such actions are to be avoided" (*TMS* III.4.7). And when, again as spectators, we observe other people behaving admirably, we can "lay down to ourselves a rule of another kind, that every opportunity of acting in this manner is carefully to be sought after." Whenever we find ourselves incapable of self-distancing, in other words, we can simply regulate our conduct by following a set of rules derived from our past assessments of others, made when we were in a calm state of mind.

It would be unwise to make too much of this strategy, however. Though Smith thinks that we can benefit from such general rules, he explicitly denies that they can provide a complete solution to the problem. General rules can provide a rough guide to virtuous conduct when no alternative is available, but they cannot substitute for genuine, simulationist, real-time self-evaluation. For, Smith argues, the dictates of nearly[78] all the virtues – prudence, charity,

[78] The one exception is justice (see *TMS* III.6.10-11).

generosity, gratitude, friendship, and so forth – are "loose and inaccurate, admit of many exceptions, and require so many modifications, that it is scarce possible to regulate our conduct entirely by a regard to them" (*TMS* III.6.9).[79] General rules of conduct can provide "a general idea of the perfection we ought to aim at," a helpful rule-of-thumb, but nothing more (*TMS* III.6.11). They cannot "afford us any certain and infallible directions for acquiring" such perfection, as there are no "rules by the knowledge of which we can infallibly be taught to act upon all occasions with prudence, with just magnanimity, or proper beneficence." There are no precise measures of the "fitness or propriety" of our conduct, other than "the sympathetic feelings of the impartial and well-informed spectator" (*TMS* VII.ii.1.49). The ideal agent must rely not on general rules, but on real-time self-distanced self-reflection.

For Smith, then, it is essential that individuals be capable of self-distancing in the present, as well as the past. Fortunately, recent research has addressed this point as well, and the results are again favorable. Mischkowski, Kross, and Bushman (2012) found that subjects who were provoked to anger *and then immediately instructed to self-distance* displayed lower levels of aggression than those who were either instructed to reflect from a self-immersed perspective or given no guidance at all. The authors concluded that by imaginatively viewing themselves from a detached perspective, people can distance themselves from their feelings *even as they feel them*,

[79] Smith defends his claim with a (strategically chosen) discussion of *gratitude*: "Of all the virtues I have just now mentioned, gratitude is that, perhaps, of which the rules are the most precise, and admit of the fewest exceptions. That as soon as we can we should make a return of equal, and if possible of superior value to the services we have received, would seem to be a pretty plain rule, and one which admitted of scarce any exceptions. Upon the most superficial examination, however, this rule will appear to be in the highest degree loose and inaccurate, and to admit of ten thousand exceptions. 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? Nor, or to-morrow, or next month? And for how long a time? It is evident, that no general rule can be laid down, by which a precise answer can, in all cases, be given to any of these questions" (*TMS* III.6.9).

and that doing so "attenuates both aggressive thoughts and angry feelings" (Mischkowski et al., 2012, p. 1189).

We may conclude, then, that self-distancing does make a difference. By imagining ourselves the spectators of our own conduct, we can detach from our "hot" emotions and facilitate a more critical and thoughtful self-evaluative process – not only with respect to our past conduct, but in the "heat of the moment" as well. Doing so will not always be easy, and general rules of conduct may sometimes be needed to keep us in line. But, given the right conditions, the available empirical evidence suggests that imagining and simulating a response to the circumstances of our spectators can indeed enhance our critical perspective and facilitate effective self-evaluation.

3. Motivating Self-Evaluation

We *can* self-distance; we can even do it spontaneously (Ayduk & Kross, 2010), and in real-time (Mischkowski et al., 2012). And, as I argued in the previous section, doing so can enhance our self-evaluative judgments. It does not follow, though, that, on any given occasion, we *will* self-distance. The default perspective from which people typically reflect on their experiences is a self-immersed one (Ayduk & Kross, 2010; Kross & Grossmann, 2011; Mischkowski et al., 2012; Nigro & Neisser, 1983). Left to our own devices, it is unlikely that we would venture beyond our own self-immersed perspectives. It is therefore not enough to show that we *can* self-distance; we need to identify something capable of *motivating* us to self-distance. In this section, I argue that Smith's solution is that we are motivated to self-distance by the presence of our peers – and that there is good reason to think he is right. There is considerable empirical evidence that we are affected by the presence of other people in roughly the way Smith suggests. We are moved to

self-self-evaluate from a self-distanced perspective, I will argue, by our feelings of accountability to those with whom we live.[80]

Let's begin with an example. Suppose that I have a tendency to anger easily, particularly when driving. Stuck behind a driver who doesn't go when the light turns green, for instance, I become irate and express myself accordingly – honking, yelling, and so forth. Suppose that I am in such a state right now. If I were to reflect on my conduct at a later point in time, I would disapprove of it. At present, though, I am not reflecting on it and do not disapprove. Caught up in the moment, I am thinking only about the driver and the delay, all from a first-person perspective. Immersed in my anger, I have no doubt of its propriety. As long as I continue to feel it, my anger appears to me both "reasonable and proportioned" to its object, just as Smith says (*TMS* III.4.3).

Suddenly, right in the middle of a particularly long honk, a car pulls up beside me. The driver is clearly disgusted by my behavior, and, catching a glimpse of her disapproving look, I feel a change come over me. My attention shifts from the driver in front of me to the one beside me. More precisely, it shifts *via* the new driver *to myself*. Instead of focusing exclusively on the object and cause of my anger, I begin to consider my anger itself, and its relations to its object and cause, all from the new driver's point of view. I begin to experience a new set of emotions: shame for overreacting to such a trivial delay, and guilt for treating the driver in front of me so rudely and with such disrespect. As Harry Frankfurt (1971) might put it, in addition to my first-order desires to get where I want to go, honk my horn, and yell, I begin to develop (or find that I

[80] As noted in Chapter 4, Darwall (2004) argues that Smith's account of propriety presupposes norms of equality and independence, and that his theory of justice adds mutual accountability. The view I develop in this chapter and the next is a variant of Darwall's. Though I think Darwall right about the importance of equality and independence, I think he understates the role of mutual accountability in Smith's work. It is central not only to Smith's theory of justice, but to his entire account of evaluative judgment.

have always had) a second-order desire *not* to act on these last two first-order desires.[81] I may continue to feel angry, as anger can be difficult to manage and my newfound perspective may not eliminate it entirely. But I will nevertheless begin to see myself in a new light.

In this example, I began by viewing things from a self-immersed perspective, focusing entirely on the object and cause of my anger, oblivious to my anger itself. I was forced out of this perspective, and triggered to self-distance, by my interaction – minimal as it was, no more than a passing glance – with an observer. This is precisely what Smith predicts. A person raised and living in isolation, he writes, would "scarce ever" consider his own passions, as the objects of his passions "would occupy his whole attention" (*TMS* III.1.3). Bring him into society, though, and everything immediately changes. Observing other people's reactions to his conduct, his "joys and sorrows" immediately become the "causes of new desires and aversions, new joys and new sorrows." His own passions, previously so uninteresting to him, begin to "interest him deeply" and "call upon his most attentive consideration." Driving alone, I honk and yell at the driver in front of me, thinking of nothing but the delay caused by his inattentive behavior. But introduce a third driver and my attention immediately shifts. Feeling accountable for my conduct, I begin to scrutinize it from a spectator's point of view. I am "led to imagine" the manner in which I would be affected if I were "only one of the spectators" of my own situation" (*TMS* I.i.4.8). I "look at it" through my spectators' eyes, "especially when in their presence and acting under their observation." The presence of others motivates me to self-distance.

Smith highlights this effect again and again. The "moment" we come into another person's presence, he writes, we "are immediately put in mind of the light in which he will view our situation," and we thus "begin to view it ourselves in the same light" (*TMS* I.i.4.9). Similarly:

[81] Charles Griswold discusses the connections between Smith and Frankfurt (Griswold, 1999, p. 105).

> In all private misfortunes, in pain, in sickness, in sorrow, the weakest man, when his friend, and still more when a stranger visits him, is immediately impressed with the view in which they are likely to look upon his situation. Their view calls off his attention from his own view; and his breast is, in some measure, becalmed the moment they come into his presence. (*TMS* III.3.23; see also I.i.4.9)

"In solitude," he writes, we "feel too strongly whatever relates to ourselves" (*TMS* III.3.38). We are "too much elated by our own good, and too much dejected by our own bad fortune." We need the presence of another to correct our initial response. The "man within, the abstract and ideal spectator of our sentiments and conduct" requires "to be awakened" by "the presence of a real spectator" (*TMS* III.3.38). Indeed, Smith goes so far as to identify "[s]ociety and conversation" as "the most powerful remedies for restoring the mind to its tranquility" (*TMS* I.i.4.10). They, more than anything else, he claims, prevent us from becoming lost in our own point of view. They change the way we view ourselves, enabling us to recognize our "real littleness," and, by countering the "natural misrepresentations of self-love," help us to make a "proper comparison between our own interests and those of other people" (*TMS* III.3.5 and III.3.1).

Once again, existing empirical evidence is broadly supportive of Smith's claim. There is a great deal of evidence that the presence of other people can, and often does, trigger us to think differently. Researchers have found, for example, that people participating in laboratory-controlled cooperative games exhibit more prosocial behavior when they speak with other people about their deliberations than when they deliberate alone (Dawes, McTavish, & Shaklee, 1977; Kurzban, 2001). In fact, although verbal communication is sufficient, it isn't even necessary. People can be triggered to behave more cooperatively by instructing them to touch or make eye contact with one another (Kurzban, 2001). Even less direct interaction can do the trick. Andreoni and Petri (2004) found that individuals participating in a repeated *Public Goods Game* exhibited greater generosity when they were told that their photographs and information regarding their

past contributions would be made available to other members of their experimental group.[82] Even the prospect of indirect social interaction can change the way we think.

In fact, recent experiments demonstrate that people are even more sensitive to the presence of others than the aforementioned results suggest. In another laboratory-controlled experiment, Burnham (2003) found that subjects playing the role of a dictator in a one-shot *Dictator Game* gave more money to their designated recipients when they were told that their photographs would be shown to their recipients (the "dictator photo" treatment) *or when photographs of their recipients were shown to them* (the "recipient photo" treatment).[83] Viewed from the perspective of narrow self-interest, both results are puzzling;[84] but it is the behavior observed in the recipient photo treatment that is most interesting. Whereas the results of the dictator photo treatment could perhaps be explained by the value of maintaining a good reputation, such an explanation could not account for the behavior observed in the recipient photo treatment. For the subjects in that treatment did not believe their actions to be public. On the contrary, they were explicitly told that they were private.[85] It follows that their behavior cannot be explained by their belief that their reputations were on the line.

Why, then, did subjects in the recipient photo treatment behave like subjects in the dictator photo treatment? Why did they act *as if* they believed their actions to be public? According to

[82] In a *Public Goods Game*, subjects choose how much of their money to keep for themselves and how much to donate to the "public good". Contributions made to the latter are multiplied by some factor and divided evenly among players.

[83] In a *Dictator Game*, each subject (a "dictator") decides how much of their money to keep and how much to give to a designated "recipient". In contrast to the situation in an *Ultimatum Game*, in which the recipient can accept or reject a proposal, the recipient in a *Dictator Game* has no choice. In Burnham's experiment, subjects playing the role of the dictator in the control group were neither shown photographs of others nor told that their photographs would be shown to others.

[84] The game was not repeated, the experiment was run under double blind conditions, and the participants knew all of this to be the case. The subjects knew, in other words, that no retaliation, whether by the recipient or the experimenter, would have been possible; and it was therefore never in their interest to give away money.

[85] Each subject in the recipient photo treatment was given the following written explanation: "neither the experimenter nor the person you are paired with will learn your identity either during or after the experiment" (Burnham, 2003, p. 143).

Burnham, subjects in the recipient photo treatment were influenced not by their *belief* that their actions were public, but by their *perception of cues* indicative of their actions' observability. We evolved in a world, he argues, in which the observation of another person typically implied that one was observable oneself. Until the advent of cameras, "the ability to see a person, particularly her or his eyes, meant that those eyes could see you" (2003, p. 137). It would not be surprising, then, if, to avoid punishment and protect our reputations, we developed a tendency to respond in certain specifiable ways to environmental cues correlated with the presence of observers. According to Burnham, subjects in the recipient photo treatment acted more altruistically than subjects in the control group because their observation of another person's photograph *activated in them a different sort of deliberative process*.

Subsequent experiments have replicated and extended Burnham's findings, demonstrating just how sensitive people are to the presence of other people. For example, Haley and Fessler (2005) found that people acted more generously when "stylized eye-like shapes" appeared on their computer screens. Rigdon, Ishii, Watabe, and Kitayama (2009) replicated these results using three dots arranged in a face-like pattern, and Burnham and Hare (2007) found that subjects in a *Public Goods* game acted more altruistically when facing a robot with human-like eyes. An experiment by Mifune, Hashimoto, and Yamagishi (2010) provides additional evidence for the "eyes effect", though the researchers found that the effect only occurs when the recipient is perceived by the subject to be a member of his or her "in-group". Finally, Bateson, Nettle, and Roberts (2006) and Ernest-Jones, Nettle, and Bateson (2011) generalized the results of the earlier laboratory experiments to non-laboratory settings. The former found that an image of a pair of eyes increases contributions to an "honesty box" used to collect money for drinks in a university

coffee room, and the latter found that cafeteria posters directing people to clean their tables yield better results when they feature an image of a pair of eyes.[86]

Strictly speaking, these results do not show that we are triggered by the presence of others to engage in self-distanced self-reflection. The observed behavior is consistent with a number of hypotheses, and additional research is needed to identify the precise psychological mechanism(s) responsible for the results.[87] But the findings are still highly suggestive. By identifying a link between the shift in our thought and the observation of cues indicative of our own observability, the results point to a possible mediatory role for the consideration of how we look from the perspective of an observer. That is to say, the results suggest not only that observing cues indicative of our own observability enhances our cooperative tendencies, but that it may do so *by triggering us to self-distance*. This thesis gains additional support from a recent study by Kross and Grossmann (2011), who found that the psychological mechanism of self-distancing can itself increase people's cooperative tendencies. We know that the observation of cues indicative of our own observability leads to enhanced cooperation, and we have evidence that self-distancing does the same. It would be natural to conclude, then, that the presence of other people influences our thought by prompting us to imaginatively view ourselves from an observer's perspective. The empirical findings suggest – if inconclusively – that social interaction triggers self-distancing.

[86] Some researchers are skeptical of the "eyes effect", and it has not been observed across all experimental conditions (Ferh & Schneider, 2012). But I will not make much of this point here. Whether or not the image of a pair of eyes in particular can, by itself, trigger a special mode of deliberation will not matter for my purposes. What matters is that *some* form of interpersonal interaction – conversation, physical contact, the presence of real observers, anything that might draw one's attention to the perspective of another – can yield this result. Whatever the status of the "eyes effect", existing empirical evidence provides strong support for the general point that social deliberation differs from deliberation conducted in isolation. That is all I need.

[87] Burnham (2003) and Haley and Fessler (2005), for instance, think that the results are best explained by reference to the experimental subjects' underlying desire to protect their reputation and avoid punishment. Though their actions are private, and neither harm to their reputation nor punishment could possibly come about, the authors suggest that the subjects' observation of cues indicative of their observability nevertheless activated subconscious mental processes that lead to conduct aimed at protecting their reputation and avoiding future punishment.

Let's review our progress up to this point. Smith thinks that the presence of another person triggers us to view ourselves from a spectator's perspective, and that doing so changes the way we think, feel, and view ourselves – moderating our emotional reactivity and reducing the weight we give to our own interests in our practical deliberations, relative to the interests of others. Recent empirical research confirms that the presence of other people tends to increase our prosocial behavior, as does the mechanism of psychological self-distancing. Moreover, self-distancing has been found to moderate emotional, behavior, and psychological reactivity. Though more research is needed, the existing empirical evidence suggests that Smith may be on to something: that the presence of other people may indeed trigger us to self-evaluate via the mechanism of self-distanced self-reflection, thereby moderating our emotions, curbing our self-love, and enhancing our prosocial behavioral tendencies. The presence of spectators may cause us to think more like an egalitarian.

4. Social Interaction and Critical Distance

I have claimed that Smith thinks we self-evaluate by viewing ourselves from the perspective of an observer; and, among other benefits, I have argued that doing so moderates our emotional, behavioral, and physiological reactivity, thereby facilitating a more critical and thoughtful self-evaluative process. Caught up in the moment, I honk and yell at the driver in front of me. I feel justified in my conduct, but only because my direct involvement in the situation has left me incapable of examining myself critically. Self-distancing can change this. By psychologically distancing myself from my circumstances, I can shield my judgments from the effects of "hot" emotional outbursts, increasing the likelihood that my real-time self-evaluations will mesh with my "cool" evaluative convictions.

This benefit is important, but limited. Complete self-evaluation involves more than just a check on our emotions. In addition to reflecting on the intensity of our feelings, we must examine the underlying factors that cause us to feel or act as we do: our beliefs and assumptions, our desires and preferences, our principles and commitments, and so forth. We must ask not only whether we are living up to our own standards of conduct (i.e., whether our real-time self-evaluative judgments mesh with our "cool" evaluative convictions), but whether we *ought to* be living up to them: whether the standards we presently endorse are *right*. We must scrutinize the psychological processes on which our practical and evaluative conclusions rely – all the while searching for any hint of bias, particularly with respect to ourselves. Though psychological self-distancing may provide an effective antidote to the effects of our "hot" emotional outbursts, I see no reason to think that, on its own, it could enable us to perform these additional tasks. In Chapter 4, I rejected the claim that Smith is committed to a form of cultural relativism. It may be thought, though, that my own proposal is no better. By insisting that we self-evaluate by imagining ourselves the spectator of our own conduct – by measuring ourselves against our own cool evaluative standards – it may seem that I have done no more than replace *cultural* relativism with an equally unattractive form of *individual* relativism.

My objective in this section is to show that this need not be the case. By supplementing the process developed in the previous sections, we can gain the distance needed to engage in a more comprehensive self-evaluative process. As in the previous section, I will once again argue that social interaction is key – though I will now put social interaction to a slightly different use. In addition to motivating us to self-distance, I will argue that interacting with our peers can enhance our critical perspective by *altering the way we see ourselves*.

How does this work? How can interacting with other people affect the way we see ourselves? Some effects are obvious. Our observations of other people's conduct, for example, can inform our own attempts to self-distance. Observing another person honking and yelling in traffic, for instance, can show us just how contemptible it really is, thereby altering the way we view our own tendency to exhibit such behavior – and perhaps even the way we view ourselves in the "heat of the moment".[88] Moreover, interacting with a group of diverse of people can provide us with an important epistemic benefit. Numerous empirical studies have found that individuals who deliberate within heterogeneous groups of people tend to exchange and discuss more information than those who deliberate alone or within homogeneous groups of like-minded individuals.[89] By interacting with a diverse group of people, individuals thus expose themselves to more information. And this suggests that, all else equal, individuals who interact with a diverse range of people will incorporate more information into their own deliberations than people who are deprived of such experiences.

What sort of information? There is a range of possibilities. In addition to the obvious sorts of descriptive facts to which an individual could be exposed, social interaction can expose one to equally important, though less easily quantifiable, facts about *what it is like* to be a different sort of person or face a vastly different set of circumstances (e.g., to be a member of an historically stigmatized group). Similarly, discussion with a diverse group of people can expose one to the reasons (good and bad) people have for endorsing rival moral or political principles, or rejecting one's own. Interacting with a diverse group of others can give one a sense of why different people subscribe to different belief systems, live different lifestyles, pursue different goals, support different causes, and so forth. In short, by interacting with a diverse set of people, one

[88] Such observations provide the foundation for Smith's account of the "general rules" of morality (see Section 2).
[89] The literature on this is enormous. For a small sample, see Hans and Vidmar (1982), L. R. Hoffman and Maier (1961), Mannix and Neale (2005), Nemeth (1995), and Sommers (2006).

can gain access to precisely the sort of information one needs to engage in a more comprehensive form of self-evaluation: factual assertions capable of challenging one's default beliefs and assumptions, and a rich and varied set of alternative perspectives from which to (try to) examine and critique one's own desires and preferences, principles and commitments, and so forth.

Smith does not mention the epistemic benefit of social interaction, but his account of self-evaluation is perfectly positioned to exploit it. His central claim, recall, is that we approve or disapprove of ourselves by sympathy with the approbation or disapprobation of our informed and impartial spectators (see *TMS* III.1.2), and he offers a simulation-based account of sympathetic experience. He thinks we self-evaluate, in other words, by attempting to see ourselves from *our spectators'* points of view. Up to this point, I have been arguing that Smith thinks we self-evaluate by imagining ourselves the spectators of our own conduct, and I have claimed that Smith's comments regarding the role of sympathy in the self-evaluative process support this view. But Smith's comments actually suggest something more precise: that we self-evaluate by imagining ourselves not merely as spectators of our own conduct, but as *particular* spectators – as *real* people, with *real* beliefs and assumptions, desires and preferences, principles and commitments, etc., which may or may not line up with our own. We don't merely self-distance, in other words; we self-distance – or try to self-distance, at any rate – *as other people.* We engage in a process of *social* self-distancing.

I am not suggesting, of course, that Smith thinks we will necessarily internalize the views of those with whom we interact. Unlike Hume's account, according to which we sympathize with others by (roughly speaking) feeling whatever they feel, Smith's account of sympathy is evaluative (see Chapter 2). When, by the use of the imagination, we simulate a response to other people's circumstances, we do so, Smith thinks, neither to identify nor to experience their *actual*

sentiments, but to determine the sentiments that it would be *proper* for them to feel. Our natural desire to sympathize with others leads us only to *consider* the views of those with whom we interact – to attempt, to the best of our ability, to see things as they do. But it does not guarantee that we will actually sympathize with any of them. By exposing ourselves to the scrutiny of others, by considering their arguments and attempting to understand their way of thinking, we *may* come to sympathize with their appraisals and change the way we see ourselves; but we may not. We will internalize only those evaluative sentiments with which we sympathize. We will internalize only the evaluative sentiments that we deem *reasonable* or *well suited* to their object (see *TMS* I.i.3.6).

Smith's claim, I want to suggest, is neither that we self-evaluate by passively internalizing the judgments typically made of us by our informed and impartial peers nor that we self-evaluate by imagining ourselves as some abstract ("fly on the wall") spectator of our own conduct. Rather, Smith's claim is that we self-evaluate by exposing our conduct to our *actual* spectators' scrutiny, observing their assessments and listening to their arguments, absorbing as much information as we can, and exploring our (limited) capacity to see ourselves as *they* see us. If I am right, then our interactions with our peers play two essential roles in the self-evaluative process: they both trigger us to self-distance and enhance our critical perspective when we do self-distance.

Is this plausible? Can interacting with people enhance our critical perspective as I have suggested – without returning us to the sort of cultural relativism rejected in the previous chapter? There is room for skepticism. Earlier, I cited evidence showing that people who deliberate in heterogeneous groups tend, on average, to exchange and discuss more information than people who deliberate alone or in homogenous groups. I claimed that this gives us reason to

think that people who deliberate in heterogeneous groups will, on average, incorporate more information into their deliberations than those who are deprived of social interaction. But this is an invalid inference. From the fact that people who deliberate in groups tend, on average, to *exchange and discuss* more information it does not necessarily follow that they will *use* such information *to correct their mistaken views*. In fact, there is reason to think that they will not. "Prejudiced people," Fleischacker correctly notes, "are notoriously impervious to better information about the objects of their dislike," however much they hear it discussed (2011a, p. 30). They do not process such information neutrally; it is "filtered through their sentiments." And rather than modifying their initial position, they often respond by attempting to explain the new information away. They highlight evidence and arguments (however weak) that support their default view and search for reasons to dismiss any evidence or argument to the contrary.[90] In short, they employ their intellectual capacities for no other purpose than to rationalize their own view.[91]

Actually, there are two problems here. People may behave as just described. They may dismiss any claim that conflicts with their own view, regardless of its strength. Given the right circumstances, through, they may also do the opposite: exposed to other people's views, they may simply adopt the positions of those around them, regardless of their plausibility. This, of course, is the core claim of the relativistic reading: that we simply internalize the dominant views of those around us. As a descriptive thesis, there is undoubtedly something to it. It is well known that individuals tend to adopt (or, at least, feign adopting) the belief of people with whom they

[90] See, for example, Jones and Kohler (1958), Lord, Ross, and Lepper (1979), and Edwards and Smith (1996). For a review, see Aronson (2011).
[91] For a discussion of this phenomenon in the context of moral judgments, see Haidt (2012). See also (Fleischacker, 2011b) for a discussion of self-deceit in the context of Smith's work.

converse whenever the latter believe with unanimity, regardless of the belief's plausibility.[92] And a similar effect has been found in the case of behavior as well: individuals tend to mirror the behavior of those with whom they interact, particularly when there is a considerable degree of uniformity in the group members' conduct.[93] When it exists, a group consensus can exert tremendous pressure on individuals to conform.

It is not enough, then, simply to observe that people who interact within diverse groups are exposed to new information and perspectives. In order for them to make productive use of this exposure, they must be open to the possibility of being mistaken; they must be willing to honestly consider opposing points of view; and their reasoning must be sufficiently free from bias. At the same time, though, they must not be too credulous, too susceptible to the influence of others, lest the process of social self-distancing dissolve into precisely that form of relativism against which I argued in Chapter 4. Ideally, the self-evaluative process should *enhance* each individual's capacity for free and independent thought – improve their reasoning, support their individual creativity, and so forth – not inhibit or destroy it. In order to show that the account that I am proposing on Smith's behalf has any hope of succeeding, I must show that the mechanism of social self-distancing is capable of supporting a mode of thought marked by independence but not obstinance, openness but not credulity, humility but not timidity.

Smith has little to say about how to solve this problem, but there is reason to think he would be optimistic. He argues, for instance, that our tendency to sympathize with others is, in part, automatic and, at least in certain cases, unavoidable (*TMS* I.i.1.3); that we *want to* sympathize with others, and are "hurt when we are unable to do so" (*TMS* I.i.2.3); and that nature has "taught

[92] See, for example, Asch (1951) and Asch (1956).
[93] For examples of this phenomenon, see Cialdini, Reno, and Kallgren (1990), Reno, Cialdini, and Kallgren (1993), and Keizer, Lindenberg, and Steg (2008).

man to respect the sentiments and judgments of his brethren" and "feel pleasure in their favourable, and pain in their unfavourable regard" (*TMS* III.2.31 & III.2.6). Smith clearly thinks that people are naturally sensitive to the views of their peers, and that they are unlikely to disregard them completely. At the same time, though, he insists that, sensitive as we are, there is a natural limit to our suggestibility, at least with respect to our moral judgments. In contrast to the "principles of imagination, upon which our sense of beauty depends," and which "may easily be altered by habit and education," he writes that our "sentiments of moral approbation and disapprobation, are founded on the strongest and most vigorous passions of human nature" (*TMS* V.2.1). Though "they may be somewhat warpt," they "cannot be entirely perverted." We would never internalize a radically objectionable moral judgment, Smith thinks, merely because it is voiced by those with whom we interact. Our natural desire to sympathize, combined with both our respect for the views of others and feelings of accountability, will ensure that we remain open to other people's judgments; but the limitations on our capacity to actively imagine deviant evaluative sentiments will, he thinks, protect us from credulity.[94]

Is Smith right? The issue is complex, but recent empirical research once again provides (qualified) grounds for optimism. Begin with the psychological mechanism of self-distancing, ignoring for the moment the effects of social interaction. I have already argued that self-distancing can enhance our real-time self-evaluative judgments by moderating our reactivity and enhancing our cooperative tendencies. There is evidence that it can do more. Kross and Grossmann (2011) found that cueing people to think about "personally meaningful issues" from a self-distanced perspective causes them to reason more abstractly, to recognize the limits of their knowledge, and to open their minds to diverse points of view. Similarly, Bremner,

[94] For a discussion of the difference between "active" and "passive" imagination, and its relevance to Smith's account, see Chapter 3.

Goldberg, and Kross (2013) found that triggering people to self-distance lessens their susceptibility to at least one form of cognitive bias, and possibly others as well.[95] In addition to protecting us from "hot" emotional outbursts, these findings suggest that self-distancing can help us to overcome two of the obstacles that threaten to prevent us from utilizing the information afforded by social interaction: our tendency to dismiss perspectives that differ from our own, and our susceptibility to self-serving cognitive biases. Self-distancing cannot, on its own, provide us with new information or novel perspectives on the basis of which to self-evaluate. But by improving our reasoning and opening our minds to other points of view, it can help to ensure that we make better use of such things if and when they become available.

What about the opposite worry: that we will simply come to adopt the views of those with whom we converse, even when they conflict with our own? The danger is real, but the outcome is by no means inevitable. According to Nemeth (1995), our response to an opposing point of view depends, at least in part, on whether it is held by the majority of our peers. Whereas majority disagreement tends to stimulate *convergent* cognitive processes, minority disagreement tends to stimulate *divergent* processes. That is to say, when exposed to a dissenting argument offered by a majority of their peers, people tend to abandon their position, adopt the majority view, and block out all other perspectives.[96] When exposed to a minority dissent, however, people tend neither to accept nor to dismiss it with prejudice. Instead, they search for more information and examine the disputed claim from a range of perspectives (including, but not limited to, that of the minority). Moreover, Nemeth found that exposure to minority dissent

[95] The study found that self-distancing reduces susceptibility to the *Fundamental Attribution Error* or *Correspondence Bias*: the "tendency to erroneously attribute the causes of people's behavior to their personalities and attitudes rather than the situation" (Bremner et al., 2013, p. 3). Though they concede that additional research is required, the authors expect self-distancing to reduce other biases as well – in particular, "any bias that is easier for an outsider to notice" (p. 13).

[96] These results are in line with those presented in Asch (1951) and Asch (1956), mentioned above.

increases individual creativity and enhances deliberative quality. People who encounter unconventional dissenting opinions tend not only to be better *informed*; they actually *think* better.

Other experiments have yielded similar findings. In a series of experiments, Nemeth and her colleagues found that, when compared to people who encounter no dissenting views, people who encounter dissent tend to search for more information (Nemeth & Rogers, 1996), think more carefully and creatively about a wider range of possibilities (Nemeth & Wachtler, 1983), and exhibit more signs of original thought (Nemeth & Kwan, 1985). More recently, Nemeth, Connell, Rogers, and Brown (2001) found that people who are exposed to *authentic* dissenting arguments give more serious consideration to opposing points of view, are more willing to alter their position, and exhibit more original thought than people who are exposed only to *inauthentic* dissent (i.e., a devil's advocate) or to no dissent at all. Whereas people in the latter category often display a systematic bias in favor of their own view – employing their creative and intellectual capacities as defensive measures only: blocking, rather than exploring, perspectives other than their own – individuals exposed to authentic dissent employ their capacities more productively. Like self-distancing, exposure to authentic disagreement can reduce "cognitive bolstering".

Interestingly, there is also evidence that exposing individuals to different sorts of *people* can have a similar effect to the one produced by exposing them to people with different sorts of *views*. For example, Sommers (2006) found that racially diverse groups tend to deliberate longer and consider a wider range of information than homogeneous groups, and that the White participants in diverse groups introduce more facts, make fewer factual errors, and are more amenable to discussing race-related issues than their counterparts in all-White groups. Even the *expectation* of deliberating with a diverse group can affect the way individuals think: there is evidence that White individuals process information more systematically and exhibit better

comprehension of topical background readings when they expect to discuss a race-relevant topic within a racially diverse group than when they expect to do so in an all-White group (Sommers, 2006; Sommers, Warp, & Mahoney, 2008). Similarly, Antonio et al. (2004) found that the presence of even a single Black collaborator in a group of White participants can improve the latter's quality of thought, stimulating greater differentiation and integration of perspectives. Exposing people to a diverse group of other people does not merely expose them to new information; it can make them better consumers of information. *It can help people think.*

Smith recognized early on that our spectators' identities affect the way we see ourselves. He writes that the presence of a "common acquaintance" composes us "more than that of a friend" and "that of an assembly of strangers still more than that of an acquaintance" (*TMS* I.i.4.9). It is the spectator "from whom we can expect the least sympathy and indulgence," he claims, that we "learn the most complete lesson of self-command" (*TMS* III.3.38).

> Are you in adversity? Do no 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. (*TMS* III.3.39).

We self-evaluate properly, Smith claims, only when our spectators are informed and impartial, and our judgments are "never so apt to be corrupted, as when the indulgent and partial spectator is at hand" (*TMS* III.3.41). That our self-evaluations are, at least in part, a function of our actual spectators' identities is central to Smith's account.

Why is this? Why do our spectators' identities affect us so? And why does Smith think that we can only evaluate ourselves properly when we are in the presence of informed and impartial spectators? Feeling accountable to our spectators for our conduct, their presence motivates us to

self-distance in a particular way: to consider ourselves not merely from *a* spectator's perspective, but from *their* perspectives. The "indulgent and partial" spectator is too close to us, too similar in thought and bias, to change the way we view ourselves. It is only an impartial spectator who, by holding us accountable for our conduct, can force us to consider ourselves from a *different* point of view: from the perspective of one who sees more broadly and feels with more circumspection, who is less vulnerable to our particular biases, who views us as merely "one of the multitude, in no respects better than any other" (*TMS* III.3.4). The impartiality of our spectators matters, Smith thinks, because it determines the perspective from which we are held accountable, and from which we ultimately judge our own conduct.

My point in this section is to suggest that Smith's claim can be generalized. Our real spectators will differ from us in more than just their degree of impartiality. They will differ in their beliefs, in their personal experiences, in their identities; and these differences matter. In addition to exposing us to new information and ways of thinking, interacting with, and attempting to view ourselves from the perspectives of, a diverse group of people can trigger us to search for information, make fewer factual errors, and process material more systematically and effectively; it can lead us to examine issues from a wider range of perspectives and demonstrate a greater capacity for differentiation and integration of diverse points of view; it can make us more open-minded, motivating us to give more serious consideration to opposing points of view and to display a greater willingness to shift our position; and it can improve the overall quality of our thought, reducing our susceptibility to cognitive bias and enhancing our creativity. Interacting with a diverse group of people can enhance our self-directed critical thought in precisely that ways that effective self-evaluation requires.

5. Conclusions

According to some, Smith thinks we self-evaluate by internalizing our society's dominant social norms. Early in Chapter 4, I mentioned four problems with this reading. First, by depicting our self-evaluative judgments as internalized social norms, the reading both denies us the capacity to criticize our own society and paints an unacceptably passive picture of the self-evaluative process – as if it were something that *happens to* us, rather than something that we *do*. Second, the reading fails to capture an important egalitarian theme embedded within Smith's moral psychology: his normative conception of society as a moral community of independent and mutually accountable equals. Third, it fails to explain Smith's claim that engaging in self-evaluation can cause us to see ourselves as equal members of such a moral community. And, finally, the reading leaves no room for Smith's distinctive notion of sympathy to play a role in the self-evaluative process.

My reading is different. I have claimed that, according to Smith, we are motivated to self-evaluate by feelings of accountability to, and respect for, our peers. We evaluate ourselves not by passively internalizing the judgments made of us by informed and impartial members of our society, but by engaging in a process of *social self-distancing*: by subjecting ourselves to the scrutiny of our peers, listening to their arguments, and attempting to understand their individual points of view. As I read it, Smith's proposal is not one of *passive* internalization, but of *active* engagement with perspectives other than our own. We evaluate our own conduct by attempting to sympathize with our spectators' assessments of us: by weighing their arguments, scrutinizing their claims, and judging for ourselves whether their conclusions about us are as they *ought* to be. Smith's principles of equality, accountability, and independence are thus built into my

interpretation from the very beginning, as is the role of his distinctive account of sympathetic experience.

Moreover, though my interpretation of Smith's account concedes that our peers' judgments affect us, it denies that our peers' judgments determine the content of our self-evaluative conclusions. My interpretation is not relativistic in this sense, nor does it commit Smith to an equally unattractive form of individual relativism. On the contrary, given the right social circumstances, I have argued that engaging in social self-distancing can significantly enhance our critical perspective by exposing us to new information and improving the way we process it, motivating us to examine issues from a wider range of perspectives (including those that differ from our own and that of our society), reducing our susceptibility to various forms of cognitive bias, and generally enhancing the quality, creativity, and originality of our evaluative thought. Our self-evaluative judgments can be independent of both our society's dominant social norms and our own default views.

Finally, I have argued that the process of self-evaluation can affect us in something like the way Smith claims: making us more cooperative and humble, more open-minded and respectful of other people's views. Given the right social conditions – a diverse and integrated society in which ideas are freely exchanged – the act of self-evaluation can move us at least part of the way towards endorsing the very principles to which Smith's general account is implicitly committed. Properly developed, I have argued that Smith's sympathy-based moral psychology succeeds in both expressing and supporting his abstract egalitarian ideals.

CHAPTER 6. CONCLUDING REMARKS

Return to the argument with which we began in Chapter 1: Hare's argument for utilitarianism. Moral language, Hare claims, is prescriptive and universal: if I were to say now that I ought to do a certain thing to a certain person, I would be expressing my all-things-considered preference that the thing be done, where this all-things-considered preference is produced by weighing and aggregating my actual preferences regarding the situation, as well as the preferences that I now have for the hypothetical scenario in which I am in the position of, and have the same personal characteristics and motivational states as, the person to whom I claim I ought to do the certain thing. Considering whether or not to play my trumpet, for example, Hare claims that I am bound by the logical properties of moral words to weigh my actual preference that I play against my conditional preference that, were I in my neighbor's situation with his aversion to the sound of the trumpet, I not play. I ought to play if and only if the former outweighs the latter.

Moreover, as we have seen, Hare places a constraint on the conditional preferences that an ideally deliberating moral agent can have. According to Hare, if I am deliberating correctly, then the conditional preferences that I now have for the hypothetical scenario in which I am in another person's shoes must perfectly correspond to that person's actual (informed and rational) preferences. For he insists that critical moral thought requires "full representation" of other people's circumstances, and full representation necessarily leads one to acquire a set of conditional preferences that track the preferences of those whose circumstances one fully represents. Combining this claim with the aforementioned properties of moral language yields

Hare's conclusion: namely, that ideal moral deliberation requires one to proceed by weighing one's own preferences against the preferences of those affected by one's actions. As Hare sees it, the moral requirement of perspective taking leads to a form of preference utilitarianism.

Smith's story is different. He begins not with an analysis of moral language, nor even with a discussion of the importance of considering other people's perspectives when deciding what to do, but with a psychological account designed to explain our evaluations of other people's feelings and behavior. We imagine ourselves in other people's shoes, complete with their personal characteristics, and attempt to simulate a response to their circumstances as if we were just like them. We try, in other words, to imagine feeling as they do. In contrast to Hare, though, Smith denies that we always succeed in reproducing other people's feelings – even when we imagine every feature of their circumstances, and every aspect of their personality, correctly. For, according to Smith, the process of imaginative perspective taking is evaluative. Imagining myself in your shoes, I *may* feel as you do, but I *may not*. If the former, I will approve of your feelings as proper or well-suited to your circumstances. But if the latter, I will disapprove of your feelings as improper or poorly suited to your circumstances.

That, according to Smith, is how our moral judgments begin. We begin by evaluating others and holding them accountable for what they have done. Very soon, though, we realize that the very people whom we evaluate are doing the same thing to us. Finding ourselves the object of other people's approval and disapproval, praise and blame, gratitude and resentment, we begin to ask ourselves whether, and to what extent, we deserve to be so judged. Driven by feelings of accountability, triggered by the presence of those to whom we are accountable, we set out to determine how we *ought* to be judged. We "suppose ourselves the spectators of our own behaviour, and endeavour to imagine what effect it would, in this light, produce upon us" (*TMS*

III.1.5). We begin to evaluate our own conduct by engaging in a process of self-distance self-reflection.

That is as far as I developed the accountability theme in Chapter 5. I claimed only that we hold others accountable for their feelings and behavior, and feel accountable to them in return. But we can push this further. Though I have argued that Smith denies that our self-evaluative judgments are dictated by other people's assessments of us (see Chapter 4), he nevertheless concedes that we care what other people think of us. We are "pleased" when they approve of our conduct and "hurt when they disapprove" of it (*TMS* III.2.31). Moreover, in addition to being affected by their assessments of our conduct, Smith notes that we hold others accountable for their judgments whenever we think them unfair. That is to say, in addition to evaluating and holding people accountable for their *feelings and behavior* (and feeling accountable to them in return), Smith thinks that we evaluate and hold people accountable for their *evaluations of our* feelings and behavior. We stand in a relation of *multi-level* mutual accountability: we take ourselves to be accountable to one another not only for our feelings and behavior, but for our evaluative judgments as well.

Consider, for instance, the case of unmerited blame or reproach. As we learned in Chapter 4, Smith insists that unmerited praise is, to anyone but the "most frivolous and superficial of mankind," irrelevant (*TMS* III.2.11). The case of unmerited blame, though, is, Smith thinks, quite different. To be unfairly blamed or punished can mortify "very severely even men of more than ordinary constancy" (*TMS* III.2.11). An innocent man

> of more than ordinary constancy, is often, not only shocked, but most severely mortified by the serious, though false, imputation of a crime; especially when that imputation happens unfortunately to be supported by some circumstances which give it an air of probability. (*TMS* III.2.11)

He is "humbled" to learn that people "think so meanly of his character as to suppose him capable of being guilty of it." And, though confident in his own innocence, the criticisms made by others throw "a shadow of disgrace and dishonor upon his character." This can cause significant harm – something worth thinking about and protecting against:

> As, of all the external misfortunes which can affect an innocent man immediately and directly, *the undeserved loss of reputation is certainly the greatest*; so a considerable degree of sensibility to whatever can bring on so great a calamity, does not always appear ungraceful or disagreeable. (*TMS* III.3.19, emphasis added)

Indeed, Smith goes further. He writes that an individual who has been unfairly blamed for some wrongdoing will be "tormented by his own indignation at the *injustice* which has been done to him" (*TMS* III.2.11, emphasis added). And this injustice, Smith argues, can itself be the proper object of resentment:

> We often esteem a young man the more, when he resents, though with some degree of violence, any unjust reproach that may have been thrown upon his character or his honour. (*TMS* III.3.19)

To be the object of unmerited blame or reproach is, Smith believes, to be the victim of an injustice. It is something to care about, to protect against; even something to resent.

It is not only unmerited blame or reproach, moreover, that Smith thinks capable of qualifying as an injustice. Even the *absence* of merited praise or approbation can produce an injustice. Though Smith insists that to "desire, or even to accept of praise, where no praise is due," is an indication of "the most contemptible vanity," he thinks it perfectly reasonable to desire praise when it is deserved (*TMS* III.2.8).

> To desire it where it is really due, is to desire no more than that *a most essential act of justice should be done to us*. The love of just fame, of true glory, even for

> its own sake, and independent of any advantage which he can derive from it, is not unworthy even of a wise man. (*TMS* III.2.8, emphasis added)

And just as we wish to be praised when we think that praise is due, Smith notes that we often wish for approval of (if not praise for) our feelings of grief, indignation, resentment, and the like, when we think that such approval is due. We may even become resentful when a companion fails to sympathize with – and thereby validate – what we take to be our own perfectly proper feelings:

> [I]f you have either no fellow-feeling for the misfortunes I have met, 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* I.i.4.5)

Just as my conduct matters to you, your sympathy and approval matters to me. You hold me accountable if you think that I have acted inappropriately, and I hold you accountable if I think that you have failed to judge me fairly. In particular, I hold you accountable if I think that you have failed to adequately acknowledge and consider my circumstances *from my point of view*. Using the language of Chapter 3, I hold you accountable for any *cognitive* errors committed in your simulation and evaluation of my circumstances and conduct.[97]

Where does this lead? Imagining other people's circumstances from their point of view, we approve or disapprove of their conduct insofar as we find ourselves capable or incapable of sympathizing with them, and we hold them accountable for their conduct when we disapprove. Realizing that they are doing the same thing to us, we set out to determine whether their assessments are justified. Imagining ourselves the spectators of our own conduct, we evaluate

[97] See Section 3 of Chapter 3 for a discussion of what I there call "*cognitive sympathetic failures*".

ourselves by evaluating our actual spectators' assessments of our conduct: approving or disapproving of – and, consequently, adopting or rejecting – their assessments insofar as we find ourselves capable or incapable of sympathizing with their evaluative feelings. And we hold them accountable for their judgments when we think them unfair. The pattern then repeats itself. Realizing that the very people whose evaluations we are evaluating are at the same time evaluating our evaluations of them, we set out to ensure that we are evaluating them fairly – just as we demand they do for us. The result is a complex web of reciprocal simulations: I evaluate your conduct by simulating a response to your circumstances, prompting you to simulate my simulation of your circumstances, leading me to simulate your simulation of my simulation of your circumstances, and so forth.[98]

For Smith, our evaluative judgments occur neither in a social vacuum nor at a single point in time. They are produced by a set of complex iterative social processes occurring against – and expressive of – a background norm of equality, respect, and multi-level mutual accountability. Observing your behavior, and finding myself incapable of sympathizing with it, I judge your conduct improper and tell you so. The process does not end there, though. You may reject my disapproval. Perhaps you think that I failed to adequately account for your circumstances or unique personal characteristics, or perhaps that I failed to recognize *the way* in which your circumstances could fittingly elicit your response. Whatever the nature of my alleged error, you respond by holding me accountable for (what you take to be) my misplaced disapproval. You try to clarify your way of seeing things, to help me feel as you do. I reach a (reasonably) firm

[98] Compare Darwall (2006, p. 44): "When I see another as a 'you', I see her as having the same relation reciprocally to me. I relate to her as relating reciprocally to me. What does this involve? Partly, it involves a rich set of higher-order attitudes: I am aware of her awareness of me, aware of her awareness of my awareness of her, aware of her awareness of my awareness of her awareness of me, and so on." Though I make no attempt to explore the connections here, there is considerable overlap between the ideas sketched in these closing remarks and Darwall's work.

conclusion that your conduct is in fact improper only if I find myself incapable of sympathizing with you, *even after* engaging in this type of back-and-forth – after listening to your side of the story, attempting to understand your way of seeing things, and so forth. I judge your conduct not only improper but unjust if, in addition to finding myself incapable of sympathizing with you, I find myself sympathizing with feelings of resentment toward you – again, after listening to your side of the story, attempting to understand and sympathize with your motives, and so forth.

To self-evaluate, we simply reverse roles. I do something and observe that you disapprove. Feeling accountable for my conduct, I imagine viewing myself from your point of view – to see for myself whether your disapproval is warranted. If I sympathize with your disapproval, I judge my conduct improper. If not, I reject your disapproval, judge my conduct proper, and attempt to change your mind – say, by providing you with additional information about my circumstances or reasons for feeling or behaving as I did. I do what I can to get you to see my point of view, to help you feel as I do, and I hold you accountable insofar as you show yourself unwilling to do so. I reach a (reasonably) firm conclusion that my conduct is in fact proper only if I find myself incapable of sympathizing with your disapproval, *even after* engaging in this type of back-and-forth.

Hare's model of ideal moral thought is fundamentally aggregative: we decide what to do by weighing our preferences against those of others, deliberating based on strength of preference alone. His account could be said to involve norms of equality and respect for individuals only in the sense that it requires everyone to assign equal weight to everyone else's preferences in their own moral deliberations. I respect and treat you as an equal, in this sense, as long as I fully represent to myself your circumstances (including your personal characteristics and motivational states), acquire a set of conditional preferences that mirror your actual (informed and rational)

preferences, and then incorporate those conditional preferences into my decision-making process.

Smith's account is not aggregative in this way, and it paints a very different picture of the nature of perspective taking and its role in moral thought. I consider your point of view, according to Smith, not because I am necessarily required to assign equal weight to your preferences in my own deliberations, but because you demand it of me – and hold me accountable if I do not. And I do the same to you. We imagine ourselves in each other's shoes not to identify and aggregate our respective preferences, but in an effort to find common ground: to construct a shared perspective from which we can collectively judge the propriety or impropriety, justice or injustice, of our (and other people's) feelings and behavior. Driven by feelings of mutual accountability, we attempt to understand and imaginatively reproduce each other's evaluative judgments as part of our search for an evaluative perspective acceptable to each participant in a community of independent and mutually accountable equals.

It is ironic that Smith is often portrayed as an early utilitarian – and particularly so that he is portrayed as such by Rawls. Rawls was committed to developing an alternative to utilitarianism, grounded in a normative conception of society as a "fair system of cooperation between free and equal citizens" (1993, p. 22). If I am correct, Smith's account of sympathy and its role in moral evaluation is much more similar to Rawls's own work than to the work of the "great utilitarians" with whom Rawls groups him (1971, p. xvii). Smith's account of sympathy and the moral sentiments expresses and supports a normative ideal very much like the one endorsed by Rawls: that of a moral community of independent and mutually accountable equals, striving to construct and employ a set of evaluative standards that is respectful of and endorsable by all reasonable

points of view. Smith's is a moral psychology well suited to support a contractualist moral theory.

WORKS CITED

Andreoni, J., & Petri, R. (2004). Public Goods Experiments without Confidentiality: A Glimpse into Fund-Raising. *Journal of Public Economics, 88*(7-8), 1605-1623.

Antonio, A. L., Chang, M. J., Hakuta, K., Kenny, D. A., Levin, S., & Milem, J. F. (2004). Effects of Racial Diversity on Complex Thinking in College Students. *Psychological Science, 15*(8), 507-510.

Árdal, P. (1966). *Passion and Value in Hume's Treatise*. Edinburgh: Edinburgh University Press.

Aronson, E. (2011). *The Social Animal*. New York, NY: Worth Publishers.

Asch, S. E. (1951). Effects of Group Pressure Upon the Modification and Distortion of Judgment. In H. S. Guetzkow (Ed.), *Groups, Leadership, and Men: Research in Human Relations* (pp. 117-190). Pittsburgh: Carnegie.

Asch, S. E. (1956). Studies of Independence and Conformity: A Minority of One Against a Unanimous Majority. *Psychological Monographs: General and Applied, 70*(9).

Ayduk, O., & Kross, E. (2008). Enhancing the Pace of Recovery: Self-Distanced Analysis of Negative Experiences Reduces Blood Pressure Reactivity. *Psychological Science, 19*, 229-231.

Ayduk, O., & Kross, E. (2010). From a Distance: Implications of Spontaneous Self-Distancing for Adaptive Self-Regulation. *Journal of Personality and Social Psychology, 98*(5), 809-829.

Baier, A. C., & Waldow, A. (2008). A Conversation between Annette Baier and Anik Waldow about Hume's Account of Sympathy. *Hume Studies, 34*(1), 61-87.

Barry, B. (1995). *Justice as Impartiality*. Oxford: Oxford University Press.

Barry, B. (2001). *Culture and Equality: An Egalitarian Critique of Multiculturalism*. Cambridge, MA: Harvard University Press.

Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of Being Watched Enhance Cooperation in a Real-World Setting. *Biology Letters, 2*, 412-414.

Baugh, D. A. (1983). Poverty, Protestantism, and Political Economy. In S. B. Baster (Ed.), *England's Rise to Greatness, 1660-1763*. Berkeley, CA: University of California Press.

Braithwaite, R. B. (1955). *Theory of Games as a Tool for the Moral Philosopher*. Cambridge: Cambridge University Press.

Brandt, R. B. (1988). Act-Utilitarianism and Metaethics. In D. Seanor & N. Fotion (Eds.), *Hare and Critics: Essays on Moral Thinking*. Oxford: Clarendon Press.

Bremner, R. H., Goldberg, A., & Kross, E. (2013). Seeing the Log in Your Own Eye: Self-Distancing Improves Judgmental Accuracy. *University of Michigan Working Paper*.

Bricke, J. (1996). *Mind and Morality: An Examination of Hume's Moral Psychology*. Oxford: Oxford University Press.

Brown, T. (1846). *Lectures on Ethics*. Edinburgh: William Tait.

Burnham, T. C. (2003). Engineering Altruism: A Theoretical and Experimental Investigation of Anonymity and Gift Giving. *Journal of Economic Behavior & Organization, 50*(133-144).

Burnham, T. C., & Hare, B. (2007). Engineering Human Cooperation: Does Involuntary Neural Activation Increase Public Goods Contributions? *Human Nature, 18*, 88-108.

Camerer, C., Loewenstein, G., & Weber, M. (1989). The Curse of Knowledge in Economic Settings: An Experimental Analysis. *Journal of Political Economy, 97*(5), 1232-1254.

Campbell, T. D. (1971). *Adam Smith's Science of Morals*. London: George Allen & Unwin Ltd.

Campbell, T. D., & Ross, I. S. (1981). The Utilitarianism of Adam Smith's Policy Advice. *Journal of the History of Ideas, 42*(1), 73-92.

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A Focus Theory of Normative Conduct: Recycling the Concepts of NOrms to Reduce Littering in Public Places. *Journal of Personality and Social Psychology, 58*, 1015-1029.

Darwall, S. (1999). Sympathetic Liberalism: Recent Work on Adam Smith. *Philosophy and Public Affairs, 28*(2), 139-164.

Darwall, S. (2002). *Welfare and Rational Care*. Princeton, NJ: Princeton University Press.

Darwall, S. (2004). Equal Dignity in Adam Smith. *Adam Smith Review, 1*, 129-134.

Darwall, S. (2006). *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge, MA: Harvard University Press.

Dawes, R. M., McTavish, J., & Shaklee, H. (1977). Behavior, Communication, and Assumptions about Other People's Behavior in a Commons Dilemma Situation. *Journal of Personality and Social Psychology, 35*(1), 1-11.

Driver, J. (2008). Imaginative Resistance and Psychological Necessity. *Social Philosophy and Policy, 25*(1), 301-313.

Duncan, B. L. (1976). Differential Social Perception and Attribution of Intergroup Violence: Testing the Lower Limits of Stereotyping of Blacks. *Journal of Personality and Social Psychology, 34*(4), 590-598.

Dworkin, R. (1977). *Taking Rights Seriously*. Cambridge, MA: Harvard University Press.

Edwards, K., & Smith, E. E. (1996). A Disconfirmation Bias in the Evaluation of Arguments. *Journal of Personality and Social Psychology, 71*, 5-24.

Ernest-Jones, M., Nettle, D., & Bateson, M. (2011). Effects of Eye Images on Everyday Cooperative Behavior: A Field Experiment. *Evolution and Human Behavior, 32*, 172-178.

Farrer, J. A. (1881). *Adam Smith*. New York: G.P. Putnam's Sons.

Ferh, E., & Schneider, F. (2012). Eyes are on Us, but Nobody Cares: Are Eye Cues Relevant for Strong Reciprocity? *Proceedings of the Royal Society B, 277*, 1315-1323.

Firth, R. (1952). Ethical Absolutism and the Ideal Observer. *Philosophy and Phenomenological Research, 12*(3), 317-345.

Fleischacker, S. (1999). *A Third Concept of Liberty: Judgment and Freedom in Kant and Adam Smith*. Princeton, NJ: Princeton University Press.

Fleischacker, S. (2004a). *On Adam Smith's Wealth of Nations: A Philosophical Companion*. Princeton, NJ: Princeton University Press.

Fleischacker, S. (2004b). *A Short History of Distributive Justice*. Cambridge, MA: Harvard University Press.

Fleischacker, S. (2011a). Adam Smith and Cultural Relativism. *Erasmus Journal for Philosophy and Economics, 4*(2), 20-41.

Fleischacker, S. (2011b). True to Ourselves? Adam Smith on Self-Deceit. *Adam Smith Review, 6*, 75-92.

Frankfurt, H. (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy, 68*(1), 5-20.

Frierson, P. R. (2006a). Adam Smith and the Possibility of Sympathy with Nature. *Pacific Philosophical Quarterly, 87*, 442-480.

Frierson, P. R. (2006b). Applying Adam Smith: A Step towards Smithian environmental virtue ethics. In L. Montes & E. Schliesser (Eds.), *New Voices on Adam Smith* (pp. 140-167). New York, NY: Routledge.

Gibbard, A. (1988). Hare's Analysis of 'Ought' and its Implications. In D. Seanor & N. Fotion (Eds.), *Hare and Critics: Essays on Moral Thinking* (pp. 57-72). Oxford: Clarendon Press.

Gibbard, A. (1990). *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge, MA: Harvard University Press.

Gibbard, A. (2005a). Angemessenheit und Mittelmas: Wie Gefuhle und Handlungen aufeinander abgestimmt werden. In C. Fricke & H.-P. Schutt (Eds.), *Adam Smith als Moralphilosoph* (pp. 277-303). Berlin: de Gruyter.

Gibbard, A. (2005b). Truth and Correct Belief. *Philosophical Issues, 15*, 338-350.

Goldie, P. (2000). *The Emotions: A Philosophical Exploration*. Oxford: Oxford University Press.

Goldman, A. (1995a). Interpretation Psychologized. In M. Davies & T. Stone (Eds.), *Folk Psychology: The Theory of Mind Debate*. Oxford: Flackwell Publishers Ltd.

Goldman, A. (1995b). Simulation and Interpersonal Utility. *Ethics, 105*(4), 709-726.

Goldman, A. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford: Oxford University Press.

Gopnik, A., & Astington, J. W. (1988). Children's Understanding of Representational Change and Its Relation to the Understanding of False Belief and the Appearance-Reality Distinction. *Child Development, 59*(1), 26-37.

Gordon, R. (1995a). Folk Psychology as Simulation. In M. Davies & T. Stone (Eds.), *Folk Psychology: The Theory of Mind Debate*. Oxford: Blackwell Publishers Ltd.

Gordon, R. (1995b). Sympathy, Simulation, and the Impartial Spectator. *Ethics, 105*(4), 727-742.

Griswold, C. L. (1999). *Adam Smith and the Virtues of Enlightenment*. Cambridge: Cambridge University Press.

Haidt, J. (2012). *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. New York: Pantheon Books.

Haley, K. J., & Fessler, D. M. T. (2005). Nobody's Watching? Subtle Cues Affect Generosity in an Anonymous Economic Game. *Evolution and Human Behavior, 26*, 245-256.

Hans, V. P., & Vidmar, N. (1982). Jury Selection. In N. L. Kerr & R. M. Bray (Eds.), *The Psychology of the Courtroom* (pp. 39-82). New York: Academic Press.

Hare, R. M. (1963). *Freedom and Reason*. Oxford: Oxford University Press.

Hare, R. M. (1981). *Moral Thinking: Its Levels, Method, and Point*. Oxford: Oxford University Press.

Hare, R. M. (1988). Comments. In D. Seanor & N. Fotion (Eds.), *Hare and Critics: Essays on Moral Thinking*. Oxford: Oxford University Press.

Harman, G. (2000). Moral Agent and Impartial Spectator *Explaining Value*. Oxford: Clarendon Press.

Harsanyi, J. (1953). Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking. *The Journal of Political Economy, 61*(5), 434-435.

Harsanyi, J. (1977). *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. Cambridge: Cambridge University Press.

Harsanyi, J. (1982). Morality and the Theory of Rational Behavior. In A. Sen & B. Williams (Eds.), *Utilitarianism and Beyond* (pp. 39-62). Cambridge: Cambridge University Press.

Hoffman, L. R., & Maier, N. R. F. (1961). Quality and Acceptance of Problem Solutions by Members of Homogeneous and Heterogenous Groups. *Journal of Abnormal and Social Psychology, 62*, 401-407.

Hoffman, M. (2000). *Empathy and Moral Development: Implications for Caring and Justice*. Cambridge: Cambridge University Press.

Hugenberg, K., & Bodenhausen, G. V. (2003). Facial Prejudice: Implicit Prejudice and the Perception of Facial Threat. *Psychological Science, 14*(6), 640-643.

Hutcheson, F. (2004 [1725]). *An Inquiry into the Original of Our Ideas of Beauty and Virtue*. Indianapolis: Liberty Fund.

Jones, E. E., & Kohler, R. (1958). The effects of plausibility on the learning of controversial statements. *The Journal of Abnormal and Social Psychology, 57*(3), 315-320.

Jouffroy, T. S. (1841). *Introduction to Ethics, including a critical survey of moral systems* (W. H. Channing, Trans. Vol. II). Boston: Hilliard, Gray, and Company.

Kant, I. (1997 [1788]). *Critique of Practical Reason* (M. Gregor, Trans.): Cambridge University Press.

Kant, I. (2002 [1785]). *Groundwork for the Metaphysics of Morals* (A. Wood, Trans.): Yale University Press.

Keizer, K., Lindenberg, S., & Steg, L. (2008). The Spreading of Disorder. *Science, 322*(5908), 1681-1685.

Kross, E. (2009). When the Self Becomes Other: Toward an Integrative Understanding of the Processes of Distinguishing Adaptive Self-Reflection from Rumination. *Annals of the New York Academy of Sciences, 1167*, 35-40.

Kross, E., & Ayduk, O. (2008). Facilitating Adaptive Emotional Analysis: Distinguishing Distanced-Analysis of Depressive Experiences from Immersed-Analysis and Distraction. *Personality and Social Psychology Bulletin, 34*(7), 924-938.

Kross, E., & Ayduk, O. (2011). Making Meaning out of Negative Experiences by Self-Distancing. *Current Directions in Psychological Science, 20*(3), 187-191.

Kross, E., Ayduk, O., & Mischel, W. (2005). When Asking "Why" Does Not Hurt Distinguishing Rumination From Reflective Processing of Negative Emotions. *Psychological Science, 16*(9), 709-715.

Kross, E., & Grossmann, I. (2011). Boosting Wisdom: Distance from the Self Enhances Wise Reasoning, Attitudes, and Behavior. *Journal of Experimental Social Psychology: General.*

Krueger, J. (2000). The Projective Perception of the Social World: A Building Block of Social Comparison Processes. In J. Suls & L. Wheeler (Eds.), *Handbook of Social Comparison: Theory and Research*. New York, NY: Kluwer Academic Publishing.

Kurzban, R. (2001). The Social Psychophysics of Cooperation: Nonverbal Communication in a Public Goods Game. *Journal of Nonverbal Behavior, 25*(4), 241-259.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37*(11), 2098-2109.

Mackenzie, C. (2006). Imagining Other Lives. *Philosophical Papers, 35*(3), 293-325.

Mackie, J. L. (1977). *Ethics: Inventing Right and Wrong*. London: Penguin Books.

Mannix, E., & Neale, M. A. (2005). What Differences Make a Difference? The Promise and Reality of Diverse Teams in Organizations. *Psychological Science in the Public Interest, 6*, 31-55.

Metcalfe, J., & Mischel, W. (1999). A Hot/Cool-System Analysis of Delay Gratification: Dynamics of Willpower. *Psychological Review, 106*(1), 3-19.

Mifune, N., Hashimoto, H., & Yamagishi, T. (2010). Altuism Toward In-Group Members as a Reputation Mechanism. *Evolution and Human Behavior, 31*, 109-117.

Mischkowski, D., Kross, E., & Bushman, B. J. (2012). Flies on the Wall are Less Aggressive: Self-Distancing "In the Heat of the Moment" Reduces Aggressive Thoughts, Angry Feelings and Aggressive Behavior. *Journal of Experimental Social Psychology, 48*, 1187-1191.

Moran, R. (1994). The Expression of Feeling in Imagination. *The Philosophical Review, 103*(1), 75-106.

Mossner, E. C., & Ross, I. S. (1977). *The Correspondence of Adam Smith*. Oxford: Oxford University Press.

Nanay, B. (2010). Adam Smith's Concept of Sympathy and Its Contemporary Interpretations. *Adam Smith Review, 5*, 85-105.

Nemeth, C. J. (1995). Dissent as Driving Cognition, Attitudes, and Judgments. *Social Cognition, 13*(3), 273-291.

Nemeth, C. J., Connell, J. B., Rogers, J. D., & Brown, K. S. (2001). Improving Decision Making by Means of Dissent. *Journal of Applied Social Psychology, 31*(1), 48-58.

Nemeth, C. J., & Kwan, J. L. (1985). Originality of Word Associations as a Function of Majority vs. Minority Influence. *Social Psychology Quarterly, 48*(3), 277-282.

Nemeth, C. J., & Rogers, J. D. (1996). Dissent and the Search for Information. *British Journal of Social Psychology, 35*, 67-76.

Nemeth, C. J., & Wachtler, J. (1983). Creative Problem Solving as a Result of Majority vs Minority Influence. *European Journa of Social Psychology, 13*, 45-55.

Nigro, G., & Neisser, U. (1983). Point of View in Personal Memories. *Cognitive Psychology, 15*, 467-482.

Peacocke, C. (1985). Imagination, Experience, and Possibility: A Berkeleian View Defended. In J. Foster & H. Robinson (Eds.), *Essays on Berkeley: A Tercentennial Celebration*. Oxford: Oxford Universitiy Press.

Popper, K. (1962). *The Open Society and Its Enemies: Volume 2*. New York, NY: Routledge.

Raphael, D. D. (1975). The Impartial Spectator. In A. S. Skinner & T. Wilson (Eds.), *Essays on Adam Smith* (pp. 83-99). Oxford: Clarendon Press.

Raphael, D. D. (2007). *The Impartial Spectator: Adam Smith's Moral Philosophy*. Oxford: Clarendon Press.

Rawls, J. (1971). *A Theory of Justice* (Revised ed.). Cambridge, MA: Harvard University Press.

Rawls, J. (1993). *Political Liberalism* (Expanded ed.). New York: Columbia University Press.

Rawls, J. (2001). *Justice as Fairness: A Restatement*. Cambridge, MA: Harvard University Press.

Reno, R. R., Cialdini, R. B., & Kallgren, C. A. (1993). The Trans-Situational INfluence of Social Norms. *Journal of Personality and Social Psychology, 64*, 104-112.

Rigdon, M., Ishii, K., Watabe, M., & Kitayama, S. (2009). Minimal Social Cues in the Dictator Game. *Journal of Economic Psychology, 30*, 358-367.

Ross, L., Greene, D., & House, P. (1977). The "False Consensus Effect": An Egocentric Bias in Social Perception and Attribution Processes. *Journal of Experimental Social Psychology, 13*(3), 279-301.

Rothschild, E. (2001). *Economic Sentiments: Adam Smith, Condorcet, and the Enlightenment*. Cambridge, MA: Harvard University Press.

Sagar, H. A., & Schofield, J. W. (1980). Racial and Behavioral Cues in Black and White Children's Perceptions of Ambiguously Aggressive Acts. *Journal of Personality and Social Psychology, 39*(4), 590-598.

Sayre-McCord, G. (2010). Sentiments and Spectators: Adam Smith's Theory of Moral Judgment. *Adam Smith Review, 5*, 124-144.

Scanlon, T. M. (1982). Contractualism and Utilitarianism. In A. Sen & B. Williams (Eds.), *Utilitarianism and Beyond*. Cambridge: Cambridge University Press.

Scanlon, T. M. (1998). *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

Seanor, D., & Fotion, N. (1988). *Hare and Critics: Essays on Moral Thinking*. Oxford: Clarendon Press.

Sen, A. (1992). *Inequality Reexamined*. Cambridge, MA: Harvard University Press.

Smith, A. (1759). *The Theory of Moral Sentiments* (First ed.). London: A. Millar.

Smith, A. (1761). *The Theory of Moral Sentiments* (Second ed.). London: A. Millar.

Smith, A. (1976 [1790]). *The Theory of Moral Sentiments* (Sixth ed.). Oxford: Oxford University Press.

Sommers, S. R. (2006). On Racial Diversity and Group Decision Making: Identifying Multiple Effects of Racial Composition on Jury Deliberations. *Journal of Personality and Social Psychology, 90*(4), 597-612.

Sommers, S. R., Warp, L. S., & Mahoney, C. C. (2008). Cognitive Effects of Racial Diversity: White Individuals' Information Processing in Heterogeneous Groups. *Journal of Experimental Social Psychology, 44*, 1129-1136.

Stueber, K. R. (2011). Imagination, Empathy, and Moral Deliberation: The Case of Imaginative Resistance. *The Southern Journal of Philosophy, 49, Spindel Supplement*, 156-180.

Van Boven, L., & Loewenstein, G. (2003). Social Projection of Transient Drive States. *Personality and Social Psychology Bulletin, 29*(9), 1159-1168.

Vitz, R. (2004). Sympathy and Benevolence in Hume's Moral Psychology. *Journal of the History of Philosophy, 42*(3), 261-275.

Walton, K. (1990). *Mimesis as Make-Believe: On the Foundations of Representational Arts*. Cambridge, MA: Harvard University Press.

Walton, K. (1994). Morals in Fiction and Fictional Morality. *Proceedings of the Aristotelian Society, Supplementary Volumes, 68*, 27-50.

Walton, K. (1997). Spelunking, Simulation, and Slime: On Being Moved by Fiction. In M. Hjort & S. Laver (Eds.), *Emotion and the Arts*. Oxford: Oxford University Press.

Weatherson, B. (2004). Morality, Fiction, and Possibility. *Philosophers' Imprint, 4*(3).

Williams, B. (1985). *Ethics and the Limits of Philosophy*. Cambridge, MA: Cambridge University Press.

Wollheim, R. (1984). *The Thread of Life*. Cambridge, MA: Harvard University Press.