How to be a Good Sentimentalist

Sveinung Sundfør Sivertsen

UNIVERSITY OF BERGEN

How to be a Good Sentimentalist

Sveinung Sundfør Sivertsen

Thesis for the degree of Philosophiae Doctor (PhD)
at the University of Bergen

Date of defense: 18.10.2019

Table Of Contents

To

Ragnhild

—not because women get too little credit for what they do (although they do),
nor because she has taught me more than anyone else (although she has), nor, finally, because she
has suffered the most for whatever ambition fuelled this dissertation (she hasn't), but because soon,
all societies in which life will be worth living will be ruled by women like her, and then this
dedication may possibly serve me for an entrance pass.

Acknowledgements

This dissertation would not exist but for other people. They all, individually, deserve thanks for the help they have provided, but are too numerous to be mentioned here. I will therefore limit myself to those who have most directly contributed to, and facilitated in, the writing of this dissertation. They present themselves before me as follows: First, there is the Department of Philosophy itself, in the person of Head of Department Reidar Lie and the members of my hiring committee. My first thanks go to them for believing in my project, even though I accidentally handed in a draft version of my proposal riddled with comments, deletions, and insertions. It took vision to look beyond that mess and see what might come of it, and I trust the outcome will not disappoint them.

Second, there is the department, this time in the body of the administrative staff. They have guaranteed the necessary material conditions for bringing this project to fruition, and have always done so with unfailing good humour, regardless of any ignorance or unreasonableness on my part.

Third, the department again, now in the body of the academics who work there. I have benefited tremendously by their accumulated knowledge, incisive criticisms, and diverse peculiarities of character and style. A special thanks goes to the Practical philosophy group for providing me with an intellectual home base. And a very special thanks goes to my peers, the PhD candidates, for their conversation, advice, friendship, and, most of all, their passion. Three deserve a special mention for the constancy of their companionship: David, Johannes, and Preben.

Fourth, there is the (now-defunct) Centre for the Study of Mind in Nature at the University of Oslo. Thanks to the kindness of Grethe Netland, I spent the first six months of my PhD program in an environment as intellectually stimulating as it was extraordinarily amicable, which period helped set the course that I would follow throughout the project. A special thanks goes to Leo Townsend for his instant friendship and invaluable help.

Fifth, there are the psychologists at the Leipzig Research Centre for Early Child Development (LFE). Never have I met a group of people more open to the odd perspectives and uninformed questions of an outsider to their discipline, and I shall count myself lucky if I ever find myself working again with a group of people like those gathered by Daniel Haun at the LFE. My thanks go in particular to Robert Hepach for bringing me into their fold and for our continued

Abstract

How can one be a good person? That, in essence, is the question I ask in this dissertation. More specifically, I ask how we, in general, can best go about the complex and never-ending task of trying to figure out what we should do and then do it. I answer that question in four articles, each dealing with an aspect of the model of morality presented by Adam Smith in *The Theory of Moral Sentiments* (TMS, 2002). The title of the dissertation, 'How to be a good sentimentalist', thus refers to that particular philosophical framework. However, the answers I give should be relevant to any person concerned with how to be a good person.

The first article, *Moral Tuning*, deals with the first part of the question, namely how we can best go about figuring out *what* it is that we should and should not do. Specifically, it deals with the question of whether, and if so how, individuals like you and I can critically reflect upon the norms of our own society, sorting the chaff of merely conventional norms from the wheat of genuinely moral ones. In brief, it is the question of whether we as individuals are autonomous in our relation to the norms of our own society. In answering this question, my co-authors and I argue that Smith's use of musical metaphors in TMS, words like *tone*, *pitch*, and *concord*, can be understood as elements of an analogical model of morality. In contrast to earlier interpretations of Smith's musical metaphors, which have seen music as an aesthetic *object*, we draw on recent developments in musicology to argue that music may also be construed as a *practice*. Construing the source domain of this analogical model as musical practice allows us to construe the target domain also as a practice—as *moral tuning*. This in turn allows us to argue that moral autonomy consists in realising the freedom inherent in the constant need to translate norms into action, and in so doing, to interpret and reinterpret, not only the actions, but the norms themselves. In other words, following the norms of our own society already implies that we are autonomous in relation to them. Being good sentimentalists thus begins with *realising* that we are free to question and reshape the moral standard of our own society.

The second article, *Love Redirected*, deals with the second part of the question, namely how we can best go about *doing* what we already think we should do. This, then, is a question of moral motivation, more specifically a question about the difference between genuinely moral motivation and other kinds of motivation, like a selfish desire for praise. Smith himself argues that we not only desire the actual praise of other people, but to be *worthy* of their praise, to be 'praiseworthy'. The desire to be praiseworthy, the 'love of praiseworthiness' is then the genuinely moral motivation, for it aims at nothing but the satisfaction of having done the right thing. The trouble with Smith's

answer is that he does not adequately connect his claim about praiseworthiness to the rest of his model. This has lead to some confusion in the secondary literature, and in the first part of *Love Redirected*, I seek to end this confusion by combining what Smith says in the various editions of TMS into a coherent argument for how the desire to be praiseworthy comes from redirecting our desire for praise from other people towards the ideal(l) of the 'impartial spectator'. I then go on to show how this reading also fits with modern psychological research on the moral development of children. Finally, I conclude that this redirection of our desire for praise requires not just negative, but positive emotional reinforcement. Therefore, becoming good sentimentalists involves taking pleasure in our moral successes, no less than we are pained at our failures.

The third article, *The Practical Impossibility of Being both Impartial and Well-informed*, makes a first pass at dealing with the idea of the impartial spectator itself, an idea that is central to the answers given in the two first articles. The problem with the impartial spectator is that she is also supposed to be well-informed about those she judges. However, the demands of impartiality and of understanding pull in opposite directions: To be well-informed—to properly understand the situation and character of the person we judge—we must, typically, be sufficiently physically close to that person to see with our own eyes what they are going through. At the same time, this kind of physical closeness tends to entangle us in the kinds of emotional bonds that hinder an impartial evaluation. One might think that this tension could be eased or eliminated by relying less on physical closeness and more on the powers of our imagination to, as Smith frequently puts it, 'bring home to us' the situation of the person we are judging. However, using Construal Level Theory, I argue that merely imagining someone's situation in detail and/or taking their perspective produces a similar effect to physical closeness, and hence that the tension between understanding and impartiality is practically inescapable. To be good sentimentalists, we must therefore recognise our limitations, and give up on the illusion of ever being fully understanding and perfectly impartial at the same time.

The fourth article, *The Partially Impartial Spectator as an Ethical Ideal*, makes a second pass at dealing with the idea, or rather ideal, of the impartial spectator. It begins with the recognition that we frequently fail to be impartial spectators, both of others, and of ourselves. This is especially true in those cases where our views conflict with those of someone else. Building on research detailing the various ways in which cognitive and affective biases impact our perception of the world, our reasoning about our own views, and our (in)ability to resolve disputes with others, I argue that merely *trying harder* to be impartial spectators is liable to backfire, rendering us just as biased as ever, to which is added an unshakeable confidence that we were right all along. Therefore, I go on to argue, we must *try smarter*. Trying smarter, I conclude, involves aiming for something

less ideal, more achievable, and, most importantly, *humbler*, namely to be *partially impartial spectators*.

Being a good sentimentalist thus beings with realising our freedom to interpret, continues in our taking pleasure in our moral successes, pauses at the realisation that we will never be truly impartial spectators, and ends with a commitment to continued improvement under the lodestar of the ideal of the partially impartial spectator.

Part I: Introduction

I. Sentiments, Sympathy, And By-Your-Own-Lights-Normativity[1]

The promise of this dissertation is that you will learn something about how to be a good sentimentalist. But what does this mean? Why should you—or I, or anyone else—want to learn how to be a 'good sentimentalist'? Why 'sentimentalist', and in what sense 'good'? We may begin by substituting 'person' for the peculiar word 'sentimentalist': I want to say something about how you, I, or anyone else can be a *good person*. By 'good person', I mean someone who in general, on balance, most of the time, or just more than they otherwise would, will tend to *do the right thing*. And by doing the right thing, I mean the *moral* thing, the thing that, given the situation and the options before you, is what you *should* do. Finally, by what you should do, I do not mean what you should do in a particular case.[2] Rather, it is a question of how you, in general, can best go about the complex and never-ending task of trying to figure out what you should do—and then do it.

This, then, is a work of *normative* ethics. As such, it enters into a long tradition of philosophical thinking. However, it does so in a particular, and somewhat peculiar, way. First, there is that word, 'sentimentalist'. In the present context, 'sentimentalist' refers to a particular view of morality, a view that emphasises the role our emotions or *sentiments* play in our judgements about what we should and should not do. To be a little more precise, my use of 'sentimentalist' refers to the compound claim, made by Adam Smith in *The Theory of Moral Sentiments* (hereafter simply 'TMS', 2002), that 1) our moral judgements are judgements about what is fitting or appropriate, 2) that our sense of what is appropriate comes from imagining how we would react to the situation in question, and 3) that our approval of how someone else in fact reacts to that situation is a reflection of the pleasure we take in observing that their actions fit perfectly with what we ourselves would do

---

1 According to University of Bergen, Faculty of Humanities guidelines (FFU, 2017), a PhD dissertation consisting in a collection of articles shall also contain an introductory chapter (the so-called 'kappe') that collates and contextualises the work done in each of the articles in such a way that it both demonstrates the unity of the dissertation and elevates the whole by contributing independent value to the research within. The present introduction is that 'kappe', and what follows is a combination of statement of my philosophical outlook and method, a summary of some of the most central aspects of Smith's model of moral judgement forming the background for the discussions in each of the four articles, and a summary of the four articles of the dissertation which contextualises and assembles the issues treated in each of the articles into a coherent (if incomplete) answer to the question implied in the title of the dissertation.

2 Such as in the face of human-caused climate change (a good start would be to vote for politicians who actually care about climate change).

in their shoes,[3] and, correspondingly, that our disapproval arises from observing a discrepancy between how we would react, and how the other in fact acts.

When I speak of being a good 'sentimentalist', I therefore mean to answer questions about 'how to be a good person' from within the framework of Smith's theory. That does not mean that what I have to say will be uninteresting for someone who does not think that morality has (or should have) much to do with emotions, imagination, or propriety. On the contrary, I believe that many of the questions that arise from the perspective of Smith's brand of sentimentalism, and many of the answers this theory offers in return, are relevant to *any* person concerned with how to be a *good* person, regardless of what one thinks 'goodness' ultimately consists in.

Which questions, and what answers, will, however, be up to you, the reader, to decide. This is not just a piece of coquetry on my part. The answers I give are only capable of guiding your actions to the extent that you find them speaking to the image you already have of yourself as a person who engages in moral reasoning and action. This is the second way in which my approach is particular: There is nothing in these articles to compel you to act in this way or that—even if you accept the basic premises of Smith's theory (of which more later). The arguments I offer and claims I make in each of the four articles of this dissertation—*Moral Tuning*, *Love Redirected*, *The Practical Impossibility of Being both Impartial and Well-informed*, and *The Partially Impartial Spectator as an Ethical Ideal*—are only things that I offer for your consideration. You must decide for yourself, *by your own lights*, what, if any, implications they have for the way you go about your own moral life.[4]

Of course, there is a sense in which any normative ethical theory is normative in a 'by your own lights'-kind of way, since their acceptance by you, the reader, as a theory about what *you* should do ultimately hinges on the extent to which they are intuitively appealing *to you*. However, as Knud Haakonssen points out, standard ethical theories like deontology and consequentialism are normative in a very *direct* way; their aim is to establish a 'criterion of right action' (Haakonssen, 2002, p. xviii) that, once established and accepted, supplants whatever pre-theoretical intuitions we may have had about morality and simply *tells* us what we should do (albeit in an abstract formula or maxim for deliberation and/or action).

In case of (Kantian) deontology, accepting the theory means accepting that morally right actions are those the maxims of which can be made universal laws applicable to all rational

---

3   To be precise, what we *ideally* would do. See Griswold (Griswold, 1998, p. 85).

4   Geoffrey Sayre-McCord uses this apt expression to capture the peculiar normativity to be found in TMS—see for example his (2013). Charles Griswold similarly speaks of the 'main source of light we possess in moral philosophy, namely prephilosophical ethical life.' (Griswold, 1998, p. 74).

creatures (or that acting morally is acting in respect of the moral law). In case of (utilitarian) consequentialism, accepting the theory means accepting that morally right actions are those that maximise the total 'utility' or happiness among all sentient creatures (or that acting morally is acting so as to maximise such utility).

In case of Smith's sentimentalism, there is no such criterion of right action that, if we accept the theory, tells us what to do. The closest we get is, as we shall see, the 'impartial spectator', but I agree with Haakonssen that the impartial spectator should *not* be understood as a criterion of right action (at least not in the standard sense); Smith, that is, is not trying to establish an 'Ideal Impartial Spectator who has the last word on what is truly proper to be done in a given situation' (Haakonssen, 2002, p. xviii); as Charles Griswold puts it, '[t]here is no such thing as "*the* impartial spectator", just this or that spectator who may be more or less impartial (in some sense of the term)' (Griswold, 2010, p. 71 emphasis in original)—'this or that spectator' being you and I.[5]

Smith's lack of interest in establishing a criterion of right action reflects the nature of the endeavour in which he takes himself to be engaged, which, as he puts it, is 'not concerning a matter of right … but a matter of fact' (TMS, II.i.5.10, p. 90). This emphasis on 'fact' in contrast to 'right', of *de*scription in contrast to *pre*scription, shows the influence of David Hume's 'science of man' (see Treatise, Introduction; Hume, 2007, p. 4) on Smith's work. The phrase 'science of man' was supposed to capture the idea of applying a characteristically *scientific* method of explanation, especially as the Scottish Enlightenment thinkers found this method exemplified by Newton, to the subject of 'man', that is to say, of people, or humans, or humanity.[6]

Several scholars have read TMS as a continuation of Hume's science of man. T. D. Campbell in particular has argued that TMS should be read, first and foremost, as a 'pioneering venture in the scientific study of morality' (Campbell, 1975, p. 68), where, as he elegantly puts it, modifying one of Smith's own phrases, the 'simple and familiar fact of sympathy' is given the same place in the system of ethics that the 'simple and familiar fact of gravity' occupies in Newton's system of natural philosophy (Campbell, 1975, pp. 69–70).

Others disagree, noting, like Charles Griswold, that, for all his veneration of Newton (see A. Smith, 2002, II.2.20, p. 144-145), Smith never writes of his own work as Newtonian (Griswold,

---

5 Which is one important difference (among others) between Smith's impartial spectator and later (normative interpretations of) so-called Ideal Observer Theory (Firth, 1952). I write more about this in *The Partially Impartial Spectator* (p. 117).

6 Indeed, to Hume, the science of man, the science of the 'principles of human nature', would be the foundation of all other sciences, since all other sciences must necessarily be founded on the application of the human understanding, and thus on these principles (Hume, 2007, p. 4). The sexist language is a lamentable feature of the time, and I have done what I can to counterbalance it in my writing.

1998, p. 72),[7] or, like Samuel Fleischacker, that Smith's approach resembles moral *phenomenology* more than psychology or sociology—modern applications of the scientific method to the subject of individuals and groups, respectively—in that he carefully examines the workings of our moral judgements from *within* lived experience (Fleischacker, 2017, section 1), rather than from the outside perspective of a scientist.[8]

Whatever the right description of Smith's empirical method in TMS might be, there can be no doubt that Smith, in contrast to much moral philosophy both then and now, is primarily concerned with examining how we *in fact* judge each other, as compared to arguing about how we *ideally* should go about making moral judgements.

Still, TMS is not a work of pure description. While Campbell disagrees with Griswold and Fleischacker on the nature of Smith's empirical endeavour, they all more or less agree that Smith's careful examination of the operations of our moral judgements results in an account of morality that also *criticises* and partly *vindicates* the 'moral attitudes' thus studied (Campbell, 1975, p. 68). The task of Smith's theory is, as Carola Freiin von Villiez puts it, to 'sufficiently *explain* why people can be expected to adhere to the moral principles advanced by the theory (or, alternatively, simply demonstrate that they in fact generally do so) as well as convincingly *justify* these principles.' (2011, p. 30). Whether that justification works, and if so, by what means, is, as we shall see in the section on *Moral Tuning*, a matter of some debate.

Without going into that debate here, we can recognise that the normativity in TMS is, as Haakonssen notes, a very *indirect* kind of normativity (Haakonssen, 2002, p. viii): We are not told what to do, but in some way brought to think that *parts* of what we already think and do are things we can *continue* to think and do, while other parts need to be scrapped or revised. Compared to the criteria of right action proffered by standard theories of ethics, this is a kind of normativity that is hard capture in a simple formula.[9]

---

7   However, see (A. Smith, 1985, pp. 144–145).

8   For more on Smith and Newton, see Deborah Redman (1993). For more on Hume and Newton, see for example Eric Schliesser (2008).

9   I am not entirely happy with the impression I have given that there is a very sharp distinction between something like Kantian deontology and Smith's sentimentalism when it comes to the question of the criterion of right action. For, while Smith is certainly not trying to establish a 'synthetic' criterion that should supplant our pre-theoretical intuitions, he is, in a sense, revealing to us a criterion by which we already evaluate the rightness of our actions, namely the idea of the impartial spectator. Kant, too, saw himself as revealing rather than constructing a criterion of right action. Perhaps, then, it is not really a matter of a single clear difference, but rather of an *attitude* towards ethics (and philosophy in general); something like that captured by Bernard Williams in his *Ethics and the Limits of Philosophy* (2006). That is why I use the formulation 'tell us what to do': In responding to the question 'what should I do?', moral philosophy *cannot* tell us what to do, nor should it try to. Instead, moral philosophy—ethics—

There are, however, ways of capturing what this indirect normativity is about. One of them, and the one I find most helpful, begins by noticing a peculiarity in the way Smith *approaches* the question of what it is that makes something right (or good, or virtuous). In Part VII of TMS, Smith gives an overview and critique of what was then the standard 'systems of moral philosophy' (TMS, VII, p. 313). Before setting out to give his overview, he makes the following distinction:

> In treating of the principles of morals there are two questions to be considerd. First, wherein does virtue consist? Or what is the tone of temper, and tenour of conduct, which constitutes the excellent and praise-worthy character, the character which is the natural object of esteem, honour, and approbation? And, secondly, by what power or faculty in the mind is it, that this character, whatever it be, is recommended to us? Or in other words, how and by what means does it come to pass, that the mind prefers one tenour of conduct to another, denominates the one right and the other wrong; considers the one as the object of approbation, honour, and reward, and the other of blame, censure, and punishment? (TMS, VII.i.2, pp. 313-314)

The upshot of the following sections is that Smith's own answer to the first question agrees more or less with those who, like Aristotle, identify virtue with propriety (TMS, VII.ii.1.12, p. 320), but that he improves upon these by giving the only plausible answer to the second question—to how propriety is 'measured'—namely *sympathy* (Raphael, 2007, p. 71).

However, if I only said that Smith improves upon extant answers to the first question by giving a good answer to the second, I would radically undersell the value of Smith's approach. For, and this is part of what makes him so interesting, Smith tackles these two questions in opposite order. Instead of trying, right away, to find an answer to what it is that characterises 'virtue' (or 'right' or 'good'), he leaves aside this thorny question (about which most moral philosophy is an endless quarrel) and begins by examining the second, empirical one. What at first may appear paradoxical—after all, if we do not know what virtue is, how do we identify the faculty of mind that responds to virtue?—soon reveals itself as one of the most inspired moves in TMS.

To see just how inspired, we shall make a tour through the first part of TMS to discover the basics of Smith's model of moral judgement and the core elements of moral psychology that go into his answer to the second question.

---

can and should, as much as possible, aid our thinking. Not, mind you, by narrowing it down to a choice between theories or principles prescribing conflicting responses to comically abstract thought experiments, but, centrally, by developing our capacity for something like what Sophie Grace-Chappell calls 'imaginative identification' (T. Chappell, 2014a), and what Smith would probably call 'brining the case home' (TMS, I.i.1.4, p. 13).

I.i. Smith's Model Of Moral Judgement

In line with his overall project, Smith begins TMS by making an empirical observation:

> How selfish soever man may be supposed, there are evidently some principles in his nature that interests him in the fortune of others, and renders their happiness necessary to him, though he derives nothing from it except the pleasure of seeing it. (TMS, I.i.1.1, p. 11)

Griswold notes that Smith starts TMS as if 'in the middle of a conversation' (1998, p. 73). That is a particularly useful perspective for thinking about the intellectual context in which TMS was published, one where the 'selfish' theory of morality presented by Bernard Mandeville in his *The Fable of the Bees* (2011) gave voice to what everyone will be quick to accept of others, and, upon honest reflection, must also accept of themselves, namely that we are selfish. We are selfish, Smith assents,[10] but, he adds, not *only* selfish, for we evidently also care about others; we are interested in each other's welfare.

He goes on to note that the most obvious example of how we are interested in the welfare of others is 'pity or compassion', which is 'the emotion we feel for the misery of others, when we either see it, or are made to conceive of it in a very lively manner' (TMS, I.i.1.1, p. 11). That we do in fact thus 'derive sorrow from the sorrow of others' is so obvious, thinks Smith that it is beyond the need for illustration. Indeed, so common is this tendency that even the 'greatest ruffian, the most hardened violator of the laws of society' is not altogether without it (TMS, I.i.1.1, p. 11)—hence, all readers should instantly recognise what Smith is talking about as a real phenomenon.[11]

However, as Smith goes on to note, while our tendency to pity and compassion is perhaps the most obvious way in which we are 'interested in the fortune of others', we also share other people's *positive* emotions (indeed, as he will later argue, we share positive emotions more readily and perfectly than negative ones, TMS, I.ii.1.1, p. 53). We feel joy at another's joy just as much as we are distressed at another's distress, and to name this general tendency of ours, Smith repurposes a term that was then already in use:

---

10 As he puts it when dealing directly with Mandeville: 'But how destructive soever this system may appear, it could never have imposed upon so great a number of persons, nor have occasioned so general an alarm among those who are the friends of better principles, had it not in some respects bordered upon the truth.' (TMS, VII.ii.4.14, p. 370)

11 Modern readers might be reminded that some of the most 'hardened violator[s] of the laws of society'—violent psychopaths—do in fact appear to be altogether without pity or compassion (Blair, 1995). However, the clinical case is less clear-cut than that, and the metaethical/moral/legal implications still less so. See for example Maibom (2008). Moreover, even if it were true that certain people are incapable of pity, the fact that those same people are also some of the most hardened violators of the laws of society would not undermine a theory that bases the sense of propriety on a capacity that includes pity.

> Pity and compassion are words appropriated to signify our fellow-feeling with the sorrow of others. Sympathy, though its meaning was, perhaps, originally the same, may now, however, without much impropriety, be made use of to denote our fellow-feeling with any passion whatever. (TMS, I.i.1.5, p. 13)

Sympathy, then, means 'fellow-feeling with any passion whatever' to Smith. For modern readers more used to talk of 'empathy', as in Barack Obama's talk of an 'empathy deficit' (2006), it is easy to substitute that latter term for the former. There is no great danger in this substitution, so long as one keeps in mind that 'empathy' is used in several, partly contradictory ways in modern philosophy and psychology (Coplan, 2011; and see the introductions to Coplan & Goldie, 2011; and to Maibom, 2017).[12] Indeed, as Remy Debes makes abundantly clear in his comprehensive review of the history of that term, there has always been confusion about the exact meaning of 'empathy' (and sympathy, for that matter; Debes, 2015). This plurality complicates reading the psychological literature on the role of empathy in moral motivation (see e.g., Batson, 1987, 1990; Batson, Duncan, Ackerman, Buckley, & Birch, 1981) and development (see e.g., Findlay, Girardi, & Coplan, 2006; Ugazio, Majdandzic, & Lamm, 2014; Zaki & Ochsner, 2012). Moreover, and more to the topic of this dissertation, the plurality in meanings ascribed to 'empathy' in the empirical literature has rendered unnecessarily complicated the debate on what role, if any, empathy *should* play—that is, whether we should embrace and develop our capacity for empathy as the foundation for (better) moral judgement (see e.g., Batson & Ahmad, 2009b; Persson & Savulescu, 2018), or reject and extirpate it as a dangerous nuisance that gets in the way of real ethical reflection (see e.g., Batson & Ahmad, 2009a; Prinz, 2011; see also the debate in Zaki, 2015).

I will not try to sort out the tangle of the modern debate on empathy here.[13] To avoid confusion, I will simply stick with Smith's original term 'sympathy', introducing cognate terms where appropriate. However, I will also not attempt a more exact positive definition of what sympathy is than what Smith himself provides. It may be that I thereby make myself guilty of contributing to the confusion, but I think not. For the purpose of this dissertation—exploring the problems of being a good person and finding possible answers to how to do better—the relative vagueness of Smith's concept is an advantage. It allows us to focus, not on whether any particular conception of empathy/sympathy is or is not supported by a given piece of empirical evidence or

---

12 If you wonder why Smith didn't just use 'empathy', the answer is that the latter term was not coined until 1909, some 100-odd years after Smith's death (Jahoda, 2005).

13 I may however say that I (partly) endorse Dan Zahavi's (in Zaki, 2015) useful list of what empathy/sympathy isn't: '1. Empathy does not dissolve the boundaries between one person and another … 2. Empathy is not about affective sharing … 3. Empathy does not require similar states in empathizer and target … 4. Empathy is not prosocial', with a possible exception for 4., for which see Meyers (2017, pp. 220–221).

philosophical argument, but on what these can teach us about ourselves. That we can learn something even with such slipshod dealings with definitions is, I hope, something the rest of the dissertation will demonstrate to the reader's satisfaction.

Some semantic problems thus squared away, we can return to Smith's procession of empirical observations and theoretical generalisations, beginning with those that lead to his positing of sympathy as a fundamental principle of moral psychology. After having established 'sympathy' as 'fellow-feeling with any passion whatever', Smith goes on to make another observation that should be familiar to those versed in the contemporary psychological theories of empathy:

> The passions upon some occasions, may seem to be transfused from one man to another, instantaneously, and antecedent to any knowledge of what excited them in the person principally concerned. Grief and joy, for example, strongly expressed in the look and gestures of any one, at once affect the spectator with some degree of a like painful or agreeable emotion. A smiling face is, to every body that sees it, a cheerful object; as a sorrowful countenance, on the other hand, is a melancholy one. (TMS, I.i.1.6, p. 13)

Some have seen this as evidence that Smith included a very rudimentary form of sympathy in his theory, a kind of proto-empathy which today is sometimes referred to as 'emotional contagion' (Hatfield, Cacioppo, & Rapson, 1993). Maria Carrasco, for example, includes what she takes to be a 'purely non-cognitive' form of sympathy in her reconstruction of Smith's index of gradually more complex kinds of sympathy (Maria A. Carrasco, 2011, pp. 10–11).[14]

However, in the next couple of paragraphs, Smith goes on first to critically reflect upon, and then reject the idea of a non-cognitive 'transfusion' of emotion from one person to another. This rejection begins with another of his keen observations, namely that someone else's anger 'serves rather to disgust and provoke us against' those who are angry (TMS, I.i.1.7, pp. 13-14). Thus, while emotions like sorrow and joy seem to pass directly from one person to another, emotions like anger and resentment evidently do not. Why not? Smith points out that anger involves two people: the angry person, and the person towards whom the anger is directed. So long as we do not know what the target of the angry outburst has done to deserve it, we are much more apt, thinks Smith, to sympathise with the fear of the one than the anger of the other:

---

14 Note that 'emotional contagion', while it is supposed to rely *less* on (explicit/conscious) cognition than the 'more cognitive, sophisticated, and 'socially beneficial' processes of sympathy and empathy' (Hatfield, Cacioppo, & Rapson, 1993, p. 96), does not exclude cognitive mediation, and is sometimes used in contexts where the contagion could *only* occur with cognitive mediation, such as in the scandalous Cornell University/Facebook 'emotional contagion' research where users' moods were subtly altered by manipulating the ratio of positive to negative posts in their 'News Feed' (Chambers, 2014; Kramer, Guillory, & Hancock, 2014).

> The furious behaviour of an angry man is more likely to exasperate us against himself than against his enemies. As we are unacquainted with his provocation, we cannot bring his case home to ourselves, nor conceive any thing like the passions which it excites. But we plainly see what is the situation of those with whom he is angry, and to what violence they may be exposed from so enraged an adversary. We readily, therefore, sympathise with their fear or resentment, and are immediately disposed to take part against the man from whom they appear to be in so much danger. (TMS, I.i.1.7, p. 14)

It is this sympathy with the presumed victim that serves to 'disgust and provoke us against' the one who is angry, and it is only when 'we are acquainted with what gave occasion to [the anger]' (TMS, I.i.1.7, p. 14) that we may begin to sympathise with the angry person instead, that is when we are in a position to sympathetically feel anger at whatever provocation might have been given.

Faced with the apparent dichotomy in our reactions to the emotional displays of others, Smith must either posit different sympathetic mechanisms for the two classes of emotions (which he later terms 'social' and 'unsocial', TMS, I.ii.3-4, pp. 41-47) or reject the appearance of direct or unmediated emotional contagion. He chooses the latter:

> If the very appearance of grief and joy inspire us with some degree of the like emotions, it is because they suggest to us the general idea of some good or bad fortune that has befallen the person in whom we observe them: and in these passions this is sufficient to have some little influence upon us. (TMS, I.i.1.8, p. 14)

In other words, there is no such thing as a purely non-cognitive sympathy.[15] Whether Smith is warranted in this choice can be debated (see for example Ruffman, Lorimer, & Scarf, 2017 for a critical review of the supposed links between 'contagious crying' and empathy), but this confusion about transfusion offers a useful opportunity for a hermeneutical lesson: In reading Smith, one ignores the many ways he qualifies his statements at one's own peril. Here, the 'may seem' should alert the reader to the possibility that Smith may not in fact endorse the way things seem at first blush—appearances can be deceiving. In this case, Smith clarifies the ambiguity by subsequently denying appearances in favour of a deeper, unifying explanation. That, however, is not always the case, and, as we shall see in *Love Redirected*, Smith's qualifications can also give rise to some prickly interpretive problems.

For now, let us return to Smith's observations on the particularities of sympathy. As the case of anger illustrates, we do not sympathise with people's emotions *as such*. Granted, we may sometimes say things like 'I'm so happy you are happy!', as when the mood of a depressed friend

---

15 Debes notes this as one of the major distinctions between Hume's view of sympathy and Smith's (2016, pp. 194–195).

appears to have lifted. While the joy of our friend is then in some sense the primary object of our sympathy, our sympathy remains contingent on contextual factors. Thus, for example, if it turned out that our friend was happy at having discovered a new way of killing himself, our sympathetic joy would presumably peter out. Even when the sentiment itself is the primary object of our sympathy, that is, we also hold some background assumption about the thing to which that sentiment is a response, such as whatever it is about life that *we* find enjoyable and genuinely wish our friend was able to equally enjoy.[16]

In other words, when we sympathise, we always, in some way, sympathise with sentiments as *responses* to a situation. While the broad smile and sparkling eyes of our partner alert us to the possibility that something good has befallen her, thus instantly lifting our spirits and bringing the beginnings of a smile to our own face, it is not until she informs us of the success of her latest green energy policy initiative that we can properly enter into her joy and share her elation in both body and mind. General ideas can inspire us with 'some degree' of the like passion, but for our sympathy to be more properly called by that name, we must know something more about the situation that gave rise to the original sentiment. As Smith puts it, 'Sympathy ... does not arise so much from the view of the passion, as from that of the situation that excites it' (TMS, I.i.1.10, p. 15).[17]

This is crucial. For, if sympathy arises from the spectator's view of the situation of the 'person principally concerned' (TMS, I.i.1.4, p. 13) rather than directly from observing the passions/emotions of that person, there arises the possibility that a spectator may come to feel something *different* to the person principally concerned—or, more precisely, something different to the feeling *expressed* by the person principally concerned. Indeed, as Smith goes on to note,

> We sometimes feel for another, a passion of which he himself seems to be 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. We blush for the impudence and rudeness of another, though he himself appears to have no sense of the impropriety of his own behaviour; 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, p. 15)

This feature of our capacity for sympathy—that our 'fellow-feeling with any passion whatever' arises not from a view of the passion itself, but from considering the situation of the person

---

16 It is something else entirely when we approve of someone's sentiments as part of a willed manipulation of their feelings to suit some purpose of ours. For example, if a demagogue, through skilful use of language, manages to get citizens riled up about some imaginary enemy in their midst, the demagogue can be fully aware of the illusory nature of this anger, but approve of it as a means to an end.

17 Hence the link to the appraisal theory of emotion noted above (see e.g., Moors, Ellsworth, Scherer, & Frijda, 2013).

principally concerned—also helps explain some things for which it would otherwise be hard to account, such as our horror at seeing somebody who has lost their mind but appears laughing and happy:

> Of all the calamities to which the condition of mortality exposes mankind, the loss of reason appears, to those who have the least spark of humanity, by far the most dreadful, and they behold that last stage of human wretchedness with deeper commiseration than any other. But the poor wretch, who is in it, laughs and sings perhaps, and is altogether insensible of his own misery. The anguish which humanity feels, therefore, at the sight of such an object, cannot be the reflection of any sentiment of the sufferer. The compassion of the spectator must arise altogether 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, p. 15)

This same feature of our sympathetic imagination, Smith goes on to argue, also explains our commiseration with the dead, which, rather conspicuously, appears to be a reflection of 'those circumstances which strike *our* senses' (TMS, I.i.1.13, p. 16, my emphasis), rather than any true reflection of the situation of the dead person:

> It is miserable, we think, to be deprived of the light of the sun; to be shut out from life and conversation; to be laid in the cold grave, a prey to corruption and the reptiles of the earth; to be no more thought of in this world, but to be obliterated, in a little time, from the affections, and almost from the memory, of their dearest friends and relations. Surely, we imagine, we can never feel too much for those who have suffered so dreadful a calamity. (TMS, I.i.1.13, p. 16)

Now, this does not seem *altogether* unreasonable, at least not the part about being 'obliterated' from the affections of our dearest friends and relations (but then there is a curious paradox involved in the thought that we 'can never feel too much' for those who we will soon forget *because* we will soon forget them). Whatever your views of death might be, Smith does have a point when he goes on to note that his mostly Christian readership thereby overlook what of their own admission should be the most important aspect of the situation of the dead, namely 'that awful futurity which awaits them' (TMS, I.i.1.13, p. 16).[18] The fact that we commiserate with the dead, as with those who have lost their mind, is in a kind of 'illusion of the imagination':

> The idea of that dreary and endless melancholy, which the fancy naturally ascribes to their condition, arises altogether from our joining to the change which has been produced upon them, our own consciousness of that change, from our putting ourselves in their situation, and from

---

18 'Awful', I assume, as in 'full of awe'.

our lodging, if I may be allowed to say so, our own living souls in their inanimated bodies, and thence conceiving what would be our emotions in this case. (TMS, I.i.1.13, p. 16)

And, as with many illusions, it is not easily dispelled, not even by religious belief. Smith argues that there is a certain sense in this. The illusion serves a function:

> It is from this very illusion of the imagination, that the foresight of our own dissolution is so terrible to us, and that the idea of those circumstances, which undoubtedly can give us no pain when we are dead, makes us miserable while we are alive. And from thence arises one of the most important principles in human nature, the dread of death, the great poison to the happiness, but the great restraint upon the injustice of mankind, which, while it afflicts and mortifies the individual, guards and protects the society. (TMS, I.i.1.13, pp. 16-17)

In other words, the illusory sympathy we feel for the dead, or rather for someone in the situation of being dead (impossible as that is), gives rise to our fear of death because we fear being in the situation of being dead (impossible as this is). This fear, while it poisons the happiness of the individual, serves in as a restraint on our willingness to sacrifice ourselves in rash fits of violence, and this restraint, finally, helps protect society by making everyone more secure.[19]

That Smith's conception of sympathy can account for these phenomena is in itself a strength of his theory. However, the most important result of recognising that sympathy arises 'from a view of the situation' is the gap it opens up between the sentiments actually felt and expressed by *ordinary* agents (in contrast to the insane and 'the' dead) and those felt sympathetically by the spectators. For in this gap, there is room for *evaluation* of the sentiments thus felt.

Before he gets to the part of his exposition that deals with evaluation, however, Smith spends some time noting how sympathy is connected with pleasure. First, he takes up the theme of the pleasure we get from seeing others sympathise with us:

> [W]hatever may be the cause of sympathy, or however it may be excited, nothing pleases us more than to observe in other men a fellow-feeling with all the emotions of our own breast; nor are we ever so much shocked as by the appearance of the contrary. (TMS, I.i.2.1, p. 17)

Smith's next observation is that sympathy from others does not work simply by enlivening the joy we are already feeling. Rather, sympathy brings a pleasure of its own that, in the case of joyous sentiments is added to the total, and in the case of sorrowful sentiments alleviates them by

---

19 Of course, this restraint can be overcome, evidence for which is plentiful, not only among modern-day suicide bombers, but in the willingness to sacrifice that is the prerequisite of any war. Griswold notes that Smith, because he bases our fear of death in the imagination; and, in contrast to Hobbes, who ties preservation to the fear of violent death; is thus able to explain how people can in fact choose to sacrifice their own life for what they imagine is the praiseworthiness of sacrifice for a greater cause (Griswold, 1998, p. 119).

'insinuating into the heart almost the only agreeable sensation which it at that time is capable of receiving' (TMS, I.i.2.2, p. 18), namely the pleasure of knowing that our friend *shares* our pain:

> How are the unfortunate relieved when they have found out a person to whom they can communicate the cause of their sorrow? Upon his sympathy they seem to disburthen themselves of a part of their distress: he is not improperly said to share it with them. He not only feels a sorrow of the same kind with that which they feel, but as if he had derived a part of it to himself, what he feels seems to alleviate the weight of what they feel. Yet by relating their misfortunes they in some measure renew their grief. They awaken in their memory the remembrance of those circumstances which occasioned their affliction. Their tears accordingly flow faster than before, and they are apt to abandon themselves to all the weakness of sorrow. They take pleasure, however, in all this, and, it is evident, are sensibly relieved by it; because the sweetness of his sympathy more than compensates the bitterness of that sorrow, which, in order to excite this sympathy, they had thus enlivened and renewed. (TMS, I.i.2.4, pp. 18-19)

In other words, although relating our misery to another person may for the moment make us more miserable—we must, yet again, bring forth all the details of the cruel fate to which we have been assigned that we, for the moment, may even have been able to suppress—the experience of relating our sorrow to another and having them understand and commiserate with us so much lightens our load that the act of telling our story, however much we may beforehand have dreaded the pain and embarrassment we imagine doing so would bring upon us, becomes on the whole a *pleasurable* experience. Of course, this outcome depends on the sympathy of our friend. If that should be lacking, we are not only disappointed, but resent the levity with which we are treated:

> The cruelest insult, on the contrary, which can be offered to the unfortunate, is to appear to make light of their calamities. To seem not to be affected with the joy of our companions is but want of politeness; but not to wear a serious countenance when they tell us their afflictions, is real and gross inhumanity. (TMS, I.i.2.4, p. 19)

Consequently, just as we desire that our friends should share in our sorrows and resentments, and not just our joys and gratitude, we ourselves

> run not only to congratulate the successful, but to condole with the afflicted; and the pleasure which we find in the conversation of one whom in all the passions of his heart we can entirely sympathise with, seems to do more than compensate the painfulness of that sorrow with which the view of his situation affects us. (TMS, I.i.2.6, p. 19)

As Dennis C. Rasmussen puts it, 'we naturally enjoy the feeling of sentimental concord—of being on the same emotional page as someone else' (2017, p. 110). Thus, the *pleasure of mutual sympathy*

gives rise to a desire both to get the sympathy of others and to sympathise with them.[20] Combined with the fact that we sympathise with a person's response to a situation rather than with the person themselves, the desire for mutual sympathy forms the foundation for the establishment of moral judgement and moral norms. To see how, consider Smith's next observation:

> When the original passions of the person principally concerned are in perfect concord with the sympathetic emotions of the spectator, they necessarily appear to this last just and proper, and suitable to their objects; and, on the contrary, when, upon bringing the case home to himself, he finds that they do not coincide with what he feels, they necessarily appear to him unjust and improper, and unsuitable to the causes which excite them. To approve of the passions of another, therefore, as suitable to their objects, is the same thing as to observe that we entirely sympathise with them; and not to approve of them as such, is the same thing as to observe that we do not entirely sympathise with them. (TMS, I.i.3.1, p. 20)

In that happy case where we have the full sympathy of our friend, or, better, where *we* are fully able to sympathise with *her*, our sentiments are also more or less the same with hers. But if our sentiments *are* more or less the same, then we implicitly *share* her evaluation or appraisal of the situation to which her sentiments are a response. Thus, if we become aware of our sympathy with another, then we also become aware that we share their sentiments, their appraisal of the situation. And to observe that we do so is to observe that these are appropriate—they are the very sentiments we would have if we were her, indeed, we now *have* those sentiments on her account. If, on the contrary, we find that we cannot entirely share her sentiments in light of the situation, that we cannot fully sympathise, then we naturally also *disapprove* of her sentiments as somehow inappropriate as a reaction to her situation—we would have reacted differently if we were her.

In the introduction to *Love Redirected*, I illustrate this process with an example from Shakespeare's *Henry IV*, in which Lord Chief Justice asks the newly crowned Henry V to:

> Question your royal thoughts, make the case yours;
> Be now the father and propose a son
> Hear your own dignity so much profaned,
> See your most dreadful laws so loosely slighted,
> Behold yourself so by a son disdain'd;
> And then imagine me taking your part
> And in your power soft silencing your son
> (Henry IV Part 2 Act 5 Scene 2)

---

20 Griswold adds to this the agent's fear of being alone and, for the spectator, the 'natural proclivity' of the imagination to 'enter into the situations of others' (1998, p. 122). For a slightly different take, see Debes' discussion of 'affective dignity' (2017).

Lord Chief Justice, that is, is asking the young monarch to set aside his (apparent) anger at the Chief Justice, who, when Henry V was still prince, had him imprisoned for breaking the law,[21] and imagine himself in the shoes of his father, Henry IV. What would he have done with a son who so 'loosely slighted' his laws? It is only by thus placing himself in the shoes of his father, the Chief Justice is implying, that Henry V can correctly judge the actions of the Chief Justice as appropriate or inappropriate reactions to the situation in question.

In Smithian terms, Lord Chief Justice appeals to Henry V's capacity for sympathy. Henry V, for his part, appears to exercise this capacity, finding that he entirely *sympathises* with the course of action taken by Lord Chief Justice. Aware of this sympathy—this conviction that he, if he were his father, would approve of the action taken by Lord Chief Justice—Henry V cannot but approve of that reaction as appropriate to the situation. And so he does: 'You are right, justice, and you weigh this well' (Henry IV Part 2 Act 5 Scene 2).[22]

The example of Henry V and Lord Chief Justice also illustrates a second aspect of the connection between sympathy and judgement. We are not so simple-minded as (always) to take a mere awareness of a discrepancy between our own sentiments and those of the person principally concerned as full and final proof of their impropriety. Motivated as we are to be *able* to sympathise with the person principally concerned, we may look for some explanation in her situation for the way she appears to feel, or ask or even demand such an explanation. If we fail to find such a thing, or she fails to provide an explanation that we can go along with—fails, that is, to put the situation in such a light as to allow us to see her reaction as one we can share, which is a major part of what Lord Chief Justice is doing here with Henry V—only then do we judge her reaction as *inappropriate*.

Of course, we frequently *do* jump to conclusions from the mere observation of differences in opinion, which, as we shall see, is a major topic of the fourth article of this dissertation. However, in this case, as in many others, children can reveal the basic mechanisms in play: In a study with three-

---

21 By, it would seem, hitting Lord Chief Justice over the head: '[Your highness] struck me in my very seat of judgment' (Henry IV Part 2 Act 5 Scene 2).

22 Whether Henry V really is angry or is only acting to test the reaction of the Chief Justice is a matter that I will leave aside here. Note also that Lord Chief Justice here invites Henry V to imagine himself, not in the shoes of the Chief Justice, but in those of the former prince's father, Henry IV. Asking the king to imagine himself in the shoes of anyone below his exalted position would probably be an insult in its own right, but note that the change of target does not violate the basic structure of sympathy identified by Smith, for, when imagining himself in the shoes of his father, Henry V imagines what he would have approved his Lord Chief Justice doing about an insubordinate son. The propriety of the actions of Lord Chief justice thus flows, in imagination as in fact, from the authority of the king.

year-old children, Robert Hepach, Amrisha Vaish, and Michael Tomasello (2013) found precisely the pattern of reactions that one would expect from Smith's insistence that sympathy arises from a view of the *situation* of the person principally concerned: When an adult confederate got his hand 'pinched' in the lid of a box, and cried out in pain, the children were quick to comfort him. When, however, the adult 'pinched' the sleeve of his shirt, but likewise cried out in pain, the children, rather than comfort him, looked around instead for something that would explain the outcry, and, failing to find anything, remained puzzled.

It might seem obvious that there would be this connection between our own sentiments and our approval or disapproval of the sentiments of another, but it is worth noting why this is the case. Our sentiments express an evaluation, an appraisal, of the situation to which we are reacting. Unless we have some specific reason to doubt our own reactions, we will naturally tend to take them as *appropriate* reactions to the situation as it (really) is (see e.g., Ward, Ross, Reed, Turiel, & Brown, 1997). When we, as spectators, sympathetically imagine ourselves reacting to a situation, our sentiments are equally an evaluation of that situation. Unless we have some specific reason to doubt our reactions, we will naturally also take our sympathetic reactions as appropriate evaluations of the situation as it (really) is.

However, in contrast to the cases of 'illusory' sympathy with the dead noted above, in most ordinary cases, when we as spectators thus apprise a situation, we can compare our sentiments with those expressed by the person principally concerned. As spectators, that is, we can compare what we feel when we imagine ourselves in the situation of the person principally concerned with the emotion actually expressed by that person. As already noted, the two sentiments may differ. Since, then, each sentiment represents the evaluation or appraisal that one person has of the situation, differences in sentiment mean differences in appraisal. If, upon trying to sympathise with someone's reaction, we find that we *cannot* sympathise, we are in effect finding that we disagree with that person's reaction, and, if the reaction is a reflection of their appraisal, then with their appraisal.

'If I were you, I would not have got so angry about that comment'. This is a kind of comment I, in the role of the spectator, frequently find myself making (no doubt to the boundless irritation of the person principally concerned). My reaction differs from yours, and because I see my own reaction as an appropriate reaction, there is an element of censure involved in this remark: If I were you, *I* would not have got so angry, and *so you should not get so angry either*. Our own sentiments, that is, are always the standard by which we judge the sentiments of others (hence the name 'sentimentalism'):

> 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, p. 23)

However, in order for my sympathetic sentiments to count as an *appropriate* standard of evaluation for your sentiments, it is clear that I must first fulfil my side of the sympathetic bargain; I must 'bring home' your case to my 'breast', that is *really* imagine myself as you, responding to the situation as it confronts you:

> In all such cases, that there may be some correspondence of sentiments between the spectator and the person principally concerned, the spectator must, first of all, endeavour, as much as he can, to put himself in the situation of the other, and to bring home to himself every little circumstance of distress which can possibly occur to the sufferer. He must adopt the whole case of his companion with all its minutest incidents; and strive to render as perfect as possible, that imaginary change of situation upon which his sympathy is founded. (TMS, I.i.4.6, p. 26)

In other words, unless and until I have brought your situation home to myself, my sympathetic reaction cannot really be said to be a reaction to the situation *as it confronts you*, and to the extent that I fail to do so, my ability or inability to sympathise with your reaction cannot really be said to be a fair or accurate evaluation of your sentiments as reactions to the situation as it confronts you. If I cannot see myself getting so worked up about that comment, it may well be that I have failed to sufficiently exert my sympathetic imagination. However, as Smith goes on to note:

> After all this, … the emotions of the spectator will still be very apt to fall short of the violence of what is felt by the sufferer. Mankind, though naturally sympathetic, never conceive, for what has befallen another, that degree of passion which naturally animates the person principally concerned. That imaginary change of situation, upon which their sympathy is founded, is but momentary. The thought of their own safety, the thought that they themselves are not really the sufferers, continually intrudes itself upon them; and though it does not hinder them from conceiving a passion somewhat analogous to what is felt by the sufferer, hinders them from conceiving any thing that approaches to the same degree of violence. (TMS, I.i.4.7, pp. 26-27)

Therefore, even if we do exert our imagination in bringing home to us the case of the person principally concerned, we will often only feel a faint copy of the clear original. It's like the line in Pulp's *Common People*: 'But still you'll never get it right/'Cause when you're laid in bed at night/ Watching roaches climb the wall/If you called your dad he could stop it all, yeah' (1995). The rich Greek girl may pretend all she likes that she is 'common people', but at the end of the day, her knowledge that her dad can, at any time, bail her out, renders her experience essentially different to that of those who, like the narrator, have no escape. Similarly, whenever we sympathise with

someone, our 'secret consciousness' that we merely *imagine* ourselves in their place not only attenuates the sentiments sympathetically felt, but also to some extent changes their quality:

> What [the spectators] feel, will, indeed, always be, in some respects, different from what [the person principally concerned] feels, and compassion can never be exactly the same with original sorrow; because the secret consciousness that the change of situations, from which the sympathetic sentiment arises, is but imaginary, not only lowers it in degree, but, in some measure, varies it in kind, and gives it a quite different modification. (TMS, I.i.4.7, p. 27)

However, the person principally concerned, having herself been a spectator on multiple other occasions, 'is sensible of this'—knows, that is, that a spectator's 'compassion can never be exactly the same with original sorrow'. He knows this, and

> at the same time passionately desires a more complete sympathy. He longs for that relief which nothing can afford him but the entire concord of the affections of the spectators with his own. To see the emotions of their hearts, in every respect, beat time to his own, in the violent and disagreeable passions, constitutes his sole consolation. (TMS, I.i.4.7, p. 27)

In order to get the sympathy of the spectators, the persons principally concerned will therefore have to modify their sentiments, or at least their display of passion, to something that spectators will be able to go along with:

> [H]e can only hope to obtain this by lowering his passion to that pitch, in which the spectators are capable of going along with him. He must flatten, if I may be allowed to say so, the sharpness of its natural tone, in order to reduce it to harmony and concord with the emotions of those who are about him. (TMS, I.i.4.7, p. 27)

Spectators, that is, will exercise their capacity for sympathy in order to be able, as much as possible, to see the situation of the persons principally concerned as it presents itself to them, and the persons principally concerned will, knowing the limitations of spectator sympathy, strive to adapt their sentiments to something that the spectators can go along with. Thus, the desire for mutual sympathy, the desire to be *able* to sympathise and to receive sympathy in turn, engenders 'a self-regulating process of sympathetic exchange' (Forman-Barzilai, 2010, p. 193) which has the effect of both awakening emotions in the spectators, and cooling—as well as assuaging—emotions in the persons principally concerned:

> In order to produce this concord, 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. As they are constantly considering what they themselves would feel, if they actually were the sufferers, so he is as constantly led to imagine in what manner he would be affected if he was only one of the spectators of his own situation. As their sympathy makes them look at it, in some measure, with his eyes, so his sympathy makes him look at it, in some measure, with theirs, especially when in their presence and acting under their observation: and as the reflected passion, which he thus conceives, is much weaker than the original one, it necessarily abates the violence of what he felt before he came into their presence, before he began to recollect in what manner they would be affected by it, and to view his situation in this candid and impartial light. (TMS, I.i.4.8, p. 27-28)

Even with all this mutual effort, the feelings of the spectator and those of the person principally concerned will never be the *same*. But as Smith also notes,

These two sentiments, however, may, it is evident, have such a correspondence with one another, as is sufficient for the harmony of society. Though they will never be unisons, they may be concords, and this is all that is wanted or required. (TMS, I.i.4.7, p. 27)

This 'is evident' because people evidently do succeed in sympathising with and getting sympathy from each other. Mutual sympathy, the state of two or more people being in harmony, of sharing a view and appraisal of something, is common.

The communication and sharing of sentiments becomes possible through a mutual effort to find some 'pitch' of passion that we can both (or all) go along with. The *appropriate* degree of passion, the appropriate sentiment, the appropriate reaction to any given situation, is thus a notion that is negotiated by the persons principally concerned and their spectators. Propriety is socially constructed.

If this is true, we would expect to find that, for example, the degree of violence in displays of passion that is deemed appropriate would vary from passion to passion, and context to context, in accordance with the ability and willingness of spectators to enter into the emotive worlds of the persons principally concerned, and of persons principally concerned to bring their passions down (or up) to a pitch that spectators can go along with. This is precisely what we find. As Smith notes, friends are more indulgent with our displays of emotion than are mere acquaintances, who, in turn, are more indulgent than perfect strangers. Consequently, we compose ourselves most in the presence of the latter, and least in the presence of the former. Moreover, the difference between 'social' passions like joy and sorrow and 'unsocial' passions like anger and hatred is also reflected in the degree to which we accept public displays of each kind of emotion (TMS, I.ii.3-4, pp. 41-52;

III.5.4-5, pp. 200-201). Finally, what counts as appropriate varies between cultures, and between historical epochs:

> The different situations of different ages and countries are apt, in the same manner, to give different characters to the generality of those who live in them, and their sentiments concerning the particular degree of each quality, that is either blamable or praise-worthy, vary, according to that degree which is usual in their own country, and in their own times. That degree of politeness, which would be highly esteemed, perhaps would be thought effeminate adulation, in Russia, would be regarded as rudeness and barbarism at the court of France. That degree of order and frugality, which, in a Polish nobleman, would be considered as excessive parsimony, would be regarded as extravagance in a citizen of Amsterdam. Every age and country look upon that degree of each quality, which is commonly to be met with in those who are esteemed among themselves, as the golden mean of that particular talent or virtue. And as this varies, according as their different circumstances render different qualities more or less habitual to them, their sentiments concerning the exact propriety of character and behaviour vary accordingly. (TMS, V.2.7, p. 239)[23]

Beyond this, our awareness of how difficult it is for a spectator fully to appreciate what it is like for the person principally concerned, and for the person principally concerned to bring her feelings down to a level that makes sympathy from the spectator possible, gives rise, in turn, to two classes of *virtues*:

> Upon these two different efforts, upon that of the spectator to enter into the sentiments of the person principally concerned, and upon that of the person principally concerned, to bring down his emotions to what the spectator can go along with, are founded two different sets of virtues. The soft, the gentle, the amiable virtues, the virtues of candid condescension and indulgent humanity, are founded upon the one: the great, the awful and respectable, the virtues of self-denial, of self-government, of that command of the passions which subjects all the movements of our nature to what our own dignity and honour, and the propriety of our own conduct require, take their origin from the other. (TMS, I.i.5.1, p. 29)

Given how difficult we know it is to bring home to our own breast the case of another, feeling her joy or grief as she feels it, we are impressed when someone displays an extraordinary facility with sympathetically understanding someone else's joys and sufferings. Given what we know of the difficulties in regulating our own emotions, bringing our anger or grief down to a level where others can sympathise with them, we are impressed when someone displays an extraordinary ability to control the expression even of violent emotions. In the first instance, such a spectator is

---

23 We shall return to this variability and what it means for moral criticisms of the standard of propriety operative in a given society in the section on *Moral Tuning*.

extraordinary in respect of 'indulgent humanity', and we call this the 'amiable virtue(s)'. In the second, the agent is extraordinary in respect of self-governance, and we call this the 'awful and respectable' virtue(s) (Raphael, 2007, p. 34).

Thus, virtues, just as judgements of propriety, are reflections or rather products of the ability and inability of spectators to enter into the sentiments of persons principally concerned:

> To be amiable and to be meritorious; that is, to deserve love and to deserve reward, are the great characters of virtue; and to be odious and punishable, of vice. But all these characters have an immediate reference to the sentiments of others. (TMS, III.1.7, p.132)[24]

Granted, it is no longer a case of a 'simple' correspondence of sentiments between spectator and agent—but there is also no other standard of judgement involved. Two things distinguish judgements of 'virtue' from those of propriety. First, virtue is something over and above the humdrum propriety of ordinary, decent human intercourse:

> The amiable virtues consist in that degree of sensibility which surprises by its exquisite and unexpected delicacy and tenderness. The awful and respectable, in that degree of self-command which astonishes by its amazing superiority over the most ungovernable passions of human nature. (TMS, I.i.5.6, pp. 30-31)

Second, judgements of virtue (normally) involve an extra imaginative change of situation: In addition to sympathising with the action of the agent, we also place ourselves in the situation of those who benefit (or are harmed by) that action (Griswold, 1998, pp. 182–183). From the perspective of the acted-upon, the 'patient', we may see the agent either as deserving of gratitude or resentment. Someone who, by their superior self-command, makes it so easy for her friends to go along with her, or, by her indulgent humanity, permits her friends to unburden their emotions on her with much less restraint than normally required, that someone, we readily imagine, would be someone towards which we were grateful, and hence we approve of her actions as both *appropriate* and *meritorious*, that is, as *virtuous*.

We approve of self-command not just to the extent that we ourselves would in fact be *able* to exhibit it, nor to any arbitrarily extreme degree, but to that degree that we would be able to approve of ourselves displaying if we were the person principally concerned. There is, in other words, a kind of double imaginative transposition: We do not simply imagine ourselves in the situation of the person principally concerned, but at the same we time imagine ourselves as the kind of spectator that we in turn, if we were spectators of ourselves, could approve of, that is, the kind of spectator

---

24 This is one of several places where Prinz' theory coincides with Smith's: 'The difference between a virtuous trait and a trait that lacks moral significance is determined by our sentiments. We value certain traits' (Prinz, 2007, p. 158).

that we *want to be*. This topic, of the kind of spectator we want to be, is, as we shall see, of central importance to understanding the indirect normativity in TMS.

I.ii. The Two Questions

Smith, then, begins by answering the second question: By closely observing and analysing actual moral practices, he arrives at a description of the 'faculty' of mind that allows us to prefer 'one tenour of conduct to another'. That faculty is our capacity for 'sympathy': our ability to imaginatively place ourselves in the situation of another, thinking and feeling as we would if we *were* them. If, by exercising this capacity, we end up *sympathising* with the other, this gives rise to approval. The absence of sympathy, after having *tried* to sympathise, gives rise to disapproval, and so are established the rudiments of moral judgement.

Notice now, that, by answering the second question, namely what psychological mechanism underlies our approval, Smith also gets an answer to a version of the first question, namely what all the behaviours we approve of have in common.[25] The psychological mechanism is sympathy, and what all behaviours we approve of have in common is thus *that we sympathise with them*.

This may seem like an empty bit of wordplay, but the critical potential of Smith's theory lies hidden within it. For behind the phrase 'that we sympathise with them' lies all the ways in which we can *fail* to sympathise with someone, *not* because that person is *undeserving* of sympathy, but because *we* make mistakes in exercising our capacity for sympathy for that person. We may fail to sympathise, that is, because we fail to be the kind of spectator we want to be. Is there, then, a kind of spectator we want to be? Is there some standard against which we already, if implicitly, measure ourselves? Smith's answer is yes: We implicitly measure *any* spectator against the standard of the 'impartial and well-informed spectator' (TMS, III.2.32, p. 150).

I.iii. The Impartial Spectator

The idea of the impartial spectator, in the form of an *ideal* for moral deliberation and judgement, is the most directly normative element of TMS. Nevertheless, for all the 'perfectionist elements' that Smith introduces in connection with the impartial spectator (Forman-Barzilai, 2010, p. 18), the idea is supposed to capture an important aspect of our moral deliberations and judgements as they in fact *are*—Smith's philosophical reflection are needed only to reveal what is otherwise overlooked

---

25 As a reminder, the original formulation of the first question is: 'wherein does virtue consist?' The reformulation of Smith's question provided here loses the critical potential of the original, but that will be added back in when we get to the impartial spectator. The narrowing of virtue to a description of certain behaviours is another shortfall of this definition, but one which needs not concern us here.

because we are so conversant with judging and justifying ourselves from a third person perspective that we do not notice the underlying structure (see TMS, III.3.2-3, pp. 156-157).

However, when Smith speaks of the 'impartial spectator' in the first parts of TMS, he appears to be evoking the image of an *actual* bystander: Someone who is a spectator to the situation and the people primarily concerned, someone well-informed about what is going on, but uninvolved, 'indifferent',[26] and therefore not partial to anyone, including herself (see e.g., TMS II.i.2.2, p. 81). In other words, the impartial spectator is a figure we can all recognise, not some abstract invention of moral philosophy. As Smith's narrative progresses, the impartial spectator comes to stand for the *idea* of such a person more than any actual spectator, but the content of the notion remains more or less the same. In the form of an idea, the impartial spectator is someone we can *imagine* being present also when there are no actual, or actually impartial, spectators around.

This shift from actual to hypothetical spectators corresponds to Smith's shift in focus from analysing our judgements of *others* in Part I of TMS, to our judgements of *ourselves* in Part III. Smith's overarching theme in Part III is that we judge ourselves analogously to how we judge others, namely as spectators reacting to the actions of an agent (and/or the reactions of someone acted upon, a patient). Of course, when judging ourselves, we are, necessarily, both a spectator and an agent (or patient). In order to judge ourselves, therefore, we must, in some sense, be splitting ourselves in two, into a spectator-aspect and an agent-aspect of ourselves. We do this, argues Smith, by imaginatively inhabiting the perspective of a *hypothetical* spectator, looking at ourselves as others would see us. By so doing, we can predict how they would react to us if we do this, that, or the other.

Since our 'original desire' is to please, and our 'original aversion' is to offend others, we initially use this trick of the projective imagination to try to predict *how* we can please, and avoid offending, those around us. However, this very process of courting the approval of actual spectators is also what gives rise to the idea of the impartial spectator, which, in turn, is what we use to critically evaluate whether such approval (or disapproval) is merited. How the idea of the impartial spectator arises out of this process is best captured in a paragraph that Smith removed from the 6th edition of TMS:

> When we first come into the world, from the natural desire to please, we accustom ourselves to consider what behaviour is likely to be agreeable to every person we converse with, to our parents, to our masters, to our companions. We address ourselves to individuals, and for some time fondly pursue the impossible and absurd project of gaining the good-will and approbation of every body. We are soon taught by experience, however, that this universal approbation is

---

26 Not in the sense of uncaring, but in the sense of having no personal or vested interests in the matter.

altogether unattainable. As soon as we come to have more important interests to manage, we find, that by pleasing one man, we almost certainly disoblige another, and that by humouring an individual, we may often irritate a whole people. The fairest and most equitable conduct must frequently obstruct the interests, or thwart the inclinations of particular persons, who will seldom have candour enough to enter into the propriety of our motives, or to see that this conduct, how disagreeable soever to them, is perfectly suitable to our situation. In order to defend ourselves from such partial judgments, we soon learn to 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. If, when we place ourselves in the situation of such a person, 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 behaviour, and regard ourselves, in spite of the censure of our companions, as the just and proper objects of approbation. (TMS, III.2.32, n/22, pp. 151-152)[27]

In other words, we begin by trying to gain the approbation, the *praise*, of everyone. Any time we endeavour to see ourselves from without, we do this to predict what will please, and how we may avoid offending, those around us. This project, however, soon reveals itself as being 'impossible and absurd', because the reactions of actual spectators we thus try to predict turn out to be inconsistent—both across time, and across spectators. For example, by doing something we are sure will please one parent, we discover, to our horror, that we displease the other. Or, by doing something which has previously pleased our parents, we now find that they are not at all pleased with us. When, in addition, we begin to have projects of our own to pursue, this chasing after the approval of actual spectators becomes completely insupportable—presumably because we are then sure to incur the displeasure of those whose *own* projects we are then no longer engaged in furthering, and far from certain of pleasing anybody but ourselves (as when we are too engaged in

---

27 This paragraph is part of a longer one reproduced in a footnote in the Cambridge edition of TMS. It was present in all editions of TMS until the 6th, and then replaced by a different one. You may wonder *why* Smith removed this paragraph (and added others and changed yet others) in the 6th edition of TMS. The short answer is that we do not know. The long answer would involve a discussion of whether Smith changes his *account* of the impartial spectator between the earlier and later editions, or merely changes *emphasis* on things that are there from the beginning. I tend towards the latter view, and develop my own take on this in *Love Redirected*. You may also wonder why I do not simply stick with the 6th edition. The answer is that I am interested in reconstructing the most plausible

play to help set the table). Therefore, we naturally seek for some means of deciding which spectators' judgements to trust, and when.

Noticing that people's reactions typically diverge from how we predicted they would react either 1) because they do not know what we have done, or mistake our motives for doing as we did; and/or 2) because by acting as we did, we somehow harmed their interests; noticing this, we begin to see the discrepancy between our predictions and their judgements, not necessarily as a reflection of our failure to predict, but as a reflection of their failure to judge as they *would* have, if they had only been sufficiently informed and indifferent to judge us fairly and impartially. We begin, that is, to use the hypothetical spectator, not just as a tool to predict, but as a means to second-guess the judgements of actual spectators. The adjectives that arrogate to the hypothetical spectator thus employed are, naturally, 'impartial' and 'well-informed', and so the hypothetical spectator becomes the 'impartial and well-informed spectator' (TMS, III.2.32, p. 150).

What was originally a prediction of what actual spectators *will* think thereby becomes a judgement of what they *would* think, if only they were 'quite candid and equitable', that is, if only they were *better* spectators. But what they would think if they were better spectators is what (we think) they *should* think. Thus, from our ardent pursuit of the 'impossible and absurd project of trying to gain everybody's good will and approbation' arises the revolutionary thought that 'everybody' might sometimes be *wrong*, and wrong as measured against a standard inherent in the act of *judging* itself.

Of course, this discovery cuts both ways, since, in recognising that moral judgements are only really valid if made by a spectator both well-informed and impartial, we must necessarily recognise that our own moral judgements, whether of others or ourselves, are also only valid to the extent that *we* are both well-informed and impartial when reaching those judgements.

Therefore, moral judgement is no longer only a matter of trying to sympathise, and, if we fail to do so, to condemn the poor persons principally concerned as necessarily blameworthy. We ourselves have been in that position, and felt the unfairness of being blamed by spectators that were ill-informed and/or partial. Just as we, in such a situation, can solicit the idea of the impartial spectator, and take solace in the thought that, had she been present, we should have not suffered such injustice, we must now recognise, if we are honest, that our judgements of others can only claim to be valid if they too would be endorsed by an impartial spectator. We recognise, that is, the difference between something merely being blamed or praised, and something being *worthy* of that blame or praise.

The two may overlap, so that what is worthy of praise is also actually praised, or what is worthy of blame is actually blamed. But they certainly need not, and we know full well that

praiseworthy actions go unrewarded, and blameworthy ones unpunished. Even worse, we know praiseworthy actions that are blamed (as is so often the case for whistle-blowers), and blameworthy ones that are praised (as is so often the case for the rich and powerful).

We cannot, that is, implicitly trust the judgements of society as expressed in praise and blame. Nor, however, can we implicitly trust our own judgements, for the moment we realise that a good moral judge must be both well-informed and impartial, we know full well that we ourselves sometimes, or rather frequently, fail to fully inform ourselves before passing judgement on the actions of others, and, no less embarrassingly, fail to regard ourselves with the dispassionate eyes of an impartial stranger when passing judgement on our own. We realise, that is, that we may fail, by our own lights, to be the kind of moral judge we think we should be.

The great strength of Smith's roundabout, empirical approach is (thus) to raise awareness about failures in our moral judgements and, possibly, give us tools to improve, *without* having to supplant our diverse and partly contradictory moral intuitions with a rationally coherent but intuitively defective synthetic criterion of rightness, *and* without requiring us all to agree *what* an ideal impartial spectator would think. All we need is to recognise ourselves—our own standard of judgement and our own failures to live up to that standard—in Smith's descriptions, and we are put in position to improve ourselves by our own lights. The indirect normativity in Smith is therefore inherent his model of moral judgement, which, in the dual sense of modelling and serving as an ideal, shows us what we should aspire to, or, rather, what we already aspire to, and how we are most apt to fall short of that aspiration.

I.iv. By Our Own Lights

The distinctive normative structure of TMS has one further implication for my project. Since the normativity in Smith's model of moral judgement is dependent on the validity of the descriptive or empirical elements of that model, those elements, whether understood as the products of first-person phenomenological or third-person psycho-sociological investigations, should be re-examined in light of findings from later empirical investigations into the operation of our 'moral judgement machinery' (Bruni, Mameli, & Rini, 2014; Haidt, 2001).

However, given the state of the debate on the nature and role of empathy/sympathy in moral judgement (see above), I will not go into any great empirical detail about how Smith's basic model of our capacity for sympathy relates to present concepts (but see e.g., Nanay, 2010). I only note that Smith's description of our capacity for sympathy is seen as a valuable resource for critique and synthesis of research on empathy (see the introduction in Maibom, 2017), so that, whatever the

ultimately correct answer to the question of what role something like sympathy does play or should play in moral judgement, then, if it plays any role at all, Smith's model is a useful framework from which to launch one's explorations into all the ways in which we fail to live up to our own standards. Betting on sympathy playing *some* role in our moral judgements, I therefore mainly concern myself with what modern psychology tells us about all the ways in which we fail to live up to our own standards.

And they are legion. Indeed, if there is one thing the development, since the days of Smith, of empirical psychology in general, and social psychology in particular, has taught us about our moral judgements, it is that we are very good at failing to live up to (what we thought were) our own standards. The most pressing question, therefore, is not what standard we should uphold, but how we can improve our ability to uphold standards we already endorse.

In the normatively inflected debates of how bias, prejudice and lack of perspective-taking drives spirals of conflict (K. A. Kennedy & Pronin, 2008; Ross & Ward, 1995), the problem is not which criterion of right action to employ, but how to improve our ability to live up to the standards we already endorse in order to achieve results that would appear good on any normative ethical theory worth its salt. If these debates need any kind of normative framework, they need something like the indirect normativity in Smith, where, from recognising new (and old) sources of by-our-own-lights failures, we can generate new ways to by-our-own-lights successes.

This dissertation, then, is truly a work of normative ethics 'by your own lights', and I wholeheartedly accept the limitation that those who honestly and without self-deceit find nothing within it to pique their own sense of praiseworthiness can put it down unaffected, without there being any grounds for me thinking them blameworthy in so doing. That said, for those who *do* find something here with which to praise or reproach themselves, there should also be something to arouse their interest in the possibility, and possibly also the means, of improving themselves.

Beyond this, the beauty of a by-your-own-lights kind of *philosophy* is that its practitioner need never feel painted into a corner by his or her speculations; need never assert things that he or she does not really believe, just because they appear to be entailed by previous claims provisionally made; need never, that is, *develop* the kind of stupid consistency that is the 'hobgoblin of little minds' (Emerson, 2005). The practitioner of by-your-own-lights philosophy need never do this because this kind of philosophy never pretends to be the whole story (see T. Chappell, 2014b). Its worth lies not in what it forces you to think, but in getting your thinking going.

II. The Four Articles

II.i. Moral Tuning

How, then, are the four articles supposed to get your thinking going on the subject of how to be a good sentimentalist? I wrote earlier that this dissertation deals with questions related to how we, in general, can best go about the complex and never-ending task of trying to figure out what we should do—and then do it. The first article, *Moral Tuning* (hereafter MT for short), deals with the first part of this question, namely how we can best go about figuring out *what* it is that we should and should not do. The topic of this article, then, is moral knowledge. The background to this article is a debate in the secondary literature on TMS on whether Smith succeeds in providing an account of how we can criticise, not only the *application* of a given standard of propriety, but the standard *itself*. Samuel Fleischacker puts the point succinctly:

> [I]n correcting for the passions and interests and misinformation of our friends and neighbors, we turn to an idealized version of our friends and neighbors who uses the same standards of moral judgment as they do. The impartial spectator is disinterested, well-informed and 'candid' (TMS, 129), but is otherwise just like actual, partial spectators. It is built out of actual spectators; it is built, in particular, out of the basic reactive attitudes, the basic modes of moral judgment, that our actual friends and neighbors have. (Fleischacker, 2011, p. 28)

Remember what I wrote above about the genesis of the idea of the impartial spectator: Faced with the inconsistent judgements of actual spectators, we begin to second-guess them by imagining what the reactions of an impartial and well-informed spectator would be in their place. Our idea of *what* those reactions would be, however, are themselves shaped by our experience of the reactions of actual spectators in the society in which we have been raised. What counts as appropriate is something we learn (and crucially, as we shall see, help shape) through observing and participating in a vast number of sympathetic exchanges from early childhood onwards. In other words, the substantive content of our moral judgements will in some sense and degree be defined by *custom*, be a matter of *convention*. The impartial spectator, or so the worry goes, may well be able to correct for partial and ill-informed applications of this conventional standard, but cannot correct for errors in that standard itself. As Fleischacker goes on to note:

> If the moral standards, the basic moral sentiments, of a society are profoundly corrupt—if a feeling of contempt for Africans or hatred for Jews or homosexuals, say, has been taken for a moral feeling, and a society's judgments of these people's actions have been comprehensively

skewed as a result—the impartial spectator within each individual will share in, rather than correcting for, that corruption. (Fleischacker, 2011, pp. 28-29)

The worry, in other words, is that we may merely be policing the application of a standard of propriety that should itself be rejected or at least revised. That is the worry, and the question is whether the idea of the impartial spectator gives us the tools we need for such revision.

Opinions among Smith-scholars are divided on this point (for a partial review, see Fricke, 2011a, pp. 47–49). Some, like Carola Freiin von Villiez, argue that the idea of the impartial spectator can in fact sort the wheat of social norms 'that have all the properties of moral norms proper' from the chaff of 'mere socio-moral conventions' (Freiin von Villiez, 2011, p. 42).

Others, like Fonna Forman, think that a universal standard of morality may well be germinating in the depths of Smith's model, but if so, it is not the idea of the impartial spectator that will bear fruit, but what she describes as the '*summum malum*' of 'the human aversion to cruelty'—an aversion that according to her reading is '*insulated from* and *prior to* the particularity of the moral sentiments' (Forman-Barzilai, 2010, pp. 24, 233).

Yet others, like Fleischacker, think Smith ultimately, if interestingly, fails to give us an adequate account of how we achieve the critical distance necessary to revise our own standard of propriety .

Fleischacker's way of *framing* Smith's failure, using the twin devices of what he calls an anthropological and a philosophical approach to morality (Fleischacker, 2011, p. 24), is especially interesting. The main difference between these two approaches is that the anthropological approach takes as authoritative a given society's own conception of which norms are morally binding, while the philosophical approach aims to separate out all and only those norms that can be rationally justified (Fleischacker, 2011, pp. 20–23). The anthropological approach has the advantage of taking seriously what most people in fact think are moral norms, but the weakness of not being able to criticise societies with moral conventions 'no rational, reflective person could endorse' (Fleischacker, 2011, p. 39). The philosophical approach has the advantage of being able to identify and criticise bad kinds of 'taboo', but the weakness of writing off much of what people in fact think are moral norms as (at best) irrelevant (Fleischacker, 2011, p. 39). Neither is anything close to perfect, and so 'we ought to want to bring the philosopher's and the anthropologist's views of morality closer together' (Fleischacker, 2011, p. 39). TMS is an important resource in that regard, thinks Fleischacker, not because Smith succeeded in bringing the two together, but because his

attempt to do so speak to problems that face moral philosophers to this day (Fleischacker, 2011, p. 40).[28]

Smith takes seriously that idea moral 'just is the social practice by which people correct one another for not adequately living up to their society's standards of conduct' (Fleischacker, 2011, pp. 24–25), while at the same time sorting misguided norms and perverted practices from those that can be justified from the perspective of an impartial spectator. The problem, argues Fleischacker, is that Smith, precisely because he takes seriously the cultural embeddedness of all standards of propriety, is unable to give an adequate account of *how* the idea of the impartial spectator can help people in a given society discover and correct for mistakes and perversions in their own moral standard.

Fleischacker allows that the impartial spectator can help us *universalise* our moral norms, and, that, in so doing, we may also discover certain kinds of bigotry, like that which justified slavery in the southern United States in the 19th century (Fleischacker, 2011, p. 35)—or, to take a modern example from the same country, the institutional racism and sexism addressed in the Black Lives Matter and #Metoo movements, respectively. Discrimination based on perceived race or gender relies on drawing purportedly morally relevant lines around these categories, and can as such be challenged from the perspective of an impartial spectator who, as Freiin von Villiez puts it, takes into consideration the perspective of *all* people within a common dimension, and hence forces those who discriminate to provide a justification for their discrimination that can be accepted by all members of a community (or else abandon their discrimination) (Freiin von Villiez, 2011, p. 41).

However, continues Fleischacker, even a realisation that we must treat all humans equally does not decide *how* we should treat them (Fleischacker, 2011, p. 36). His case in point: the killing of infants in ancient Greece, a practice Smith himself spends a significant amount of time examining and condemning in TMS (V.2.15, p. 246). For, while we can certainly share Smith's abhorrence at this practice,

> [t]he Greeks … presumably believed that any child could be exposed. They were willing to extend their practice of infanticide equally to all human beings. Knowing that human beings are all equal is therefore not enough: we need also to know what sort of treatment befits these equal beings. (Fleischacker, 2011, p. 35)

One of the things that makes Smith's criticism of Athenian infant exposure so interesting is that he explicitly criticises two of history's greatest philosophers—Plato and Aristotle—for failing to realise the wrongness of that practice:

---

28 Fleishcacker assumes that there is a large gap between morality as described by anthropologists and philosophers (and he is not alone in so doing), but see Michele M. Moody-Adams' thoughtful critique of this assumption in her (2002).

Aristotle talks of it as of what the magistrate ought upon many occasions to encourage. The humane Plato is of the same opinion, and, with all that love of mankind which seems to animate all his writings, no where marks this practice with disapprobation. (TMS V.2.15, p. 246)

These great thinkers had let themselves be misled by 'uninterrupted custom' (TMS, V.2.15, p. 246), but Smith does not see this as an excuse: Their doctrines 'ought to have been more just and accurate' (TMS, V.2.15, p. 246). In other words, Smith both upholds a trans-historical moral judgement on the wrongness of 'the murder of new-born infants' from 'views of remote interest or conveniency' (TMS, V.2.15, p. 246), and suggests that people, or at the very least philosophers, ought to have realised this.

Presumably, Fleischacker's worry is that Smith's condemnation of Plato and Aristotle cannot be sufficiently justified by appeal to the impartial spectator.[29] However, Fleischacker's supposition that Athenians were willing to extend their practice of infanticide equally to all human beings does not, at least, appear to apply to these two thinkers. As indicated in the Cambridge edition of TMS, Smith may have been referring to a comment in *The Republic* (2000), where Plato has Socrates say the following about the ideal state:

> The children of good parents will be taken, I think, and transferred to the nursing-pen, where there will be special nurses living separately, in a special part of the city [c]. The children of inferior parents, on the other hand, or any deformed specimen born to the other group, will be removed from sight into some secret and hidden place, as is right. (V, 460c)

And to one in *Politics* (1995), where Aristotle decrees:

> As to the exposure and rearing of children, let there be a law that no deformed child shall live. But as to an excess in the number of children, if the established customs of the state forbid the exposure of any children who are born, let a limit be set to the number of children a couple may have; and if couples have children in excess, let abortion be procured before sense and life have begun; what may or may not be lawfully done in these cases depends on the question of life and sensation. (VII.16, 1335b20-21)

Here, however, we see Plato restricting infanticide to the children of 'inferior' parents and 'deformed' children of 'good' parents, and Aristotle restrict it to 'deformed' children in general (although he can also be read as implicitly opening up for the possibility that all infants in 'excess' could be killed if 'established custom' allowed it). Therein lies the germ of a universalising critique of the practice. If the impartial spectator teaches us that, for the purposes of moral judgements, all

---

29 Interestingly, where I read Smith as passing a relatively harsh judgement on Plato and Aristotle here, Fleischacker thinks that Smith 'regards [their acquiescence] as fully understandable if not admirable.' (Fleischacker, 2011, p. 16 n/31)

humans are equally valuable, then there are also no morally relevant ways in which some parents are 'inferior' or some children 'deformed'. Since Plato and Aristotle at least superficially justify the practice of infant exposure by reference to such categories, then the abolition of these categories also invalidates their justification for that practice. The problem is not that they do not know *how* to treat infants, but that they recommend treating some infants barbarously on the basis of a dubious categorisation. Hence, their doctrines both could and ought to have been more just and accurate.

Be that as it may, Fleischacker *is* right to point out that the universal form of the impartial spectator does not determine the substantive content of our judgements *by itself*. That content must, as Freiin von Villiez argues, be supplied by interpretations of the strictly formal '*Grenzideal*' (Freiin von Villiez, 2006a, p. 206, 2006b, pp. 130–134) or 'limiting ideal' of what Smith calls 'exact propriety and perfection' (TMS, VI.III.23, p. 291)—interpretations undertaken by people like you and I, drawing on the resources of our individual and collective experiences.

The central question therefore becomes whether we, in giving such substantive interpretations of 'exact propriety and perfection', are (wholly) beholden to the customs of the society in which we have grown up and into whose moral standard we have been socialised—or, somehow, free to interpret and reinterpret these customs and this standard in light of new experiences and the customs and standards of other times and societies.

In MT, my co-authors and I argue for the latter conclusion: We are in fact free to interpret and reinterpret the standards of propriety dominant in our own society, and free in such a way as to render it possible, if not easy, to discover and reject moral perversions embodied in that standard. However, our route to that conclusion is somewhat circuitous.

Our point of departure is Smith's well-known aptitude for metaphor. However, we are not interested in the 'invisible hand' this time around, but in Smith's extensive and elaborate use of musical metaphors to explain the movements and interplay of the sympathetic exchange.

Charles Griswold (1998) has done a tremendous job of elevating these metaphors from the category of rhetorical flourishes (into which they too easily could be dismissed) to the level of an important feature of Smith's system—indeed, in taking issue with the claim that Smith aims to be Newtonian, Griswold suggests that, if there is any system in TMS, it is more like that of a 'well-composed concerto' (A. Smith, 1982, pp. 204–205; Griswold, 1998, pp. 75, 360).

However, Griswold treats the source domain of these metaphors, the *music*, as an aesthetic *object*. There is nothing wrong with that, but, we argue, it does impose a particular interpretive framework on the target domain, on *morality* (MT, pp. 71-72). The most important constraint of that framework is the one indicated by the words just used: If music is (first and foremost) an object, then morality is also, by virtue of the analogy, (first and foremost) an object. Of course, morality *is*

an object in the sense that moral norms and whole standards of propriety are 'things' picked out by nouns. But morality is something else as well, namely *actions*. And if there was ever a theory of 'morality' that emphasised the acting over the thing, the 'doing' of morality over the codification of doing into norms, it must be Smith's TMS.

Inspired by this and by recent developments in musicology, we propose 1) to see Smith's musical metaphors as elements of an analogical model of morality, and 2) to locate the source domain of that model, not in music as an aesthetic object, but in music as a practice, in what Christopher Small calls 'musicking' (1998). This opens up a number of new interpretive possibilities, but we focus on what this shift might do to our understanding of the problem of conventionality in TMS.

The first upshot of our metaphorical remapping is to change the lexical ordering of morality as object and as practice. In other words, using musicking as the source domain allows us to see morality as (first and foremost) a *practice*, as a kind of 'moral tuning' (MT, p. 81). We pick the element or sub-metaphor of 'tuning' to talk of morality as a practice in part because the closest verb form, 'moralising', has unfortunate connotations, but also because, as we show with a practical example, the act of tuning in to a shared standard fits nicely with Smith's model of the sympathetic exchange: Musicians who want to play together must tune in to each other, and tuning in involves relating both to a conventional standard, such as that which holds A'=440 Hz, and to each other, that is to the particularities of their instruments and the venue in which they play (MT, pp. 68-70). Moreover, in tuning, the conventional standard is itself a product of previous generations of musicians gradually 'tuning in' to a standard that works for the purposes of musicking (within a particular tradition of music). Music, that is, is thoroughly conventional yet unquestionably valuable. Acknowledging the primacy of practice; of musicking over music, of playing over the score; allows us to see musical conventions as simultaneously subservient to the needs and desires of those who engage in musicking and guiding and even binding for the same.

This in turn allows us to look for the wellspring of *moral autonomy*—the thing that (supposedly) allows individuals, even in the grip of a strong conventional standard of propriety, to critically reflect upon and even alter that standard in pursuit of the limiting ideal of perfect propriety and perfection—in the source domain of musicking. We do so by building on Small's main example of a musical practice, of musicking, which is no less than the well-composed concerto favoured by Griswold—but now as the *performing* of a well-composed concerto within the peculiar ritual that is a symphony concert (MT, p. 76, Small, 1998). We argue that the constant need to translate the musical score into movements renders even the strictly regimented concert musician free to interpret that score (MT, pp. 79-80). (Indeed, the very idea, central in the western annotated music

tradition, that one should aim for the note-perfect play-through of a score, or the recreation of the composer's 'original intention', is itself just one interpretation among many possible of what musicking should be.)

Analogically, we argue, our moral autonomy can be found in the constant need to give substantive interpretations of the limiting ideal of exact propriety (MT, pp. 81-84). In this interpretive endeavour, we may draw (more or less freely) on other such substantive interpretations that is on other conventional standards of propriety. The limiting ideal sets the (constantly receding) limit to this inquiry, and our choices will always be guided by our own lights, even as those lights are shaped and reshaped by the interpretations we give. The freedom lies in the *constant* need to interpret and reinterpret, and in thus interpreting, shaping and reshaping the standard of propriety. No standard of propriety *need* ever ossify, and *can* in any case never claim to be the one and only moral truth. Yet, by virtue of the kinds of interactions it allows and facilitates, it may rightfully claim some allegiance and value as a particular, substantive interpretation of the ideal of exact propriety and perfection—just never ultimate value or unquestioning allegiance.

How, then, can we best go about figuring out what we should and should not do? Well, the counsel from our work in MT is to realise that we are *already* always free to interpret and reinterpret any given standard of propriety. In that interpretive endeavour, we can draw on different traditions, different 'styles', guided by the idea of the impartial spectator, and measuring our success or failure by our own lights; the ultimate, if ephemeral, yardstick of normativity. Most importantly, we should realise that *not* questioning established norms and ideals, *not* using our moral and musical autonomy, is, as we put it, 'nothing more than to propagate, borrowing a phrase from Herbert Spencer, 'the rule of the dead over the living'. There is certainly nothing inherently noble in that' (MT, p. 84).

II.ii. Love Redirected

The second article, *Love Redirected* (LR for short), deals with the second part of the question put forth in the introduction, namely how we can best go about *doing* what we already think we should do. The topic of this article, then, is moral motivation. The background to this article is the debate in the secondary literature on TMS on the question of whether Smith succeeds in providing a coherent explanation for the appearance of genuinely moral actions, that is actions that are genuinely other-regarding, either in the sense of not just being the issue of unalloyed selfishness cloaked in a socially respectable garb of good will and beneficence, and/or in the sense of being the result of a genuine consideration of the situation of the person principally concerned (Griswold, 1998, pp. 76–113).

The background to *that* debate, and indeed to Smith's own writing on the topic, is a very old engagement of moral philosophy with the question 'why be moral'. Depending on whom you ask, you will get different answers to what this question is about. I quite like how Plato, in Book II of *The Republic* (2000), has the character Glaucon relate the underlying problem in the form of a parable about 'the ancestor of Gyges the Lydian':

> They say he was a shepherd, and that he was a serf of the man who was at that time the ruler of Lydia. One day there was a great rainstorm and an earthquake in the place where he grazed his sheep. Part of the ground opened up, and a great hole appeared in it. He was astonished when he saw it, but went down into it. And the legend has it that among many marvels he saw a hollow horse made of bronze, with windows in it. Peeping through them, he saw inside what appeared to be a corpse, larger than human, wearing nothing but a golden ring on its hand. They say he removed the ring, and came out. 'The shepherds were having one of their regular meetings, so that they could give the king their monthly report on the flocks. And the man turned up as well, wearing the ring. As he sat with the rest of them, he happened to twist the setting of the ring towards him, into the palm of his hand. When he did this, he became invisible to those who were sitting with him, and they started talking about him as if he had gone. He was amazed, and twisted the ring again, turning the setting to the outside. As soon as he did so, he became visible. When he realised this, he started experimenting with the ring, to see if it did have this power. And he found that that was how it was. When he turned the setting to the inside, he became invisible; when he turned it to the outside, he became visible. Once he had established this, he lost no time arranging to be one of those making the report to the king. When he got there, he seduced the king's wife, plotted with her against the king, killed him and seized power. (II, 359d-360b)

The moral of the story is that people act justly 'on account of public opinion' (358a)—if given half a chance, 'the just man would be acting no differently from the unjust' (360c). What stops us is our 'want of power to do wrong' (359b), that is, our lack of power to take the things we want, and do with others as we please. Justice is not valued in itself, only on account of its effects: We act justly in fear of the social consequences of acting unjustly. As such, the image of morality presented by Glaucon is one of social conformity. Remove the social pressure, give us a ring of invisibility, and we are all, at heart, equally unjust.

Glaucon does not himself hold this view of justice, he tells Socrates, but he is 'dismayed by the unending sound in my ears of Thrasymachus and thousands like him' who proffer it (358c). It is Thrasymachus who, in Book I of *The Republic*, argues (very roughly) that justice is just for show, something the powerful, for their own benefit, impose on the powerless (see T. D. J. Chappell, 1993). Similarly, in Smith's own time, Bernard Mandeville was (in)famous for promoting a 'selfish' account of morality, according to which virtue and vice are inventions of 'wary politicians' for the purposes of controlling and directing the labour of large numbers of people (Mandeville, 2011, p. 80). Through flattering our pride, they have got some of us thinking that it is noble—an expression of our higher nature, of what separates us from all other animals—to restrain the gratification of our appetites; and the rest of us realising that we avoid a world of trouble if we just go about that gratification less openly (Mandeville, 2011, p. 77).

Why be moral? Well, Thrasymachus would say, unless you are a dictator (or happen to find a ring of invisibility), then you have precious little choice in the matter: Being, or at least *pretending* to be moral, is what you have to do to survive in society—besides, would Mandeville add, 'private vices' make for 'public benefit', not least because the desire for praise causes conspicuous consumption, from which everyone benefits (Mandeville, 2011, pp. 73–74).[30]

Those in the other camp, like Glaucon in *The Republic* and Smith in TMS, think that we behave morally (without the scare quotes this time) because doing the right thing is the right thing to do. We do not merely pretend to be moral, we *are* moral, or, at last, we *desire* to be (*The Republic* 361b, TMS III.2.7, p. 136). The corollary is that people with genuinely moral motives will (sometimes) do the right thing even if it does not bring them *anything* in terms of external benefits, and even if it will cost them dearly.[31]

---

30 A claim Smith, despite all his other disagreements with Mandeville, would himself make in very similar terms: TMS VI.1.10, pp. 214-216. See also Smith's *Letter to the* Edinburgh Review, where he comments on both Mandeville and Rousseau (and first funnily misspells Mandeville as 'Dr. Mandevil'): (A. Smith, 1982, p. 250)

31 The example of whistle-blowers comes to mind, given the high price they often pay for their choices (Peters et al., 2011)—less dramatically, but no less important, is the fact that children, from a young age, have been shown brave social exclusion for the sake of acting justly (Engelmann, Herrmann, Rapp, & Tomasello, 2016).

Some might think that even this kind of moral motivation cannot fulfil the most rigorous philosophical demands that moral actions be *truly* other-regarding and independent of *any* kind of benefit on the part of the agent—including, that is, inward satisfaction at having done the right thing. To that I would reply: If the inward satisfaction of having done the right thing motivates you to do the right thing independently of any external reward you might get from so acting, and even in spite of the external costs you will likely incur, then your motivation deserves the appellation 'moral' *if anything does*. Good on you![32]

This was also Smith's view. His name for the genuine kind of moral motivation was the 'love of praiseworthiness'—the desire, not to actually *be* praised, but to know oneself *worthy* of praise whether such praise is forthcoming or not.[33] Since praiseworthiness, in Smith's model, is the mark of the moral, the *love* of praiseworthiness (together with the corresponding 'dread of blameworthiness') is the motive force of genuinely moral actions. Why be moral? Smith's descriptive answer is that we (evidently) *want* to be; we desire not only to be praised but to be worthy of praise. As for the *validation* of this desire, Smith is less clear than we might expect him to be given the prominence of the love of praiseworthiness in his rebuttal of selfish theories of morality. Specifically, he does not say overly much of how, exactly, the love of praiseworthiness fits in with the rest of his theory. The secondary literature is also relatively silent on the topic, with a few but notable exceptions. The main exceptions I found were Griswold (1998):

> The self-approbation that derives from knowing oneself to be praiseworthy is a natural outgrowth of the process by which, through sympathy, we approve of others. Were this process impossible, we would be fit only for the 'affectation of virtue' and the 'concealment of vice,' rather than for the 'real love of virtue' and the 'real abhorrence of vice,' and would therefore wish merely to appear to be fit for society (III.2.7). The love of virtue is not the love of the approval of some other person, called the 'impartial spectator,' but of an aspect of ourselves with which we 'sympathize.' At this level it is a question of the self's relation to itself. As we become habituated to observing ourselves from the impartial point of view, our emotions are themselves shaped so as to diminish the motivation to act from self-love alone, and our loves are consistent with our love of virtue; for we are impartial spectators of ourselves. The love of virtue is an outgrowth of sympathy. (p. 133);

Lauren Brubaker (2006):

---

32 Kantians might disagree, but when it comes to being suspicious of any kind of 'inclination' towards moral action, I think the crucial distinction is not between being inclined and not being inclined, but rather between 1) being inclined to, for example, render help *yourself* and 2) seeing help rendered regardless of who renders it (thanks to Robert Hepach for this distinction). The latter kind of inclination deserves the appellation 'moral'.

33 Or, as Debes suggests, the desire to be '*worthy to be cared about*' (2016, p. 204, emphasis in original).

Smith has a nuanced understanding of the perfection of human nature. This is clear from his accounts of the gradual development of the idea of an impartial spectator from our experience of the original actual spectators of our actions, and of the development of the love of virtue as the basis of morality rather than merely the love of praise, to name only two of the most crucial discussions. While these start in our immediate natural sentiments, their maturation requires judgment and reflection. Through such 'slow, gradual and progressive work' we all develop to some extent an 'idea of exact propriety and perfection' (TMS VI.iii.25, 247). As a result our evolving understandings of the impartial spectator's viewpoint and of the love of virtue for itself come to restrain some of our other natural sentiments. These developed or perfected sentiments can even come into conflict with the natural sentiments from which they first arose. The impartial spectator can render a judgment contrary to the judgment of actual spectators, the original or natural tribunal. The love of virtue or the desire to be praiseworthy can sustain us even under conditions in which we lose the praise of actual spectators, our original or natural desire. (p. 181);

and Ryan Patrick Hanley (2009; see also Maria A. Carrasco, 2011, p. 22):

First, his distinction [between love of praise and of praiseworthiness] testifies to the persistence of his concern with one of the aspects of commercial society that Rousseau, as we have seen, found most troubling: its propensity to separate concern for appearances from concern for character, *paraître* from *être*. Smith's solution to this problem rests on his postulation of a love of praiseworthiness both logically and temporally prior to a love of praise, the former being necessary to mitigate the excesses of the latter. Herein lies Smith's Rousseauan solution to the characteristically Rousseauan problem identified earlier. Having already established that individuals in commercial societies are uniquely sensitive to the opinions of others, Smith now makes the Rousseauan claim that nature in its wisdom antecedently invested man with a love of praiseworthiness capable of withstanding and mitigating civilization's corruptions (cf. TMS I.iii.3.7). (p. 140)

While all three put forth a view on the relationship between the love of praiseworthiness and the rest of Smith's model (notably the love of praise), none go into any great detail about this relationship, nor do any of them deal with the apparent tensions in Smith's own statements about it. That there *is* such a tension is evident from the divergent interpretations given by, on the one hand, Griswold and Brubaker—who see the love of praiseworthiness as in some sense developed from the love of praise—and, on the other, Hanley—who sees the love of praiseworthiness as 'both logically and temporally prior to the love of praise'.

Part of the problem is the shift in Smith's emphasis on the independence of the love of praiseworthiness from the love of praise between the 2nd and 6th editions of TMS (Raphael, 1975, p.

94, 2007, pp. 44–45). The Griswold-Brubaker interpretation is most directly supported by what Smith says of the love of praiseworthiness in earlier editions:

> But though this tribunal within the breast [the idea of the impartial spectator] be thus the supreme arbiter of all our actions, though it can reverse the decisions of all mankind with regard to our character and conduct, and mortify us amidst the applause or support us under the censure of the world; 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. (III.2.32: 152, n/22).

Read in context, Smith is saying that the love of praiseworthiness, though it may appear something entirely distinct from the love of praise, is in reality derived from the more basic or 'original' love of praise (TMS, III.2.6 p. 135, LR, p. 108).

The Hanley-interpretation, by contrast, is supported by what Smith writes in the 6th edition. Here, Smith replaces the whole passage in which the above paragraph appears with one that says, among other things, the following:

> The jurisdictions of those two tribunals are founded upon principles which, though in some respects resembling and akin, are, however, in reality different and distinct. 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. (III.2.32, p. 150)

In an earlier paragraph also added to the 6th edition, Smith writes:

> The love of praise-worthiness is by no means derived altogether from the love of praise. 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. (III.2.2, p. 132)

The difference is remarkable. Smith goes from describing the love of praiseworthiness as 'in a great measure derived' to 'by no means derived altogether' from the love of praise. Nevertheless, there is no actual *contradiction* between these two ways of putting it; the love of praiseworthiness can be *both* in a great measure *and* not altogether derived from the love of praise at the same time—this is, strictly speaking, a difference in emphasis rather than substance. Still, the fact that Smith shifts his emphasis so dramatically invites the question of what exactly the relation between these two loves is supposed to be. What was missing, I thought, was a thorough discussion of the degree to which the love of praiseworthiness could be said to be derived from that of praise—a discussion, more broadly, of the place of the love of praiseworthiness in Smith's account of moral psychology. LR is that discussion.

Griswold's point about the love of praiseworthiness being 'a question of the self's relation to itself' is central to my understanding of the relationship between this love and the 'original' love of praise. In LR, I argue that the love of praiseworthiness should be understood as the *desire to achieve* mutual sympathy with oneself that is mutual sympathy between the spectator and agent aspects of ourselves. If, moreover, we understand the spectator aspect of ourselves—the idea of the impartial spectator, the 'man within the breast, the great judge and arbiter of [our] conduct' (TMS, III.2.32, p. 150)—as an image for our *conscience*, then the love of praise is the desire to act in accordance with one's conscience, that is to do what one believes is the right thing to do (LR, p. 52).

After LR was published, I was made aware of an earlier text by Michele Bee (2014) that discusses the relationship between the love of praise and the love of praiseworthiness in similar terms. I can only apologise for not finding that text sooner, and I hope that the present acknowledgement can help remedy the impression I have given of being the first to deal exhaustively with this topic. Bee argues persuasively that the fundamental motive to moral action in Smith is 'self-love', which Bee construes as the 'mutual sympathy between us and ourselves' (2014, p. 20), that is, between the spectator and agent aspects of our split, self-regarding selves. This is in line with Griswold, and also my own conclusion in LR. In matters of the *structure* of self-love/love of praiseworthiness, that is, Bee and I agree with Griswold and Brubaker.

There is, however, a difference in emphasis in our arguments. Bee is interested in social motivation more broadly whereas I am interested in moral motivation more narrowly. Moreover, and perhaps because of this difference in emphasis, we also seem to end up assigning different *ontogenies* to self-love/love of praiseworthiness. Bee appears to view self-love as the *foundation* for human sociability (Bee, 2014, pp. 23, 34, 35, but cf. the qualifier 'mature' on p. 30), which, paradoxically, would align his interpretation with that of Hanley when it comes to ontogeny (see Bee, 2014, p. 12). In contrast, I argue that the love of praiseworthiness is an ontogenetically later development, and that the foundation for human sociability is the original love of praise, the desire to be approved of by others (LR, pp. 97-98). The love of praiseworthiness, I claim, is the original love of praise from actual spectators *redirected* towards the hypothetical praise of the impartial spectator (hence the title, LR, p. 101); an interpretation that is in line with those of Griswold and Brubaker both structurally and ontogenetically. Truly moral motivation may be derived from purely selfish motivation.

This interpretation is, I argue, supported by a close reading of the different editions of TMS (LR, pp. 97-101). However, given that TMS is to a large extent a descriptive endeavour, I argue that an exegetical argument about the source of our love of praiseworthiness in the love of praise should

be checked against modern empirical research on the development of moral motivation from infancy onwards (LR, p. 101). To that end, I also engage with recent psychological research on the genesis of pro-social motivation in infants and (proto-)moral judgements young children to evaluate the plausibility of my construal of the love of praiseworthiness (LR, pp. 101-105).

The introduction of non-verbal paradigms (e.g., Premack & Premack, 1997) has allowed psychologists to trace the beginnings of prosocial preferences to within the first year of life (e.g., Hamlin, Wynn, & Bloom, 2007), possibly even as early as 3 months after birth (Hamlin, Wynn, & Bloom, 2010). What is perhaps most remarkable about these studies is that they appear to upend the traditional view of morality as something a child must learn over many years of socialisation, starting, as it were, from scratch, or, rather, from a 'pre-moral' obedience to authority (Vaish & Tomasello, 2013, p. 279). Instead, these and other studies of prosocial judgement and action in infants and children seem to indicate the existence of something like an inborn moral compass. The idea of an 'original' love of praiseworthiness might be supported by the evidence after all (LR, p. 102).

However, these studies also show that infants and children are highly imperfect moral judges, most notably in their tendency to in-group partiality and social comparison (LR, pp. 102-103). Infants are, in particular, not impartial spectators of *themselves* (LR, p. 103). In other words: although we may have an inborn moral compass *of sorts*, the kind of convention-defying, stand-by-your-beliefs, *impartial* love of praiseworthiness to which Smith appeals in TMS does not come quite as easily as the notion of a moral compass would seem to suggest (parents among you will not find this surprising, I imagine, LR, p. 104). Specifically, it takes about three to five years for children to start viewing themselves from something like the perspective of an impartial spectator, and it takes about eight years to develop a robust desire to act on what one then sees—and, as Smith wryly notes, 'the practice of the longest life is very seldom sufficient to bring [this] to complete perfection' (LR, p. 104; TMS, III.3.22, p. 168).

While this evidence does not offer any direct *proof* of my hypothesis, the results I have found are compatible with my interpretation, and, importantly, appear to be *in*compatible with the interpretation that a richly reflective kind of self-love or love of praiseworthiness is ontogenetically primary to the love of praise. The evidence, that is, supports the exegetical conclusion. What remains to be done is to translate this into practical advice on how we can best go about *doing* what we already think we should do.

This is simple enough: If the love of praise is the original desire for praise from other people redirected towards the impartial spectator, then we presumably also derive some pleasure from finding ourselves praiseworthy (able to sympathise with ourselves), and some pain from finding

ourselves blameworthy (unable to sympathise with ourselves). That we feel pain at finding ourselves blameworthy is evident (TMS, III.2.9-15, pp. 137-142), but I think the positive dimension of self-sympathy has been somewhat neglected. To be sure, feelings of guilt and/or shame can be effective motivators to change our behaviour in accordance with what we think we should do. If my argument in LR is right, however, part of the answer to how to be a good sentimentalist is to cultivate the love of self-sympathy through taking real pleasure in our moral successes, no less than we are pained at our failures.[34]

---

34 To which the cynic Mandeville would no doubt have remarked: 'the humblest man alive must confess, that the reward of a virtuous action, which is the satisfaction that ensues upon it, consists in a certain pleasure he procures to himself by contemplating on his own worth: which pleasure, together with the occasion of it, are as certain signs of pride, as looking pale and trembling at any imminent danger, are the symptoms of fear' (Mandeville, 2011, p. 83). Pride perhaps, but *appropriate* pride if so.

II.iii. The Practical Impossibility Of Being Both Impartial And Well-informed

The third article, with the impractically long title of *The Practical Impossibility of Being both Impartial and Well-informed* (PIBIW), makes a first pass at dealing with the idea of the impartial spectator *itself*, an idea that, as we have seen, is central to the first two articles. The background to this article is a problem pointed out by Fonna Forman (2010), namely that, within the framework of TMS, the two qualities supposedly belonging to an ideal spectator, impartial and well-informed, are to some extent opposed in what they demand of us: To be well-informed—to properly understand the situation and character of the person we judge—we must, typically, be sufficiently physically close to that person to see with our own eyes what they are going through (TMS, I.i.1.4, p. 13; Forman-Barzilai, 2010, p. 142). At the same time, this kind of physical proximity tends to entangle us in precisely the kinds of emotional bonds that hinder an impartial evaluation.

Thus, according to Smith, we must turn to close friends to be understood (TMS, I.i.4.9, p. 28), and to distant strangers to be judged impartially (TMS, III.3.38, p. 178). However, the close friend, who understands our situation only too well, will tend to judge us partially, while the distant stranger, who does not know our situation well enough, will tend to make an ill-informed judgement. In other words, neither friends nor strangers fulfil the ideal of the spectator who is both well-informed and impartial, and it is difficult to see *who* could fit the bill. Therefore, if it really is the case that the ideal of judgement is to be both impartial *and* well-informed, then it seems we have before us an insurmountable task, in as much as attaining these two traits requires us to be both close and removed, both friends and strangers, at the same time.

Forman does not pursue this problem any further, nor have I been able to find any attempts at solving it in the secondary literature. As far as I can tell, there are two ways of responding to the problem as Forman presents it. The first is to deny that it is a problem in any *practical* sense, since, in addition to our own efforts to see ourselves from without, we can consult both friends *and* strangers, thus triangulating something like the reaction of an ideal spectator between the three of us. The problem with this first response is that it is far from clear how this triangulation would take place. How do we disentangle the positive effect of understanding from the negative impact of partiality in the judgements of our friends? How do we know which parts of a stranger's reaction to us is the result of her limpid vision, and which are distortions from her ignorance?[35] It may well be that such triangulation is the closest we can get to seeing ourselves as an impartial and well-

---

35 See also John McHugh's related discussion of a tension between identification and evaluation in TMS (2011).

informed spectator would see us, but simply to dismiss the tension between understanding and impartiality on the basis of the *possibility* of such triangulation seems premature.

The second way of responding to the problem is to note that the entanglement of understanding and impartiality is normally the result of close and repeated physical contact (PIBIW, p. 112).[36] Put simply, friends become friends (in part) through repeated interactions in physical spaces, interactions that *simultaneously* engender mutual understanding and the kind of affective bonds that precludes dispassionate judgement. Thus, the tension we observe between understanding and impartiality, between the roles played by friends and strangers as spectators of our conduct, may be a *contingent* effect of interactions in physical space, rather than a necessary tension between impartiality and understanding. In other words, we can recognise the tension between understanding and impartiality as a real, practical problem affecting our moral judgements, but deny that it is inescapable. If Smith is right that sympathy is based on an act of the imagination, regardless of whether we are friends or strangers, then it seems we could be both impartial *and* well-informed spectators of others if we rely on our powers of imagination to bring their cases home to our own breasts while remaining strangers to them, thus becoming well-informed without the emotional entanglement that precludes indifferent impartiality (PIBIW, p. 112).

The problem with this response, is, as I show in PIBIW, that the tension between understanding and impartiality runs deeper than previously appreciated. Indeed, this tension seems to be a product, not of interactions in the physical space, but of the entanglement of information with space in our *minds*.

My argument for this claim is built upon a psychological theory on the connections between construal level and psychological distance, known as Construal Level Theory (CLT, Trope & Liberman, 2010). Briefly put: Proponents of CLT maintain, on the basis of a wide range of experiments, that the *level of detail* in our thinking of something and the *perceived distance* between ourselves and that thing are *linked*. If we construe something abstractly, we will also tend to perceive that thing as distant to ourselves. If we construe something concretely, we will tend to perceive it as close. Moreover, if we perceive something as close to us, whether in time, in physical or in social space, or in how likely it is to happen (or have happened), then we will also tend to construe that thing more concretely. Conversely, if we perceive something as being far away from us in one or several of these dimensions, our construal will also tend to be more abstract. Smith's favourite metaphor for the imaginary work that goes into sympathy, the metaphor of 'bringing the

---

36 Today, that contact may also take place predominantly or even exclusively in virtual spaces, but the mechanisms appear to be similar (Wilson, Boyer O'Leary, Metiu, & Jett, 2008).

case home', turns out to be an accurate description of how a detailed construal of something also tends to bring that something psychologically closer to us.

And so what, you may wonder? Well, as it turns out, construal level and psychological distance also affects our moral judgements (PIBIW, pp. 117-118). There are, that is, clear experimental indications that we judge moral transgressions differently depending on whether we construe them abstractly or concretely, and, correspondingly, whether we perceive them as distant or close to us (Mårtensson, 2017). Notice the implication of this: Bringing the case home, which was supposed to be a way of keeping our disinterested distance while also becoming sufficiently well-informed, necessarily involves a reduction in *psychological* distance, which in turn affects our moral judgement (typically by making us more lenient, but if we are emotionally entangled in the case, so that bringing the case home inflames our resentment, our judgements may also become harsher than they otherwise would be, PIBIW, p. 118).

Bringing someone's case home to our own breast, then, makes us *partial* to—or, sometimes, biased against—that someone, even if she was initially a perfect stranger to us. In other words, the tension between understanding and impartiality is not merely an accidental by-product of interactions in physical space. The tension appears to be rooted in the very functioning of our minds. Hence, the tension between understanding and impartiality appears all but inescapable.

As I intimated above, we might be able to deal with this inescapable tension in practice through a process of triangulation between our own efforts at seeing ourselves from without, our friends' overindulgence, and the insensitivity of an impartial stranger. However, the very idea of the impartial and well-informed spectator plays such a central role in Smith's model, not to mention in the solutions proffered to the problems surveyed in the two first articles of this dissertation, that to simply leave it in place, after acknowledging the deep tension between the two qualities this ideal asks us to aspire to, strikes me as not very satisfactory.

Moreover, the whole point of the impartial spectator was that this was not supposed to be some philosopher's invention, some abstract and unreachable ideal. The figure of the impartial spectator was supposed to be a person we were all familiar with, and the idea of this spectator as an ethical ideal for which we should aspire something achievable, something within reach—at least for some of us, some time. To admit that the impartial spectator is 'essentially Janus-faced', as Forman puts it, and still go on to hold it up as the ideal for which we should strive seems somehow disingenuous. For, if the impartial and well-informed spectator really is an unreachable ideal, why settle for second-rate idealism? Why not go for something like Roderick Firth's 'ideal observer' (1952) or Richard M. Hare's 'Archangel' (1981, p. 44)? At least they make the most of their ideality (PIBIW, p. 121).

The right response is not, however, to abandon the ideal of the impartial and well-informed spectator. Smith's choice of a non-ideal ideal was, I think, essentially right. Only, later research has revealed that we are *still less ideal*, or non-ideal in different ways, than Smith recognised (PIBIW, p. 121-122). The correct response, I suspect, is not to abandon our ideal, but, somehow, to *remould* it.

II.iv. The Partially Impartial Spectator As And Ethical Ideal

The fourth article, *The Partially Impartial Spectator as an Ethical Ideal* (PISEI), not only continues the trend of long titles, but also makes a second pass at dealing with the ideal of the impartial spectator. The argument in PISEI is completely independent of that in PIBIW, down to the empirical research on which it is based. Nevertheless, the upshot of my argument in PISEI can also be taken as an answer to the question implied by the conclusion of PIBIW: If we must remould the ideal of the impartial spectator, *how* should we do it?[37]

The empirical results on which my answer is based comes from the long and prolific tradition within psychology (especially within social psychology) of studying how our reasoning is impacted by various *biases*. Two biases are particularly relevant to my argument in PISEI: the so-called 'naïve realism' bias (Ward et al., 1997), and the 'fundamental attribution error' (Ross, 1977). The first points to our tendency to assume that our perception of the world is direct and uncomplicated, and the second to our tendency to explain other people's actions by reference to their character rather than the situation they are in (note the 'other': the whole point is that the opposite pattern holds for our explanations of our own actions).

These biases are relevant for the question of how to be a good sentimentalist because 1) our naive realism leads us simply to *assume* that any impartial spectator would agree with us, and 2), when someone disagrees, we tend to explain that disagreement by thinking that they simply fail to be impartial spectators (PISEI, pp. 130-133).

Additionally, I review evidence of how motivated reasoning—reasoning governed either by the desire to find the right conclusion, no matter what it might be, or the desire to reach a particular conclusion, no matter what the 'right' one might be—impacts our striving to be more like impartial spectators (Kunda, 1990; PISEI, pp. 133-135). What the evidence reveals is that, in situations where we are antecedently invested in some particular conclusion—such as when we are trying to be impartial spectators of ourselves—all the extra effort we put into trying to reach the right conclusion tends only to go towards building the case that our original conclusion was right all along (PISEI, p. 134).

Put together, evidence from research on biases and motivated reasoning reveal that *trying* to be more like impartial spectators will frequently backfire, merely appending to our pre-existing bias an unfounded confidence that we were right all along (PISEI, p. 135). If, then, we want to be more

---

37 Therefore, the background to this article, to the extent that one can say there is any in the field of Smith-studies, is that which is made up of PIBIW and the sources mentioned in the summary of that article above.

like impartial spectators, we cannot simply try harder, but must, somehow, *try smarter* (PISEI, p. 137).

As I go on to argue, 'trying smarter' must involve shifting our focus away from our own construal of what any impartial spectator would think, and back towards what actual other spectators and persons principally concerned think (PISEI, pp. 137-138). 'Back towards' because this is how it all starts out: the genesis of the idea of the impartial spectator is our desire to be approved of by others (and to avoid their disapproval). The impartial spectator, as laid out above, is something we 'invent' to anchor us against the changing winds of other people's opinions of us (PISEI, p. 129). Somewhere along the way, we become too reliant on what we imagine this hypothetical spectator would think. Because of our naïve realism and tendency to commit the fundamental attribution error, the impartial spectator gradually becomes *us*, instead of the other way around; becomes a partial spectator under another name, instead of serving as a real ideal for which we can strive.

Untying this knot requires, or, so I argue, that we supplement the ideal of the impartial spectator with a second, foreground ideal in the form of the 'partially impartial spectator' (PISEI, p. 140). To counteract our naïve realism, our naïve assumption that we are right, the ideal of the partially impartial spectator embraces the possibility that others might *also* be right; embraces, that is, the possibility of several, incompatible, and *valid* construals of the same situation (PISEI, pp. 139-140). It is, I argue, only through thus undercutting our naïve realism by an embrace of (a kind of) relativism that we can hope gradually to develop the courage and sophistication to admit, as a point of departure, that we might not be right after all. In other words, even if you only want to rid yourself of your naïveté, becoming, as it were, a *sophisticated* realist (or sophisticated *anything*), you will achieve this more readily by *aiming* to be a (kind of) *sophisticated relativist*. Put differently: In order to approach closer to the background ideal of the impartial spectator, we should aim for the foreground ideal of the partially impartial spectator (PISEI, p. 140).

Ultimately, then, the last answer to the question of how to be a good sentimentalist that I offer in this dissertation is the following: For reasons both of humility—we cannot be both fully informed and completely impartial (PIBIW)—and efficacy—aiming to be impartial spectators will often take us further away from that goal (PISEI)—, we should aim to be partially impartial spectators; moral actors who recognise both the ever-present potential for improvement in what are necessarily only partially impartial moral judgements, and that, in order to improve, we need to listen to those who disagree with us.

III. Conclusion

The four articles, seen under one, go some way toward answering the question of how we can be good sentimentalists. First, in *Moral Tuning*, my co-authors and I argue that moral autonomy consists in realising the freedom inherent in the constant need to translate norms into action, and in so doing, to interpret and reinterpret, not only the actions, but the norms themselves. Being a good sentimentalist thus begins with realising that you are free to question and reshape the moral standard that you and those around you abide by.

Second, in Love Redirected, I argue that genuine moral motivation comes from redirecting our desire for praise from other people towards the ideal(l) of the impartial spectator. This redirection requires, not just negative, but positive emotional reinforcement, and so becoming a good sentimentalist involves taking pleasure in your moral successes, no less than you are pained at your failures.

Third, in *The Practical Impossibility of Being both Impartial and Well-Informed*, I argue that we should recognise the tension between understanding and impartiality as a fundamental feature of our striving to be more like impartial spectators. To be a good sentimentalist, you must therefore recognise your limitations, and give up on the illusion of ever being fully understanding and perfectly impartial at the same time.

Fourth, in *The Partially Impartial Spectator as an Ethical Ideal*, I argue that, despite our frequent failures to live up to the ideal of the impartial spectator, trying harder is not likely to get us any closer, and may in fact make matters worse. Recognising our limitations, we should aim for something less ideal, more achievable, and, most importantly, *humbler*, namely to be partially impartial spectators.

Being a good sentimentalist thus beings with realising our freedom to interpret, continues in our taking pleasure in our moral successes, pauses at the realisation that we will never be truly impartial spectators, and ends with a commitment to continued improvement under the lodestar of the ideal of the partially impartial spectator.

Part II: Articles

I. Moral Tuning[†]

Sveinung Sundfør Sivertsen, Jill Halstead, and Rasmus T. Slaattelid

Abstract: Can a set of musical metaphors in a treatise on ethics reveal something about the nature and source of moral autonomy? This article argues that it can. It shows how metaphorical usage of words like *tone*, *pitch*, and *concord* in Adam Smith's *Theory of Moral Sentiments* can be understood as elements of an analogical model for morality. What this model tells us about morality depends on how we conceptualise music. In contrast to earlier interpretations of Smith's metaphors that have seen music as an aesthetic object, this article sees music as a practice. Understood in this way, the analogy allows us to see morality too as a practice—as *moral tuning*. This in turn reveals a novel answer to the intractable problem of conventionalism: moral autonomy consists in the freedom inherent in the constant need to interpret and reinterpret the strictly formal ideal of perfect propriety.

Keywords: music, metaphor, model, ethics, Adam Smith, tuning, practice, musicking, sentimentalism, conventionalism, autonomy.

I.i. Introduction

In *The Theory of Moral Sentiments*, the Scottish Enlightenment philosopher Adam Smith admonishes Plato and Aristotle for failing to condemn the then common practice of killing unwanted babies by abandoning them outside (2002 V.2.14, pp. 246-47). Rightly so, we would venture to say, but if even Plato and Aristotle were blind to such a morally corrupt practice among their contemporaries, how can we lesser mortals ever hope to discover morally corrupt practices among ourselves? In this article, we find an answer by studying Smith's use of metaphor.

Smith's aptitude for metaphor is well established. To many, Smith is known primarily for the metaphor of 'an invisible hand,' which, although associated with his economic theory, first appears in *The Theory of Moral Sentiments* (2002, III.6.10, p. 215). We focus on a different cluster of metaphors in the book, which have music as their source domain and morality as their target. We propose to treat these metaphors as elements of an analogical model for morality, rather than purely rhetorical devices, and suggest that this analogy can tell us something significant about the nature of moral interaction.

What the analogy tells us depends on our understanding of what music is. In contrast to earlier interpretations of Smith's musical metaphors that have seen music primarily as an aesthetic object, we propose to consider music as a practice, as something we *do*. After refocusing the musical metaphor in this way, we reconsider morality as described by Smith, and find reason to think of this too as a kind of practice, as a kind of *moral tuning*. This in turn opens new possibilities for understanding moral judgement and action. We single out and pursue one of these possibilities, showing what the musical analogy offers in terms of resources to understand how individuals can rebel against the conventional norms of society.

I.ii. Modelling Morality

Smith makes extensive use of musical metaphors in *The Theory of Moral Sentiments*.[38] Are the metaphors there for purely rhetorical purposes, or do they also have a more systematic function? Maria Semi's *Music as a Science of Mankind in Eighteenth-Century Britain* (2012) tends towards the latter conclusion by showing how Smith and his contemporaries used music as a model for thinking about philosophical subjects. In keeping with her aesthetic theme, Semi discusses Smith's essay 'Of the Imitative Arts' and does not mention *Moral Sentiments* (Semi, 2012, pp. 93–102; A. Smith, 1982, pp. 176–215). There is, however, an affinity between the way Smith uses musical terms in *Moral Sentiments* to describe morality and the wider tendency of his day to use musical concepts to elucidate philosophical subjects. Building upon Semi's observation that music in eighteenth-century Britain provided a rich source of metaphorical transfer between art and philosophy, we interpret Smith's musical metaphors in *Moral Sentiments* as elements of a model for morality.

In so doing, we must tread carefully. Much misrepresentation of Smith's work has its source in an overemphasis on the metaphor of 'an invisible hand' (G. Kennedy, 2009), and Smith himself, in an essay on scientific explanation, writes disparagingly of those who fall for the temptation of letting a nice analogy become the 'great hinge' upon which everything in a 'system' turns (1982, p. 42).

Proceeding with these cautions in mind, our exploration of Smith's musical metaphors is primarily aimed neither at the question of how Smith himself intended these metaphors to be understood nor at the question of how his contemporaries might have interpreted them. Instead, we use the opportunity afforded by his use of musical metaphors to explore an aspect of moral psychology and ethics that he himself did not adequately address: namely, how we are to free

---

38   For an overview of just how extensive Smith's use of musical metaphor is, see Klein and Clark (2011).

ourselves from the grip of convention. The test for our proposal thus lies in what insights we gain by pursuing it.

I.ii.1. Metaphor And Model

To see the musical metaphors in *Moral Sentiments* as elements in an analogical model is to see them as cognitive tools for interpreting phenomena in one domain (the target) in terms of phenomena in another (the source). More precisely, by *model* we mean a cognitive device that extends our capacity to understand and manipulate complex phenomena by reducing their complexity, often to a very limited set of features. An *analogical model* is a cognitive device that is used to understand phenomena in domains other than its domain of origin. The transfer of a model from one domain to another resembles how metaphor transfers a word from one domain to a new one. Just as the success of a metaphor depends on how well it captures salient features of that for which it is a metaphor, the adequacy of an analogical model depends on how well the salient features of the model fits the salient features of the target domain (Nersessian, 1999, p. 16).

As for what distinguishes the use of metaphor in modelling from their poetic use, Mary Hesse proposed that the 'truth criteria' for the former type of metaphor, while not 'rigorously formalizable,' are generally much clearer than for the latter (1966, p. 169). A similar distinction can be made in the case of philosophy by stipulating that whatever metaphors are central to a philosophical endeavour 'are best viewed as theoretical rather than poetic or rhetorical' (Thagard & Beam, 2004, p. 504). Given how central the musical metaphors are in *Moral Sentiments*, it is reasonable to see them as theoretical. Moreover, because Smith's use of these metaphors is sustained and systematic, they suggest an 'analogical transfer of vocabulary' (Black, 1962, p. 238) from the domain of music to the domain of morality. On this basis, we interpret Smith's musical metaphors as elements of an analogical model for morality.

I.ii.2. Modelling Morality With Music

We can observe the transfer of vocabulary in the systematic mapping by Smith of musical terms such as *pitch*, *beat*, *tone*, *unison*, *harmony*, and *concord* onto some of the most central terms in his sentimentalism, such as *sympathy*, *affects*, *emotion*, *passion*, *society*, and *sentiments*. Most of this mapping, and thus the modelling of morality in analogy to music, happens in a single paragraph in the first part of *Moral Sentiments*, where Smith describes the interaction between agent and spectators:

The person principally concerned . . . longs for that relief which nothing can afford him but the entire *concord* of the *affections* of the spectators with his own. To see the *emotions* of their hearts, in every respect, *beat time* to his own . . . constitutes his sole consolation. But he can only hope to obtain this by *lowering his passion to that pitch*, in which the spectators are capable of going along with him. He must *flatten*, if I may be allowed to say so, *the sharpness of its natural tone*, in order to reduce it to *harmony and concord* with the *emotions* of those who are about him. . . . These two *sentiments*, however, may, it is evident, have such a correspondence with one another, as is sufficient for the *harmony of society*. Though they will never be *unisons*, they may be *concords*, and this is all that is wanted or required. (2002, I.i.4.7, 27, emphases added; see also I.i.3.1, 20; I.i.5.2, 20; and I.i.3.1, 20)

The musical terms Smith uses relate to sound qualities (pitch, tone, unison, concord, and dissonance), rhythmic qualities ('to beat time'), and the wider organisational systems of sound (harmony and pitch).

Pitch, by way of an example, is a term used by Smith in connection with the action of adjustment, as in the phrase 'lowering his passion to that pitch, in which the spectators are capable of going along with him.' His concept of pitch adjustment appears to refer to the way musicians alter the pitch of their instruments, to be in tune with each other, where being in tune is the result of an agreement on, and adherence to, a *shared* notion of pitch. But how are we to understand pitch as a musical concept? What meanings are evoked in Smith's statement? On one level a pitch is a musical note, a particular quality of sound defined by the rate of vibrations producing it. In this light, a pitch can be seen as holding a particular quality, such as high or low. This simple definition implies a stable, universal standard dictated by the laws of acoustics. If pitch is a stable, universal standard, then being in tune should simply be a matter of adjusting to this standard. Agreement on pitch is not, however, simply set via universal acoustical norms. It is negotiated between musicians in particular contexts.

A short practical example might best illustrate the point. A bass guitarist and a pianist are about to perform together. Before they can play, they must tune their instruments. The pianist plays on the piano provided by the performance venue, which was tuned by a professional instrument tuner earlier in the day. The bass player tunes her instrument with an electronic guitar tuner. When both players are satisfied that their instrument is in tune, they play together, at which point the clashing, dissonant sounds make it immediately apparent that their instruments are not in tune with each other. The bassist points out that her instrument is in tune because she has tuned it with the aid of an electronic tuner that measures precisely the number of vibrations per second occurring when each string is played. But the pianist believes the piano is also in tune because it has been

professionally tuned and sounds in tune when she plays alone. Who is in tune? Who should adjust her pitch and how? In this case, the reason the two instruments sound in tune when played alone but out of tune when played together is because the piano had been tuned to a pitch whereby the note a´ (A above middle C) occurred at 432 vibrations per second (Hertz or Hz) in order to be in tune with an old pipe organ located in the same venue. The bassist, on the other hand, had tuned to a pitch whereby a´ occurred at 440 Hz, the modern-day standard programmed into the electronic tuner. Both instruments are in tune, but not with each other. The only way to resolve the pitch problem is for the two players to reach a consensus about which pitch they should use in this context and then make adjustments according to the agreed principle. The players then agree that the bass player would retune to the piano's a´ = 432 Hz, since retuning a piano is much harder than retuning a bass.

Even after retuning to a´ = 432 Hz, however, the two instruments remain slightly out of tune when played together. This final tuning discrepancy occurs because the bass player tuned her instrument according to the natural harmonic series found on the instrument, whereas the piano, as pianos always are, was tuned according to the system of equal temperament. The tuning system of equal temperament had to be developed because if pianos are tuned to the natural harmonic series, the instrument ends up being out of tune with itself across its wide range. When tuning the bass guitar using natural harmonics and starting from a´ = 432 Hz, the G string (96 Hz) will be about 0.22 Hz out of tune with the corresponding G on the piano tuned to equal temperament (96.22 Hz). This would not be audible to all people listening, but to some it would be sufficient to spoil their enjoyment of an otherwise successful musical performance.

This situation describes some of the challenges faced by musicians today, in relation to the concept of pitch, pitch adjustment, and its relationship to tuning. Pitch is a quality of sound, but it is set through an agreement among musicians working in a particular context. In Smith's era, consistent and precise measurement of pitch was difficult, and therefore practices were both extremely varied and always a matter of negotiation. Indeed, in European history pitch has fluctuated widely according to time and place, to the point where 'it is rarely possible to generalize about pitch standards. Even when the exact period and location are known, different kinds of music often had their own standards [of pitch]' (Haynes and Cooke, 2015). In the modern era, pitch has become easily standardized via technologies that can reliably produce and measure pitch. Nevertheless, the modern consensus on pitch expressed as a´ = 440 Hz was only established in 1939 and can be considered 'no less artificial and unrealistic' than the differing pitch standards that preceded it or continue to coexist with it (Haynes & Cooke, 2015). Consequently, pitch should not be thought of as a stable, universal, or unchanging essence, around which adjustments can be made. Even with the broad adoption of a´ = 440 Hz and the advent of electronic tuning meters, pitch

remains socially and culturally negotiated. Rather than a universal and unchanging essence, pitch is a relational process that must necessarily take place whenever people want to sound together.

When Smith talks of the agent 'lowering his passions to that pitch, in which the spectators are capable of going along with him,' we can see that this pitch is not itself independent of the process described: what pitch the spectators will be able to go along with depends on who they are, who the agent is, what passions are involved, and what features of the situation are most salient. Furthermore, even if the interaction in question takes place in a society that has adopted a single standard of 'pitch,' a single standard of propriety, the people directly involved will, like the two musicians in the example above, have to agree, explicitly or implicitly, on a pitch that is suitable for the specific situation. There is, of course, nothing mysterious about this: what counts as appropriate behaviour or sentiment differs according to the situation. Understanding pitch in a way that enables the analogical model to account for this is, in all fairness, a very marginal gain.

However, the analogy between standards of pitch and standards of propriety holds intriguing possibilities for understanding the relationship between morality as a natural and as a conventional phenomenon. On the one hand, pitch can be measured in terms of vibrations per second, and it is defined by the physical constraints of an instrument and the auditory capacities of a listener. Accordingly, pitch is thoroughly natural. On the other hand, pitch, as we have seen, is a matter of social negotiation. Even if a′ = 440 Hz has been widely adopted as a standard pitch for tuning, musicians will deviate from the standard when the situation demands it. Likewise, what we consider right and wrong seems tightly connected to the kind of creature we are: what things are likely to hurt us, what pleases us, what our basic needs are, and how they may be met. At the same time, the particular standards of propriety that are operative in society or that we negotiate in particular interactions are mostly conventional.

Where few if any seem bothered by the role conventional standards play in music, the idea that morality is somehow conventional is often considered highly problematic. Perhaps a better understanding of how the conventional aspects of pitch relate to the natural ones—as well as to the enjoyment or even value of music—could help us navigate the perceived problems of conventionality in ethics? We could formulate similar questions from the other musical concepts that Smith uses to describe the interactions underlying morality, such as harmony, concord, and beat. Seeing Smith's musical metaphors as an analogical model for morality, however, also raises a more fundamental question about the model itself: How do we understand music, the source domain of the musical metaphors?

I.ii.3. Music As Aesthetic Object Versus Music As Practice

In *Adam Smith and the Virtues of Enlightenment*, Charles Griswold analyses Smith's work in the light of its key metaphors. According to Griswold, Smith's juxtaposition of art and life through metaphors such as *theatrum mundi* may lead us to wonder 'whether our lives are in some peculiar sense like works of art, so that evaluating them, like evaluating a play, blurs the line between aesthetic and moral categories' (1998, p. 67). Smith's concept of 'sympathy,' like the allied desire for mutual sympathy with others, 'responds to the disinterested pleasure that arises from the apprehension of concord. . . . The pull of sympathy in our lives testifies, in short, to our love of beauty' (1998, pp. 111–12). The pleasure we get from seeing someone who is in concord with his fellow human beings springs from the same source as the pleasure we get from seeing works of art. Griswold regards Smith's 'striking fondness for musical metaphors' as an expression of this general tendency to *aestheticize morality* (1998, p. 183). These metaphors 'pervade [*The Theory of Moral Sentiments*] and express Smith's conviction that life is suffused with a spontaneous love of beauty' (1998, p. 300).

Griswold's interpretation is both plausible and informative. It rests, however, on the implicit assumption that the relevant aspect of music, in this case, is the one we study in aesthetics: the work of art as an intentionally produced artefact. For music to aestheticize morality, the music itself must first be defined as an aesthetic artefact. There is, to be clear, nothing illicit about this. Nevertheless, in the particular context of Smith's use of these metaphors, framing music in terms of aesthetics presents some problems for our understanding of morality.

To see this, consider a distinction drawn by Knud Haakonsen between practical and theoretical imagination in Smith's *Moral Sentiments* ((Haakonssen, 2002, p. xiii). Although the two are expressions of the same 'desire for order,' they are also 'fundamentally different' (2002, p. xiii). Practical imagination—'sympathy' is Smith's word—is responsible for 'creating the moral world,' writes Haakonssen, by allowing us to ascribe actions to persons, including ourselves, and to evaluate these actions as appropriate or not, based on our ability or inability to sympathize with them (2002, p. xiii). The theoretical imagination, on the other hand, is concerned with bringing 'order and system into things and events' and is thus 'the foundation for all the arts and sciences' (2002, p. xiii). If we apply this distinction to Smith's use of musical metaphor while at the same time defining music as an object of aesthetics, we seem forced to conclude that when Smith characterizes moral sentiments as harmonious, in concord, discordant, in tune, and so on, he is bringing the theoretical imagination to bear, urging us to view the personal interaction from without: an *event* to be contemplated and judged, much as we would a work of art.

It is trivially true that any single interpretation makes sense of a phenomenon in part by excluding alternate interpretations. In the case of Smith's musical metaphors, however, understanding them in aesthetic terms excludes what we take to be a particularly important alternative interpretation: namely, that they can help us explore the intricacies of the *practical* imagination. To learn something from these metaphors about the practical imagination, about what it means to be *in* a moral world, we have to construe their source domain not as a realm of artefacts and events but as a realm of personal interaction between agents that act in concord. We must, in short, consider music as *practice*.[39]

I.ii.4. Music As Practice

In the Western philosophical tradition, the term 'music' has often been taken to refer to a defined and bounded collection of acoustic materials conceptualized and reified into the form of a musical work. Music is seen as an aesthetic object, a thing to be contemplated. Traditionally, the largely unquestioned thingness of a musical work has formed both the basis of music's self-contained autonomy and its ability to create meaning (Small, 1998, p. 4). Musicology has a long-standing preoccupation with music as an object, something that can be measured, described, analysed outside and beyond the people who make it and experience it. Yet such understanding of music fails to account adequately for the enormous variety of sounds, structures, practices, and experiences commonly included within the term 'music,' globally and historically. The traditional understanding of music has been challenged in several ways, and we shall restrict our discussion to briefly mentioning two important redefinitions of music as practice.

John Blacking led the challenge to the standard definition of music detailed above (Blacking, 1973). A pioneer in the field of ethnomusicology, he observed that music was neither an elite skill nor a 'sonic object,' suggesting instead that music was better defined as a kind of social action that had consequences for other kinds of social action (Blacking, 1995, p. 223). Blacking's work points to the fundamental connection between musicality, musical thinking, and the dynamics and organisation of human social life, what he termed the 'musicosocial' (1995, p. 231). Music is here foregrounded as a vital capacity rather than ancillary or abstract, 'a basic human mode of thought by which any human action may be constituted' (1995, p. 224). Blacking's extensive work with the Venda people of South Africa led him to believe that music should be considered 'a primary modelling system' for human thinking, 'generative' as a cultural system and as a human

---

39 Smith himself seems to have had little interest in technical or strictly aesthetic aspects of music. When, as in (Smith 1982), he writes about music directly, it is the emotional and social effects of music, in other words, music as a social practice, that is the focus of his interest.

capability (1995, p. 223). In this understanding, music is a way of being in the world and, importantly, a way of being with others in the world. Our innate musicality forms the roots of our sociability and the dynamic structure of our relational capacity.

Evolving Blacking's work, Christopher Small begins his study of the meaning of musical performing and listening with the statement, 'there is no such thing as music' (1998, p. 2). Small's rejection of 'music' as a term is based on the fact that in the English language 'music' is a noun commonly used to refer to 'the thing music,' which, for Small, is inadequate because it is a 'figment, an abstraction of action' (p. 2). Instead, he proposes that we can only attain a greater understanding of music if the noun is recast as a verb, 'to music' or simply 'musicking,' Employing music as a verb removes it from the autonomous and abstract, locating it emphatically as an active process, contextual and relational, located in people and practices rather than existing beyond and outside them. The term 'musicking' directs our attention to the way music resides 'in actions, in what people do,' and in what they do together (p. 8).

Small's work critiques the abstract 'music as object' position as growing from an ethnocentric conventionalisation of music bound to the European notated musical tradition of the period stretching from the sixteenth to the mid-twentieth century. Challenges to this definition of music and the value system it generates have often been developed through the validation of music from outside this canon, as in Blacking's work. Small, on the other hand, develops his concept of musicking through the example of 'the total experience of a symphony concert' (1998, p. 184), a genre usually held as the very epitome of the European notated musical tradition. He stresses that even in this context music is neither an object nor a rare skill residing only in the highly trained performers on stage but a distributed human capability, a form of action or behaviour, related to the uniquely human capacities for communication and relationship building. Accordingly, taking part in a musical event in any capacity is an instrument of relational 'exploration' (p. 183). 'By bringing into existence relationships that are thought of as desirable, a musical performance not only reflects those relationships but also shapes them. It teaches and inculcates the concept of those ideal relationships,' with relationships created not only among the sounds as they are created and performed but also 'among the people who are taking part' (p. 184).[40]

40   Most recently Small and Blacking's claims for music have found increasing resonance in the umbrella concept of *communicative musicality*, a theory that positions human musicality as the pre-linguistic basis for human thought and action rather than just the basis for all forms of musicking (Malloch & Trevarthen, 2009). Malloch and Trevarthen refer specifically to the work of Adam Smith as consistent with their own insights into the fundamental nature of human musicality: 'It is our common musicality that makes it possible for us to share time meaningfully together, in its emotional richness and its structural holding, and for us to participate with anticipation and recollection of pleasure in the 'imitative arts' as explained by Adam Smith' (2009, p. 5).

In what follows, we argue that Small's analysis of the late twentieth-century symphony concert—with all its norms, ideals, and codes of conduct—provides insight into the source and nature of what we call 'musical autonomy.' This insight can, in turn, give us a new perspective on the place of moral autonomy in Smith's theory of moral judgement—a problem that has engendered a significant amount of debate in the secondary literature on *Moral Sentiments*.

I.iii. Music, Morality, And The Freedom Of Interpretation

Briefly, the issue of moral autonomy in connection with *Moral Sentiments* is the question of whether Smith's theory of the 'impartial spectator' (Smith 2002, passim; e.g., III.1.2, p. 129) admits moral agents sufficient autonomy to criticize not only the application of moral norms in particular situations but also the validity of the norms themselves. In order to understand why this question arises, it is helpful to have a basic grasp of Smith's account of moral autonomy.

I.iii.1. The Moral Blindness Of The Impartial Spectator

Smith starts out with some simple observations, one of them being that we tend to be anxious that others should like us. Because of this, we try to imagine what others would think of us if we acted in such and such a manner. By analogy to how we use mirrors to check our own appearance, we set up a metaphorical mirror to our own behaviour in the form of an imaginary spectator (2002, III.1.5, p. 131). Taking the perspective of this spectator, we can predict what others will think of us based on our experience of what their judgements are usually like. Through the habit of viewing ourselves from the perspective of an imagined spectator, however, our predictive judgements gain a certain independence from the actual judgements of real spectators. Based on our experience of being unfairly treated by real spectators who are often either ill informed, partial, or both, we gradually form an idea of what 'ought to be the judgement of others' (III.1.2, p. 128), namely, what an 'impartial and well-informed spectator' would judge, if such a one were present (III.2.32, p. 150).

Even though the perspective of the impartial spectator gives us a degree of autonomy from actual spectators, it seems that the standard of propriety we apply will still only be an idealized version of the standard of propriety followed by actual spectators. As Samuel Fleischacker puts it, 'The impartial spectator is disinterested, well-informed and "candid" . . . but is otherwise just like actual, partial spectators. It is built out of . . . the basic modes of moral judgment that our actual friends and neighbors have' (Fleischacker, 2011, p. 28). If the impartial spectator is no more than an idealized version of 'our friends and neighbors,' it likely also conserves or even distils whatever biases and prejudices might be endemic to the moral culture in question. If that is the case, taking

the perspective of the imagined impartial spectator will allow us to see ourselves from without, but it will not allow us to step outside the standard of propriety of our society. If there is anything wrong with the standard of propriety in the society to which we belong, we shall not be able to discover it. Our autonomy as individual moral agents would then be limited to our specific society, and we end up with a kind of cultural relativism that, among other things, seems difficult to combine with any vision of moral progress.

Smith did not think that this was all there was to the impartial spectator. This is most evident in his admonition of Plato and Aristotle for their support of the practice of infant exposure, or the indirect killing of unwanted newborns by abandoning them out of doors. Smith accepts that the practice might have been justified by the extreme hunger and constant threat of death of an earlier age 'of the most savage barbarity' (2002, V.2.15, p. 246). By the time, however, of 'the latter ages of Greece . . . the same thing was permitted from views of remote interest or convenience, which could by no means excuse it' (V.2.15, p. 246). The reason even great thinkers like Plato and Aristotle failed to see this was that 'the uniform continuance of the custom had hindered them . . . from perceiving its enormity' (V.2.15, p. 246). In other words, two of history's greatest thinkers were blinded to the moral wrongness of killing infants by the mere fact that it was an established practice.

In Smith's own day, the transatlantic slave trade was similarly permitted for reasons that could by no means excuse it, and Smith does his part in *Moral Sentiments* to argue against it on this ground (V.2.9, pp. 240–42). With the benefit of hindsight like Smith's vis-à-vis the Greeks, we in the early twenty-first century can see that Smith himself had his own moral blind spots—for example, on the question of the equality of the sexes. No doubt we ourselves are similarly blind to or only dimly aware of aspects of our conduct which are equally unjust and unreasonable, and for which our descendants will harshly condemn us. In all these cases, Smith's conclusion is clear: being blinded by tradition is no excuse for failing to correct the moral perversions of our particular practices. We ought to recognize this perversion even if it is difficult to do so when we are accustomed to them. Because of this, we need to answer the question of how we can come to realize that a particular practice is perverted. We need, that is, an understanding of the source and nature of moral autonomy.

This is where the analogy with musical practice comes into play. The reasoning is as follows: If we take *musical autonomy* to be the ability to critically reflect on musical practice in such a way as to render musicians capable of changing this practice through practising music, we can look for the source of this autonomy in an analysis of musical practice. Having located this source, we have discovered, by virtue of the analogical model, a candidate source of moral

autonomy. What we need, therefore, is an analysis of a musical practice in which we can locate the desired kind of musical autonomy. We find this in Small.

I.iii.2. Musical Autonomy In The Symphony Concert

On the face of it, Small's analysis of the peculiar ritual that is the late twentieth-century symphony concert presents us with a portrait of musical practice strictly bounded by convention. Every aspect of the performance is tightly choreographed and regulated by explicit and implicit norms.

Small describes the grandiose concert hall, 'designed down to the last detail to house not just musical performances but performances of a very specific kind' (1998, p. 20); the audience, which, physically separated from the musicians but forcefully directed towards them by the orientation of their seats, 'knows it is to keep still and quiet' (p. 26); the uniformed musicians, whose evening wear locates them 'in a social between-stairs, on the one hand proclaiming their social equality with the members of the audience and on the other suggesting their continuing status as providers of services for the upper classes' (p. 66), each in possession only of a small part of the complete score (p. 110); and, finally, the conductor (usually a he) on his dais, 'the centre of attention . . . of this whole vast space' (p. 25), who presides over the ritual and directs the individual efforts of the musicians into a coherent whole, but whose apparently absolute authority only extends as far as the score—that enduring testament to the sonic intentions of the composer—permits: the conductor 'can make no gesture that is not inspired by those instructions, make no demands on the players that is not sanctioned by them' (p. 115). The score itself is something of a sacred text. Certain musicologists, in the fashion of religious scholars, seek out the most 'authentic' versions they can find (p. 90), thus subscribing to the idea that the meaning of art resides in the art object. In the case of music, however, that idea has some unfortunate corollaries.

One of them is that the performance of the work, the actual playing, is secondary, even incidental, to the work 'itself.' The performance only exists as 'the medium through which the isolated, self-contained work has to pass in order to reach its goal, the listener' (Small 1998, p. 5), and so 'the quality of the work sets an upper limit to the possible quality of the performance' (p. 6). Any given performance can only be as good as the score that is performed.

Moreover, given the technical limitations and brute contingencies of a physical performance, a performance will only ever imperfectly approach the perfection that the score may embody. Thus, if one does not side with Johannes Brahms in preferring to stay at home reading a great work of music, one might agree with Igor Stravinsky, according to whom the 'execution' of a musical work should be nothing but 'the strict putting into effect of an explicit will that contains nothing beyond what it specifically commands' (Stravinsky, 1947, qtd. in Small, 1998, p. 6).

Therefore, the performers, the living musicians, 'can clarify or obscure a work, present it adequately or not, but . . . have nothing to contribute to it; its meaning has been completely determined before a performer ever lays eyes on the score' (Small 1998, p. 5) If there is any musical autonomy in the symphony concert, it is certainly well hidden.

I.iii.3. The Moral Score Of Society

Interpreting Smith's use of musical metaphor in the light of the symphony concert would lead us to think that moral action is a matter of finding and following the behavioural script laid out by the 'score' of society's standard of propriety. The more exactly this can be followed, the more perfect the propriety of the behaviour.

One need not think that a divine composer has written this score; the idea of a 'moral score' is equally amenable to thinking in terms of a collective endeavour, the result of which is the moral norms of a given society—a behavioural script for proper action and sentiments in that society. In either case, moral education and individual moral development would amount to a rigorous training in the execution of extant moral norms—be these what they may.

Of course, as Smith notes, no moral society can subsist on thoroughly perverted moral norms (2002, V.2.16, p. 247). Human societies have basic needs that a system of morality must fulfil, and if 'custom' and 'fashion' are allowed to pervert the usages of otherwise suitable moral norms to the point where the norms themselves become perverted, that society is already far along on its way to self-destruction.

Even with these checks in place, however, moral development, moral education, and moral action would all be measured by the degree to which they approach a pre-set ideal of perfect propriety. Moral autonomy would be wholly restricted by the moral conventions into which one is socialized. Fleischacker's criticism that the impartial spectator is no more than an idealized version of 'our friends and neighbors' would be supported also by Smith's musical metaphors. There would be no answer here to the question of how we are to go about discovering our moral blind spots.

This, however, is not the whole of what Small's analysis shows us. If Western classical music were the only thing we recognized as music and the symphony concert the only kind of musicking, it would indeed be difficult to escape its confines. Luckily, that is not the case. We use the term 'music' for an incredibly diverse set of sonic relationships, and, if Small is right, what unites them is that they 'explore, affirm, and celebrate' sets of human relationships that those taking part in the performance 'feel to be ideal' (1998, p. 49). Therefore, even if we are thoroughly socialized into a specific tradition, we can recognize radically different ways of musicking as the celebration of alternative, and possibly valuable, sets of human relationships.

Small's analysis of the late twentieth-century symphony concert allows us to step outside whatever presuppositions we might have about the nature or quality of classical music in the Western tradition, and critically reflect on the practice and the sets of relationships it celebrates. If we free ourselves from Western classical music as the paradigmatic example of what music is, we can come to see that the strong authoritarian bent of this tradition is a contingent feature of one way of musicking, rather than a general feature of musicking as such.

That is not to say that other forms of musicking are fundamentally free in a way the symphony concert is not, or that they afford a musical autonomy essentially different from the one available to those partaking in an instance of musicking in the European annotated music tradition. Small warns against the kind of 'neat antithesis' (1998, p. 44) one might be tempted to postulate between a bourgeois symphony concert, celebrating the values and relationships of the industrialized society, and a phenomenon like 'the great rock festivals of the 1960s and 1970s' (p. 45). While the latter became famous for creating, temporarily, something like a parallel society founded on tolerance and love, Small continues (p. 45), they did so not by escaping constraints but by establishing new ones: 'At rock festivals, as at any other kind of musical event, there were, and are, right and wrong ways to behave, right and wrong ways to dress, to speak and to respond, both to one another and of course to the musical performances. To dress or behave there in ways that come naturally in Symphony Hall would be to invite ridicule, if not downright hostility. That these codes were felt by those present not as constraints but as liberation only goes to show how lightly norms fall on those for whom they represent ideal social relationships' (p. 46). The point, therefore, is not that there is a kind of musicking that, if used as a model for morality, would reveal the source of moral autonomy. Rather, any kind of musical performance, however constrained it might appear, provides affordances of musical autonomy. To see this, we shall consider the role *interpretation* plays in musicking.

I.iii.4. Imperfection And Interpretation

The starting point for finding the source of musical autonomy lies in the realization that no two musical performances, no two instances of musicking, will ever be the same—even if the two are instances of the same symphonic orchestra playing the same work by the same composer. The reason for this is trivial. Playing a symphony requires a concerted effort of perhaps a hundred musicians, with none of the thousands or millions of bodily motions performed by them ever an exact replica of any other, nor the sounds produced ever the same. Moreover, each instance of musicking is constituted also by the relations between the musicians, between them and the conductor, between all of them and the audience, between all of them and the building in which

they play, and so on; and by second-, third-, and even higher-order relations between these relations —patterns that defy description but can be experienced and explored in the musicking itself (Small 1998, p. 200). While the general form of these relationships may be recognizably similar, perhaps even indistinguishable, depending on the granularity of our analysis and the aims of those engaged in the two instances in musicking—a symphonic concert is, after all, as Small argues, an enactment of stability (1998, p. 90)—their constitution, for the reason noted above, is not.

This unavoidable variability is the flip side of the fact that no instance of musicking, not even the professional performance of a work of Western classical music, is ever the mere following of a score but always an interpretation of it. How the score is interpreted—that is, how it is performed—might, in the case of a symphonic orchestra, in large measure be up to the conductor. Still, however small we make the space between direction and execution, there will always be a gap, a need for translation of one thing into another: the notation into musicking. Where there is translation, there is always, no matter how accurate it aspires to be, space for interpretation. Where there is interpretation, there is always also freedom.

Granted, the freedom of interpretation is bounded by the possibilities offered or realized in the domain from which you translate and the domain into which you translate, and so the score, the musical genre, the direction of the conductor, and the skill of the individual musician all put bounds on the freedom of interpretation.

But there is a deeper, more general point in this: performance is always interpretation. However constrained, a performance of a score will only ever be one of endless possible variations. The variations may sound alike, at least to the untrained ear, but none of them will ever be the same. Trying to weed out all errors of interpretation or trying to arrive at an authentic performance of a score is, therefore, in a sense, senseless. At 'best,' one will arrive at an interpretation shared by the entire orchestra for the duration of the performance. The deeper point is thus that the style of musicking represented by the symphony concert is itself just one of many that are possible. The desires for note-perfect, authentic, or otherwise perfected ways of performing an orchestral piece is itself an interpretation of what musicking should be.

By realizing that there is such a need for interpretation, we can come to see even the most forcefully protected musical convention as just that—one convention among many possible. Being one of many possible does not mean that the convention is without value or some claim to allegiance. It does mean, however, that it has no absolute value or unquestionable authority. Musicking is about exploring, affirming, and celebrating a set of human relationships: how we relate to each other and to the world. The relationships celebrated in the symphony concert—sonic, social, commercial, and cultural—constitute one very particular set of relations among those

possible. This set has some things to recommend it, and others that count against it. Realizing your interpretational freedom is, therefore, a way to realize the interpretational nature of the practice itself. When combined with the realization that fundamentally the same is true for everyone, this leads to the conclusion that you are not relegated to merely following the rules set by others. On the contrary, you are in on the making of them.

That said, the freedom thus realized does not put you suddenly outside all convention, free to create, from nothing, a new set of ideal relationships. Small's image of a 'herdsman playing on his flute . . . in the African night' (1998, p. 201) both reinforces and nuances this point. The solitary flute player stands in sharp contrast to the collective conventionality of the orchestra musician, but his freedom, though real, is not as radically unbounded as it might first appear. His simple flute 'is as much a product of technology and of technological attitudes and choices as is the Western orchestral instrument that goes by the same name, and it is as finely adapted as the Western instrument to the musical and social purposes for which it is intended' (p. 202). As for the music he plays, it will almost certainly sound strange to Western ears adapted to Western notions of beats and harmony. Failing to find familiar rhythms, the sounds may appear to such ears as free in the sense of unpredictable or chaotic. Nevertheless, 'we may be sure that they are always being measured against a rhythm that is going on in his head,' and 'whatever it is he is playing, it will not be invented from nothing. No human being ever invents anything from nothing but is guided always in his invention by the assumptions, the practices and the customs of the society in which he or she lives—in other words, by its *style*. A person may rebel against the assumptions of the society, but the style of the rebellion will inevitably continue to reflect those assumptions. It is inescapable' (p .203, emphasis in original). Style is inescapable. Nevertheless, the recognition of other sets of relationships as explored in other kinds of musicking provides us with stylistic tools other than those most readily available in the society in which we live. These tools can be used to embellish or criticize the style we are familiar with, whether as individuals or as members of a collective. We cannot escape style in musicking, but we are not confined to just one. Nor does any style ever stop evolving. Not even Western classical music, 'with the repertory held steady and with the authenticity movement thriving' (Small 1998, p. 90), has achieved stasis. 'Of course such a thing is impossible. Each generation of musicians and listeners remakes the culture in ways that will support and sustain their values' (pp. 90–91). The source of musical autonomy thus lies in the need for interpretation, both in the following and in the shaping of different styles of musicking.

I.iii.5. Moral Autonomy And The Ideal Of Perfect Propriety

Is moral autonomy similarly founded on the freedom of interpretation? Before we can consider this question, we must step back and consider a more general one: namely, whether conceiving of the source domain of Smith's musical metaphors as music-as-practice really is compatible with the target domain as described by Smith.

We think it is. The principal role played by sympathy in Smith's theory suggests that what we usually reify as 'morality' is originally and primarily a practice: the complex ways in which individuals 'feel in' to each other in a 'self-regulating process of sympathetic exchange' (Forman-Barzilai 2010, p. 193). Morality understood as a practice is a sort of *moral tuning*.

Evidence in favour of this interpretation can be found throughout *Moral Sentiments*, but is most clearly expressed in what Smith says about the 'general rules of morality': far from being the foundation of our moral judgements, they are 'founded on experience of what, in particular instances, our moral faculties, our natural sense of merit and propriety, approve, or disapprove of' (2002, III.4.7–8, pp. 184–85). Take murder as an example. The first person who saw 'an inhuman murder' needed no divine command to grasp its wrongness; on the contrary, the general rule against killing arose from the 'detestation' that this person 'felt necessarily arise . . . at the thought of this, and every other particular action of the same kind' (III.4.8, p. 185). Such general rules may in time become 'universally acknowledged and established,' and they are thus frequently cited as the foundation of our moral judgements (III.4.11, p. 186). The general rules, however, are really just shorthand summaries of human experience, rules of thumb that we can use to guide ourselves when we are too pressed or hot-headed to truly take the perspective of an impartial spectator and properly survey the situation (III.4.12, pp. 186–87). Sympathy and the moral judgements issuing from it are thus primary to the general rules of morality. Moral tuning is primary to the moral score.

This primacy of practice is true also in the case of the impartial spectator. Even though Smith often personifies it as 'the great demigod within the breast' (2002, IV.iii.25, p. 291), the idea of the impartial spectator arises out of a particular act of the imagination, namely, the taking of an outside perspective on ourselves (III.1.2–7, pp. 128–32). As we have already argued, moral autonomy, to the extent that we have any, comes from this ability to see ourselves from without, and through the perspective of the hypothetical impartial spectator to second-guess the judgements of the actual spectators surrounding us. Therefore, if interpretation plays any role in moral autonomy, this is where we would expect to find it.

When looking for such interpretation, we find a good starting point in Smith's observation that we always have a kind of double vision when judging our own merit: 'In estimating our own

merit, in judging of our own character and conduct, there are two different standards to which we naturally compare them. The one is the idea of exact propriety and perfection, so far as we are each of us capable of comprehending that idea. The other is that degree of approximation to this idea which is commonly attained in the world, and which the greater part of our friends and companions, of our rivals and competitors, may have actually arrived at' (2002, IV.iii.23, p. 291). We hardly ever, contends Smith, consider our own merit only by comparing ourselves to what we ideally ought to do; we almost always also compare ourselves to our friends and companions. The main effect of this comparison may be the comforting thought that, while no saint, I am at least better than my friends are. But the 'idea of exact propriety and perfection' can also function as something towards which to strive, and with which to criticise our own and other people's shortcomings. This includes shortcomings in moral reasoning. When Smith criticises Plato and Aristotle for failing to condemn the practice of child exposure, he is in effect appealing to his own idea of exact propriety and perfection to criticise that of those venerable philosophers. If we in turn criticise Smith for his failure to grapple adequately with the issue of gender inequality, we apply our own idea of such perfection to Smith.

The trouble is that this idea is itself a product of our moral surroundings: the idea of exact propriety and perfection is, Smith admits, 'gradually formed from [our] observations upon the character and conduct both of [ourselves] and of other people' (2002, VI.iii.25, p. 291). If we have no other standard by which to judge the propriety of our actions, we risk ending up with just the kind of social or cultural relativism that Fleischacker was worried about. Of course, there are other people and cultures with other standards, and we could draw on these in criticising our own. If, however, we have no third, overarching standard with which to adjudicate, it is hard to see how we could even get this process started.

But reconsider for a moment the very thought that there is something you *ideally* ought to do. As Carola Freiin von Villiez points out, this thought is strictly formal, a '*Grenzideal*' or limiting ideal transcending any particular interpretation of it (2006a, p. 206; 2006b, pp. 130–34). The thought that there is something you ideally ought to do contains no reference to what this something is. Accordingly, the idea of exact propriety and perfection could play the role of the third, overarching standard that we can use to adjudicate different conceptions of propriety.

For this normative limiting ideal to be of any use in guiding our judgements and actions, we have to give it a substantive interpretation, filling in the blank 'something' with concrete particulars. In giving such an interpretation, we are inevitably drawing on our own, limited experience (Freiin von Villiez 2006a, p. 203; 2011, p. 41), and this experience will be shaped by the style (in Small's

sense) of the society in which we have matured. Therefore, that style will put bounds on the freedom of our interpretation of what that something is.

Nevertheless, consider what we established about musical autonomy above. Although style itself is inescapable, a musician is never limited to a single style. Nor does any style ever remain constant; it is constantly reinterpreted by those engaged in applying it in practice. Small's solitary flute player is bounded by convention but free to interpret and reinterpret this in response to different situations and the conventions of others. When interpreting the normative limiting ideal of perfect propriety, moral agents appear to be in an analogous position. Interpretation is necessary, and so a certain freedom of interpretation—the freedom to draw on the standards of propriety of different people and different moral cultures, adapting these to the particular situation at hand—is built into the foundations of morality-as-practice.

There is, however, a problem with this analogy: the 'idea of exact propriety and perfection' has no obvious analogue in musicking. A disanalogy for such a central concept could undermine the comparison between music-as-practice and morality-as-practice. If we have ideas of perfect propriety against which we test imperfect manifestations of it, then critically reflecting on moral norms appears to be essentially different to the process of interpretation in musicking, in which individuals can simply draw on different styles to embellish or criticize the style they are most familiar with.

Notice, however, that if we reverse the analogy, we can see different ideals of musicking as representing imperfect interpretations of a formal limiting ideal of musical perfection. The ideal of performing an orchestral piece the way the Great Composer intended would then be one such substantial interpretation, the rock festivals of the 1960s and 1970s another, and so on for any substantive ideal of musicking.

Here the reader may object that we have merely traded one problem for another. For what is a 'formal limiting ideal of musical perfection' supposed to be? Can we even imagine such a thing? The answer is yes. If we accept Small's broad characterisation of musicking as a way of 'exploring, affirming, and celebrating ways of relating to one another and to the world' (1998, p. 87), then musical perfection is not an unknown after all. It is simply the thought that there is a perfect way of relating to one another and to the world. What this perfection consists in is available to us only in imperfect interpretations—namely, in specific ideals of musicking. Wondrously, it then turns out that specific ideals of musicking and specific ideals of propriety are different kinds of answers to very same question: How are we to live as humans among humans? Considered as practices, that is, music and morality are two sides of the same many-faced die.

We can criticize, revise, and (temporarily) justify particular moral norms from the imagined point of view of what we imperfectly conceive of as an impartial spectator. At the same time, the ideal of perfect propriety that guides us in so doing is itself is open to revision in the light of new information and the different perspectives on propriety that you may meet with in others (Freiin von Villiez 2006b, p. 132). This mutual exchange can potentially take the form of a reflective equilibrium (pp. 130–34) between the formal ideal and various imperfect substantive interpretations of it. Such an equilibrium would provide a standard that is stable enough for judging the propriety of our actions, without ever ossifying, and thus never pretending to be the one and only truth about what is right and wrong.

The mirror in which we see ourselves distorts our vision with the assumptions of our society. Even norms justified through reflective equilibrium will reflect this distortion. Nevertheless, in the constant need for interpretation, there is room for revision, rebellion, and even reconstruction of our set of moral norms. The freedom of interpretation—bounded as it is, but freedom nonetheless—appears built into the very fabric of morality, just as it is in musicking. The question, therefore, is not how the individual musician or moral agent can come to have this freedom but how we can come to realize that we already do. If we do, the realization that interpretation is essential to both musicking and morality also allows us to see that *not* questioning established norms and ideals, not using our moral and musical autonomy, is nothing more than to propagate, borrowing a phrase from Herbert Spencer, 'the rule of the dead over the living' (1899, p. 515). There is certainly nothing inherently noble in that.

I.iv. Conclusion

By treating Smith's musical metaphors in *The Theory of Moral Sentiments* as elements of an analogical model in which the source domain is musical *practice* rather than musical works, we open new possibilities for interpreting Smith's model of moral judgement, as well as new paths to discovering and exploring affinities between music and morality more generally.

That there are affinities between Smith's model of moral judgement and the domain of music becomes particularly clear when we look to recent developments in musicology. From Blacking (1973; 1995) through Small (1998) to the current interest in communicative musicality (Malloch and Trevarthen 2009), the turn from object to process in music studies shifts perception of music-as-source-domain from aesthetics to practice. This in turn warrants a closer look at Small's analysis of Western classical music and the particular ritual that is a symphony concert.

By showing us the set of relationships celebrated in the symphony concert, Small allows us to step outside whatever presuppositions we might have about the nature or quality of Western classical music, thus giving us the distance we need to reflect critically on these relationships. Small achieves this not just by pointing out to 'us'—Westerners steeped in Western traditions—the alternative represented by the ways of the 'other' but also by engaging in the same way with the very tradition in which we are steeped. To cultivate moral autonomy, moral philosophy should similarly encourage such 'fieldwork in familiar places' (Moody-Adams, 2002, p. 224).

In order to do so, we must tread the line between what Fleischacker has called the anthropological and philosophical approaches to morality: recognising the norms of different societies as actual moral norms while also providing the philosophical tools necessary to critique them (2011, p. 25). Fleischacker laments what he sees as Smith's failure to combine these two approaches (p. 40), but seeing morality as a practice in analogy to musicking allows a reconsideration of Smith's purported failure. The analogy established by Smith's musical metaphors then reveals the freedom inherent in the constant need to interpret and reinterpret the strictly formal ideal of perfect propriety.

II. Love Redirected: On Adam Smith's Love Of Praiseworthiness‡

Abstract

Why be moral? Why, in the language of Adam Smith, act on what you think is praiseworthy even when it does not get you praise from other people? Because, answers Smith, you love praiseworthiness. But what is this love of praiseworthiness, and where does it come from? In this article, 1) I argue that we start to love praiseworthiness when we redirect our love of praise away from other people toward the 'impartial spectator'-aspect of ourselves, and 2) show how this fits with evidence that the rudimentary moral compass which guides us early in childhood needs correction through socialisation to develop into a mature moral conscience.

II.i. Introduction

> Sweet princes, what I did, I did in honour,
> Led by the impartial conduct of my soul:
> And never shall you see that I will beg
> A ragged and forestall'd remission.
> (Henry IV Part 2 Act 5 Scene 2)

If Adam Smith had wanted a literary figure to illustrate his *Theory of Moral Sentiments* (TMS, 2002),[41] he could have done much worse than to pick Shakespeare's Lord Chief Justice. In the course of a single scene towards the end of *Henry IV Part 2*, this figure plays out most of the major themes of Smith's moral theory.

In the scene in question, Henry IV has just died, and his son, Prince Hal, is set to succeed him. Although Hal has shown valour in war, he is still best known for his debauched lifestyle and questionable companions. The court is in a state of anxiety over the prospect of his coronation, and Lord Chief Justice is told that he stands 'in coldest expectation' of the new monarch's grace due to his previous run-ins with the prince's oft-unlawful coterie. On one occasion, the play hints, the

---

41   All references to *TMS* will be to the Cambridge edition (2002).

Chief Justice even sent the prince himself to prison. In spite of this, and going against the advice of Prince Hal's 'sweet' younger brothers, Lord Chief Justice refuses to resort to flattery or to beg forgiveness for his actions. Confronted with the anger of the soon-to-be-crowned Henry V, who has not forgotten his time behind bars, Lord Chief Justice instead asks the nascent king to

> Question your royal thoughts, make the case yours;
> Be now the father and propose a son
> (Henry IV Part 2 Act 5 Scene 2)

Read through the lens of Smith's *Theory*, we can see Lord Chief Justice in the twenty-nine-line passage as appealing to Henry V's capacity for 'sympathy' (TMS I.i.1.5, p. 13), asking him to bring 'the case home to [his] own breast' (I.i.3.9, p. 23) by 'changing places in fancy' (I.i.1.3, p. 12) with his father, Henry IV. Only by so doing, the Chief Justice is saying, can Henry V judge truly whether it was right of Lord Chief Justice to imprison him. He must think whether he himself would have had his own Chief Justice imprison such an insubordinate son:

> Hear your own dignity so much profaned,
> See your most dreadful laws so loosely slighted,
> Behold yourself so by a son disdain'd;
> And then imagine me taking your part
> And in your power soft silencing your son
> (Henry IV Part 2 Act 5 Scene 2)

The spiel works. Henry V finds that he would have had his own insubordinate son imprisoned in such case, and asks Lord Chief Justice to 'still carry the balance and the sword' under his reign. In thus standing up to the king, and in his reply to the 'sweet princes', Lord Chief Justice also embodies something of Smith's ideal 'wise man' (TMS III.2.7, p. 136): A person who so thoroughly identifies with the point of view of the 'impartial spectator' (III.2.9, p. 137) that he cares little for *actual* praise and blame, being instead guided by the desire to be *worthy* of praise—the 'love of praise-worthiness' (III.2.25, p. 147).

 In comparison to the other elements of Smith's theory, the notion of a love of praiseworthiness has received little attention in the secondary literature. This is a significant lacuna. First, without a love of praiseworthiness, the ability to distinguish what is truly praiseworthy from what is merely praised is, morally speaking, worthless. One would know what is right but have no desire to pursue it. Second, Smith's most substantial discussion of the love of praiseworthiness, added to the sixth and final edition of *TMS*, is mostly concerned with distinguishing this virtuous love from the vain desire for praise; we are not clearly told how we come to have this love, nor how

we can foster it in ourselves and in others. Combined, the result is potentially disastrous: Unless we are able to show how the love of praiseworthiness fits within Smith's wider theory, we are in effect allowing that its plausibility as a normative theory rests on the postulation of an inborn moral compass.

I say 'potentially' because, while psychologists up until the late twentieth century were fairly certain that we had to be taught the difference between right and wrong, recent research on the moral psychology of infants actually lend some credence to the notion of an inborn moral compass. Infants as young as six months apparently consider helping others achieve their goals to be good and hindering them to be bad. However, research on infant and child morality also reveals that our early moral compass is rather wonky, its guidance distorted by egotism and in-group partiality. It needs correction through moral education if it is to aspire to anything like a mature morality, not to mention the love of praiseworthiness displayed by Shakespeare's Lord Chief Justice. If, then, we are not born with a love of praiseworthiness, how does it develop?

In what follows, I will build on insight gleaned from moral psychology, scholarship on Smith, and what Smith himself writes in earlier editions of *TMS* to argue that the development of a full-blown love of praiseworthiness requires the redirection of our naturally strong desire for praise from others toward the imagined impartial spectator, our conscience. The love of praiseworthiness is not an inborn or original love, but a love redirected. Realising this not only helps us make sense of Smith's theory but aligns it with the current understanding of morality as part natural endowment, part human education. By showing how the love of praiseworthiness relates to the desire for mutual sympathy, my argument also provides a testable hypothesis about moral agency: Developing the ability to experience self-sympathy will correlate positively with individual moral autonomy.

II.ii. Love Of Praiseworthiness

> How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it except the pleasure of seeing it. (TMS I.i.1.1, p. 11)

> [Man] desires, not only praise, but praise-worthiness; or to be that thing which, though it should be praised by nobody, is, however, the natural and proper object of praise. (III.2.1, p. 132)

Both these sentences appear purely descriptive in form. They are statements of fact. In both cases, the appearance of facticity hides normative implications: We might be selfish, but we are not entirely selfish, so when we care for the fate of others, we at least sometimes do so from a genuine

concern for them. Likewise, we might be vain, but we are not entirely vain, so we sometimes act with genuine concern for what is right. To the extent that the two statements are similar, it is the second that echoes the first. The first statement is what meets us when 'the curtain goes up' (Griswold, 1998, p. 44) on Smith's *Theory*. It's been there since the first edition. The second appears as part of the first paragraph in a chapter that is almost entirely new to the sixth and last edition of *TMS*. In both cases, Smith starts out with a striking claim about moral psychology and proceeds to build upon it a cornerstone of his theory.

In the first case, Smith uses our capacity for sympathy, or 'fellow-feeling with any passion whatever' (TMS I.i.1.5, p. 13) to explain our interest in others. Sympathy, in turn, forms the basis for moral judgements since our inability to sympathise with someone equals disapproval of their sentiments as either excessive or deficient (I.i.3.1, pp. 20-21). When Smith opens *TMS* with the claim that we are not wholly selfish, he is confronting the 'selfish' systems of morality (VII.ii.4.6-8, pp. 353-55) associated with Thomas Hobbes' bleak 'state of nature' (Hobbes, 2009). Smith's main target is Bernhard Mandeville's interpretation of this in his *Fable of the Bees* (Mandeville, 2011; Frazer, 2010, pp. 19–22; TMS VII.ii.7, p. 364). Whereas these authors see humans as thoroughly selfish, and any morality as an artificial arrangement of society, Smith is convinced that humans are fundamentally concerned with the well-being of others. Smith knows his claim is controversial, and he backs it up with detailed and convincing analysis of the role that sympathy plays in our ability to put ourselves in other peoples' shoes and to evaluate their actions and reactions on the basis of whether we can or cannot go along with them.

In the second case, Smith builds on his theory of sympathy and the imagined impartial spectator this spawns (TMS III, p. 128-229), and introduces the notion of a 'love of praise-worthiness' (III.2.25, p. 147) to explain how we are able to act in defiance of the opinions of others when our judgements about what ought to be praised differ from theirs. On the one hand, Smith's second claim—that we not only desire praise but also desire to be worthy of praise even if none is given—can be seen as a corollary of his first: If vanity is a form of selfishness, then the claim that we are not wholly vain is a corollary of the claim that we are not wholly selfish. On the other hand, the claim that we love being praiseworthy carries an explanatory burden not carried by the claim that we are interested in the well-being of others. Smith needs this second notion in part because he thinks our first inclination is to seek harmony or concord with those around us. The desire to sympathise, to agree with others, and to have others sympathise with us is what drives the process underlying morality. Our desire for 'mutual sympathy' (I.i.2, p. 17-20) naturally inclines us toward seeking common ground with other people. This explains how the capacity for sympathy can lead to

the construction of shared moral norms, and it is, therefore, central to Smith's analysis of how morality arises from human sentiments.

Morality, however, is not only about agreeing with others, or about following shared norms. A central aspect of moral discourse as we know it is moral disagreement. We are not just trying to get along, but trying to do so in the right way, guided by judgements, arguments and rules that are somehow truer or better than others. This aspect must be accounted for even if the final analysis of what 'truer' means refers exclusively to what people happen to feel; we need to understand what it is that allows people to disagree just as much as we need to understand the mechanisms of agreement, regardless of what the agreement or disagreement is about.[42] While the desire for mutual sympathy carries within it both these tendencies – it is, after all, the compound desire of understanding others and having others understand us – it cannot, on its own, explain how people sometimes completely forego the sympathy of others to pursue what they take to be right. Going against the opinion of people who are close enough to us that we notice their disapproval requires some other motive than just the desire for mutual sympathy with them; it requires an inner strength of some kind that sustains us in our resolve to act on our convictions even when these turn out to be unpopular. When Shakespeare's Lord Chief Justice stands up to Henry V even though he fears that it may cost him his life, he cannot merely be driven by his belief that he has acted in a manner that is praiseworthy. Coupled only with the desire for mutual sympathy, his belief would fold under the weight of being in discord with the king. What sustains Lord Chief Justice's resolve faced with the prospect of condemnation from the kingdom's highest authority is his love of praiseworthiness.

The love of praiseworthiness is in some ways Smith's answer to the age-old question 'why be moral?'—at least if this is understood as a question about why we should care about what is right, or, alternately, as a request for an explanation for why we, in fact, do care. The love of praiseworthiness is our 'natural incentive to be virtuous' (Griswold, 1998, p. 130). As such, Smith's claim that we are not wholly vain is controversial in its own right. The duty to account for this is not discharged simply by explaining how it is that we are not wholly selfish. If we are to believe that the second claim is a corollary of the first, we need to know how we get from the one to the other. Smith, unfortunately, does not tell us. In the sixth edition of *TMS*, where Smith writes most about the love of praiseworthiness, he tells us almost nothing about how this love connects to our ability to sympathise, our desire for mutual sympathy, and the imagined impartial spectator.

---

42 Even if, as an emotivist or error theorist would claim, the disagreement is not about anything, we would still like to know, for example what people are thinking when they are disagreeing. The psychology is interesting independently of the metaphysics.

Connecting what Smith says about the love of praiseworthiness to the other main elements of his theory is, first of all, an exegetical problem. As an exegetical problem, it is compounded by the relative lack of scholarly work on the love of praiseworthiness. While the notions of sympathy and the impartial spectator have been extensively explored in the secondary literature, only a handful of scholars have written explicitly about the nature of this love. One of them is Ryan Patrick Hanley, and although his claim that the love of praiseworthiness is 'logically and temporally prior' to the love of praise is almost certainly false (Hanley, 2009, p. 140), the possibility that the love of praiseworthiness might an 'original desire' on a par with the love of praiseworthiness deserves further scrutiny (TMS III.2.7, p. 135).

II.iii. An Original Love?

In his virtue-ethical reading of Smith, *Adam Smith and the Character of Virtue* (2009), Hanley connects Smith's discussion of the distinction between praise and praiseworthiness, and the love of each, to a problem first formulated by Rousseau. Society, thought Rousseau, tends to foster a certain kind of self-love, *amour-propre*, that privileges 'a concern for appearances over a concern for the truth of one's character' (Hanley, 2009, p. 41). The consequence of this is that seeming becomes more important than being: 'in Rousseau's terms, *paraître* supplants *être*'(Hanley, 2009, p. 41). Smith, writes Hanley, saw the same problem arising from certain features of commercial society, features which tend to tie advancement to recognition whether or not that recognition is merited (Hanley, 2009, p. 41).

Smith even makes a distinction quite similar to Rousseau's, namely between *appearing* to be fit for society and actually *being* fit. 'Nature', Smith says, has 'formed man for society', and in so doing has 'endowed him with an original desire to please and an original aversion to offend his brethren' (TMS III.2.7, p. 135). However, 'this desire of the approbation, and this aversion to the disapprobation of his brethren, would not alone have rendered him fit for that society for which he was made'; it 'could only have made him wish to appear to be fit for society' (III.2.8, p. 136). The original desire to please, which Smith calls the 'love of praise' (III.2.25, p. 147), will drive people to act in ways that make others see them in a positive light, but only to the extent that there are actual spectators to see and praise them. The act would be just that, an act, a put-on, 'the affectation of virtue, and … concealment of vice' (III.2.8, p. 136). The solution to the problem of the separation of *être* from *paraître*, or the explanation for why people are not merely pretending, but genuinely concerned with being virtuous, is the desire to be praiseworthy: 'Nature … has endowed [man], not

only with a desire of being approved of, but with a desire of being what ought to be approved of; or of being what he himself approves of in other men' (III.2.8, p. 136).

In discussing this part of Smith's argument, Hanley writes that Smith's solution to the Rousseauan problem of the separation of *paraître* from *être* 'rests on his postulation of a love of praiseworthiness *both logically and temporally prior to a love of praise*' (Hanley, 2009, p. 140, my emphasis). To say that our love of praiseworthiness is 'logically and temporally prior' to our love of praise is to say that love of praise depends on the prior existence of a love of praiseworthiness. In other words, if there does not already exist a love of praiseworthiness, neither can there be a love of praise. Does this fit with Smith's theory? It is difficult to see how it could. Indeed, Smith's account of the structure and development of moral agency suggest the opposite ordering. It is our desire to be approved of by others that drives us to view ourselves from without, which in turn lets us see the difference between something being praised and something being worthy of praise. Only after having realised this distinction can we be said to be able to desire to be worthy of praise, and so be able to love praiseworthiness. In this sense at least, it seems that the love of praise is logically and temporally prior to the love of praiseworthiness.[43] However, there is a slightly weaker and more plausible interpretation to be made of Hanley's statement. At the end of the paragraph in which he invokes the logical and temporal priority of love of praiseworthiness, Hanley says that Smith 'makes the Rousseauan claim that nature in its wisdom antecedently invested man with a love of praiseworthiness capable of withstanding and mitigating civilization's corruptions' (Hanley, 2009, p. 140). This, I take it, is a claim that the love of praiseworthiness, like the love of praise, is an 'original desire' (TMS III.2.7, p. 135), a basic part of human nature not derived from something else, and certainly not derived from the love of praise.

Much of what Smith says about the love of praiseworthiness in the sixth edition of *TMS* can be taken to support this interpretation. For example, after confidently stating that '[man] desires, not only praise, but praise-worthiness' (III.2.1, p. 132), Smith goes on, in the very next paragraph, to say that '[t]he love of praise-worthiness is by no means derived altogether from the love of praise'

---

43   At the Smith and Virtue workshop in Oslo, august 2015, Hanley responded to a version of this criticism by saying that, had he written the paragraph again, he would have said that love of praiseworthiness is lexically prior to the love of praise. Lexical priority is a common way of ranking principles in ethics, especially in the tradition after John Rawls. If we interpret Hanley's claim in this light, we could say that love of praiseworthiness is morally more important than love of praise. If in a particular situation there is a conflict between acting so as to be worthy of praise and so as to be praised, one ought to act from a love of praiseworthiness rather than a love of praise. This, I think, is a wholly unproblematic claim about Smith's model of moral judgement. However, it fails to answer the question of why or how it is the case that humans not only love praise, but also love praiseworthiness, and, even more in need of an explanation: how we come to hold praiseworthiness to be more important than praise.

(III.2.2, p. 132). As an example of how the love of praiseworthiness cannot be derived altogether from the love of praise, the third paragraph, also new to the sixth edition, discusses how we judge of our own praiseworthiness by becoming 'the impartial spectators of our own character and conduct' (III.2.3, p. 133). If what we see from this point of view is 'as we wish', that is, if we can consider ourselves praiseworthy, 'we are happy and contented' (III.2.3, p. 133). This contentment does not depend on any actual praise from real spectators, but if we were to find that other people see our conduct and character 'in the same light' as we do, '[t]heir praise necessarily strengthens our own sense of our own praise-worthiness' (III.2.3, p. 133). Smith concludes, '[i]n this case, so far is the love of praise-worthiness from being derived altogether from that of praise; that the love of praise seems, at least in a great measure, to be derived from that of praise-worthiness' (III.2.3, p. 133). Finally, in the paragraph about appearing versus being fit for society quoted above, Smith says that 'Nature, accordingly, has endowed him, *not only* with a desire of being approved of, but with a desire of being what ought to be approved of' (III.2.8, p. 136, my emphasis).

The 'original love' interpretation seems to stand on solid ground, exegetically speaking— except, of course, for the fact that Smith never refers to the love of praiseworthiness as 'original', which is an adjective he reserves for our 'original desire to please [our] brethren' (III.2.6, p. 135). Could this be because the love of praiseworthiness has to be developed from more basic desires? In an endnote to his discussion of the love of praiseworthiness, Hanley points the reader in the direction of Lauren Brubaker's essay *Does the 'wisdom of nature' need help?* (2006) for a fuller treatment of the relationship between nature and the love of praiseworthiness in *TMS* (Hanley, 2009, p. 174, endnote 12). Interestingly, Brubaker's take on the place of the love of praiseworthiness in Smith's theory is almost exactly opposite to that suggested by Hanley. Brubaker says of the impartial spectator and the love of praiseworthiness that they are 'developed or perfected', sentiments that require 'judgement and reflection' to mature (Brubaker, 2006, p. 181). Once perfected, they 'can even come into conflict with the natural sentiments from which they first arose' (Brubaker, 2006, p. 181). Case in point: a properly developed love of praiseworthiness 'can sustain us even under conditions in which we lose the praise of actual spectators, our original or natural desire' (Brubaker, 2006, p. 181).

Brubaker's insight springs from an analysis of Smith's seemingly contradictory uses of the term 'Nature'. Nature, to Smith, is sometimes 'Darwinian' (Brubaker, 2006, p. 177), with its 'favourite ends' being the 'self-preservation and propagation of the species' (TMS II.i.5.10, p. 90). At other times, Smith imputes to nature a concern for human happiness and perfection (see for example III.5.7, p. 193)—a concern that sits uneasily with the stark logic of evolutionary adaptation (Brubaker ,2006, p. 177). Brubaker's claim is that this is not a contradiction, but rather the sign of a

'conflict within nature', a conflict between 'nature simply' and (mature) human nature (Brubaker, 2006, p. 178, emphasis in original; see also Griswold, 1998, p. 313–30). In Smith's own words,

> [M]an is by Nature directed to correct, in some measure, that distribution of things which she herself would otherwise have made. The rules which for this purpose she prompts him to follow, are different from those which she herself observes. (TMS III.5.9, p. 195)

Smith at one point uses the example of an 'industrious knave' and an 'indolent good man' and notes that while nature rewards industry and punishes indolence, it is human nature to prefer the virtuous person to the knave, rendering unjust a natural order in which the knave comes out ahead (III.5.9, p. 195, see also II.ii.2.1, p. 97). In this way and others, the rules of nature are liable to 'shock and offend the 'natural sentiments' of man' (Brubaker, 2006, p. 178). But nature simply and human nature are not constant enemies. On the contrary, while each follows rules that are fit for them, both sets of rules, Smith tells us, are in fact 'calculated to promote the same great end, the order of the world, and the perfection of human nature' (TMS III.5.9, p. 196).

Both sets of rules—the laws of nature and the 'law and morality' with which humans intervene in the natural course of things—are needed to promote human happiness for two reasons (Brubaker, 2006, p. 180). First of all, there is no human happiness without humans, and so self-preservation is a precondition for the promotion of human happiness. For our preservation, we depend on 'appetite and instinct' (Brubaker, 2006, p. 180), which are the work of our narrowly Darwinian nature. These internal representatives of nature simply ensure not only our struggle to survive, but also our constant striving to better our condition. The rules of nature are 'useful and proper for rousing the industry and attention of mankind' (TMS III.5.10, p. 168) since they typically reward industry (Cropsey, 1975, p. 141). Secondly, humans are also naturally social animals, and so the human concern to promote the perfection and happiness of human nature 'through morality, law, and society' is also, in a sense, nature's concern (Brubaker, 2006, p. 181). Immediate appetite and instinct, however, are insufficient to promote human happiness, and '[n]ature simply needs the help of human efforts to correct nature and human nature' (Brubaker, 2006, p. 181). The pursuit or realisation of human happiness 'is "prompted" by nature … but depends on the wisdom of human efforts' (Brubaker, 2006, p. 181). The impartial spectator and the love of praiseworthiness are core parts of this 'wisdom', so while nature 'sets us on the course to 'happiness and perfection' through human society' (or, perhaps: human sociality), achieving that goal means developing standards of judgement that will put us at odds with both nature simply and our own, basic desires, and the motivation to follow these judgements (Brubaker, 2006, p. 181).

A continuation of this conflict between nature simply and human nature can be found in Smith's discussion of the two metaphorical 'tribunals' in which questions of propriety are settled (TMS III.2.32, p. 150). If we understand the fundamental conflict highlighted by Brubaker as one between egotism and altruism (broadly construed), then the conflict of the two tribunals can be seen as a conflict between the first level of 'law and morality' established by human socialisation and a second level of morality, where the latter is needed to correct for errors in the former. It is a conflict between conventional morality and individual moral autonomy. Smith establishes the moral authority of the first tribunal in Part I of *TMS*, where he shows how, on the basis of sympathy, each of us is made the 'immediate judge' of others (III.2.32, p. 150). This is the 'inferior tribunal' of 'the man without' (III.2.32, p. 151, footnote 22; III.2.32, p. 150), Smith's image for the judgement of society. Smith establishes the moral authority of the second tribunal in Part III, where, on the basis of the distinction between praise and praiseworthiness, he shows how each of us judge ourselves, and in turn each other, from the assumed point of view of an impartial spectator. This is the tribunal of 'the man within' (III.2.32, p. 150), Smith's image for the judgement of our own conscience. The tribunal of the man within functions as the 'court of appeals' (Freiin von Villiez, 2011, p. 39) for ill-informed or partial judgements by actual spectators. It is therefore only by appeal to this 'much higher tribunal' (TMS III.2.32, p. 150) that we can achieve some measure of independence from common opinion, some measure of moral autonomy.

In the sixth edition of *TMS*, Smith goes on to say that '[t]he jurisdictions of those two tribunals are founded upon principles which, though in some respects resembling and akin, are, however, in reality different and distinct' (III.2.32, p. 150). The principles corresponding to the two tribunals are of course the love of praise and the love of praiseworthiness, and Smith is here reiterating his earlier statements about these being 'in many respects, distinct and independent of one another' (III.2.2, p. 132). In the second edition of *TMS*, however, Smith writes instead that 'if we enquire into the origin of [the tribunal within the breast], 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' (III.2.32, p. 152, footnote 22). In other words, the love of praiseworthiness is in a great measure derived from the love of praise. The difference between the second and the sixth edition is striking. Does it signal a change of heart on Smith's part? Or is he, more conservatively, emphasising the independence of the love of praiseworthiness as a moral motive? D.D. Raphael has argued that the differences between the first, second and sixth editions are differences in emphasis rather than in the elements of Smith's theory (Raphael, 1975, p. 94, 2007, pp. 44–45).[44] Taking this

---

44 Note that Raphael is primarily talking about the status of the impartial spectator and how Smith's descriptions of this changes throughout the different editions. Given the intimate connection between the impartial spectator and

view, I will draw on the full spectrum of Smith's varying emphases in reconstructing his argument. The result renders moot the question of Smith's intention in revising: We get a way of seeing the love of praiseworthiness as both derived and independent.

II.iv. A Love Redirected

> To be amiable and to be meritorious; that is, to deserve love and to deserve reward, are the great characters of virtue; and to be odious and punishable, of vice. But all these characters have an immediate reference to the sentiments of others. (TMS III.1.7, p. 132)

Apart from the identification of virtue with praiseworthiness, the most important thing in this passage is Smith's insistence that virtue has an *immediate* reference to the sentiments of others. He continues, 'Virtue is not said to be amiable, or to be meritorious, because it is the object of its own love, or of its own gratitude; but because it excites those sentiments in other men' (III.1.7, p. 132). This is probably one of the clearest examples of Smith's non-objectivism about moral value (his skepticism, as Griswold calls it, 1998, pp. 155–73), and it is important to understanding the relationship between love of praise and love of praiseworthiness.

In the paragraphs preceding this, Smith has sketched out the mechanism by which we make judgements about our own conduct and sentiments by analogizing it to how we judge others. We do so, Smith writes, by imaginatively inhabiting the perspective of a spectator: '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, p. 131). We divide ourselves 'as it were, into two persons … the examiner and judge [and] the person whose conduct is examined and judged of' (III.1.6, p. 131). By seeing ourselves from without in this manner, we are able to predict the judgements of others. The analogy is that of a mirror, a 'looking-glass', in which we see how we appear to others, and with the help of which we may adjust that appearance in expectation of their opinion (III.1.4, p. 130). Thanks to the 'looking-glass' of the imagined spectator, we can congratulate ourselves upon doing something for which we think 'other men' would praise us:

> The consciousness that [virtue] is the object of such favourable regards, is the source of that inward tranquillity and self-satisfaction with which it is naturally attended, as the suspicion of the contrary gives occasion to the torments of vice. What so great happiness as to be beloved, and to know that we deserve to be beloved? What so great misery as to be hated, and to know that we deserve to be hated? (II.2.7, p. 132)

---

the love of praiseworthiness, to be explored later in this article, I think the same argument can be extended to Smith's changing descriptions of this love.

In the first to fifth editions of *TMS*, the paragraph following this, and thus the first paragraph of the chapter entitled 'Of the love of praise, and that of praise-worthiness …', begins with an assertion: 'The most sincere praise can give little pleasure when it cannot be considered as some sort of proof of praise-worthiness' (III.2.4, p. 133, see footnote 4, p. 132). Having explained how the pleasure of being praiseworthy arises from the knowledge that others would praise us if they had the chance, Smith proceeds to discuss how actual praise can give little pleasure if it is not accompanied by an awareness of praiseworthiness.

At first pass, this seems paradoxical: The 'other men' are praising us, and yet we can derive no satisfaction from this because we think that they *would* not. But think of a situation where we are being praised for something we have not in fact done. In that case, our awareness that we have not done anything for which 'other men' would praise us renders the actual praise from those same 'other men' worthless. As Smith writes, '[t]he man who applauds us either for actions which we did not perform, or for motives which had no sort of influence upon our conduct, applauds not us, but another person' (III.2.4, p. 133). Here we see clearly how Smith's discussion of praiseworthiness connects to Rousseau's distinction between *être* and *paraître*, as Hanley (2009, p. 41) points out: To accept unmerited praise is to accept appearing to other people as someone we are not. Letting praise trump praiseworthiness is vanity, plain and simple (TMS III.2.4, p. 133-34). The obverse of this effect is that the mere knowledge of being praiseworthy 'often gives real comfort … though no praise should actually be bestowed upon us' (III.2.5, p. 134), and so we can experience the 'inward tranquillity and self-satisfaction' associated with the knowledge that we have acted virtuously even in the absence of actual spectators.

All the while, the tranquillity offered by our knowledge that we are either praiseworthy or at least not worthy of blame keeps its 'immediate reference to the sentiments of others' (III.1.7, p. 132). When we do not let ourselves be pleased by unmerited praise, it is because we think that those who praise us *would not* do so had they known what we know. And to the extent that we can endure unmerited blame, it is likewise because we think that all their accusations would fall to the ground *if only they knew*. So, praise- and blameworthiness trump actual praise and blame, not by referring to some objective standard of propriety that exists independently of what anyone might think of it, but by combining what we take to be the actual standard of propriety held by others with the knowledge that we ourselves have of our situation. In other words, if a *well-informed* spectator would praise us, we are praiseworthy. In this manner, our judgements from the point of view of the imagined spectator come not only to predict but to some extent to *supplant* the judgements of others. The imagined spectator thus gains a measure of independence from the actual spectators on which it is modelled.

However, other people will sometimes, even much of the time, fail to praise and blame as they would, not just because they are ill-informed, but because they are *partial*. We are partial to ourselves, and the same is, of course, true of others to themselves. But we are also partial to those close to us, people with whom we identify, certain ideas, books, hairstyles, foods, amongst other things. With these diverse partialities comes an equally diverse set of potential distortions in the way a spectator perceives and judges the conduct or sentiments of a given agent in a particular situation. Thus we often find ourselves in a situation where those 'other men' fail to judge us as they would, not because they lack some crucial bit of information about what we have or have not done (and our motives were), but because their partialities stop them from seeing the situation in the manner that they would, had they not been so partial. And so is added another adjective to the description of the supposed spectator: It must not only be well-informed but also *impartial*.

This, of course, is a double-edged adjective. In the case of being well-informed, there is really no such thing as too much information. The closer we are to the situation and the person judged, the more likely we are to be well-informed. Thus, when predicting what a well-informed spectator would think of our conduct, the fact that we are ourselves, with first-hand knowledge of our situation and our motives, offers a clear advantage. When it comes to being impartial, on the other hand, being ourselves presents more of a problem. Impartiality implies a certain degree of disinterestedness: the ability to judge 'without any prejudice generated either by one's own private emotions of the moment or by any narrow desire to 'better one's condition' through manipulation of the situation' (Griswold 1998, p. 136). That is easier said than done when that 'one' is oneself. Smith has a nice way of putting it when he, in the first edition of *TMS*, discusses the difficulty of regarding oneself with such disinterest:

> Unfortunately this moral looking-glass is not always a very good one. Common looking-glasses, it is said, are extremely deceitful, and by the glare which they throw over the face, conceal from the partial eyes of the person many deformities which are obvious to every body besides. But there is not in the world such a smoother of wrinkles as is every man's imagination, with regard to the blemishes of his own character. (TMS III.1.5, p. 131, footnote 3)

In judging ourselves, the distorting effects of self-love will often cancel out whatever benefit we get from being well-informed about the person we are judging. Nevertheless, the initial independence of the imagined spectator is at least doubled by the addition of 'impartial' to its description. Not only can judgements made from the perspective of a well-informed, impartial spectator supplant the inaccurate judgements of ill-informed, actual spectators, but taking this perspective can also serve as a check both on their *and* on our own self-preferences. Here we see emerging one of the main

'perfectionist' elements of Smith's theory (Forman-Barzilai, 2010, p. 18), namely 'the man within the breast, the great judge and arbiter of [our] conduct' (TMS III.2.32, p. 150). However, the impartial spectator is not, as Forman implies, 'exogenous' to Smith's empirical descriptions (Forman-Barzilai, 2010, p. 18). Rather, it springs forth in the mind of the individual agent as a result of the sympathetic process itself.

When Smith in the sixth edition of *TMS* begins the chapter on the love of praiseworthiness by saying that 'Man naturally desires, not only to be loved, but to be lovely; or to be that thing which is the natural and proper object of love … He desires, not only praise, but praise-worthiness', rather than asserting that 'The most sincere praise can give little pleasure when it cannot be considered as some sort of proof of praise-worthiness', (III.2.1, p. 132), we now know that this is not indicative of the 'originality' of the love of praiseworthiness. Instead, Smith is bolstering its claim to independence. With this in mind, the paragraph immediately following can also be seen in a new light:

> The love of praise-worthiness is by no means derived altogether from the love of praise. 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. (III.2.2, p. 132)

Considering this a description of the relationship between the two loves after they have entered the scene, Smith's care to distinguish the love of praiseworthiness from the love of praise makes perfect sense.[45] The love of praiseworthiness does not simply reduce to a love of praise; the judgements of a well-informed, impartial spectator not only predict, but to some extent supplant the judgements of actual spectators. The imagined spectator represents what real spectators would judge if only they had been *better* spectators. And so emerges not only the independence of the imagined impartial and well-informed spectator but also the normative priority of the judgements of the impartial spectator over those of actual spectators.[46] To be precise, the reason an impartial judge is normatively superior to a partial judge is that the impartial judge avoids giving preference to any particular agent's interest in a way that conflicts with the idea that everyone is of equal worth. That idea is inherent in the sympathetic process itself, since by seeing ourselves as a stranger sees us, we come to realise that we are neither more nor less important in their eyes than they in ours: 'When he views himself in the light in which he is conscious that others will view him, he sees that to them he is but one of the multitude in no respect better than any other in it' (II.ii.2.1, p. 97). Our ability to escape the self

---

45 Compare also Smith's use of 'altogether' here and in VI.ii.2.4, p. 270.

46 Or, as I assume Hanley would phrase it: the lexical priority.

(Griswold, 1998, p. 78) inexorably puts us on to the truth about our worth relative to that of other people, and, of course, theirs relative to ours.

Put differently, love of praiseworthiness is not derived *altogether* from the love of praise, since if praise comes apart from praiseworthiness, love of praiseworthiness will guide you away from actual praise toward self-approbation through the judgements of the impartial spectator. Love of praiseworthiness is not reducible to the love of praise; our moral motives are not fundamentally selfish. However, love of praiseworthiness *is* derived from love of praise in the sense that it is our desire to be approved of by others that teaches us, by turns, to view ourselves as others see us from the point of view of an imagined spectator, predicting what others *will* judge, then what they *would* judge had they only been well-informed, and, finally, what they *should* judge, as impartial spectators. The love of praiseworthiness is the love of the impartial spectator aspect of ourselves (Griswold, 1998, p. 133). As such, it is a love redirected.

Exegetically speaking, it might be true that the love of praiseworthiness is a love redirected. Smith, however, was an empiricist, and *TMS* is to a large extent an early work of moral psychology, analysing the mechanisms underlying moral judgements. The love of praiseworthiness is no exception to this: It is the name Smith chose for what he took to be one of the driving forces behind the kind of behaviour we usually consider to be praiseworthy or virtuous. The problem of the love of praiseworthiness is therefore also a problem of what empirical support it might have. To go beyond exegesis, and to treat the love of praiseworthiness as more than a theoretical curiosity, we should, therefore, look to modern moral psychology to test whether the argument that this love is a love redirected fits with our best current understanding of moral development.

II.v. An Inborn Moral Compass?

In the tradition established by the pioneering work of Jean Piaget and Lawrence Kohlberg, moral psychologists up until the late twentieth century tended to regard moral development as a stadial progression, beginning from an 'essentially pre-moral' obedience to authority (Vaish & Tomasello, 2013, p. 279). The question of whether morality might have some innate basis was perhaps not as much dismissed as simply passed over as empirically untestable.[47] Infants, after all, cannot speak. The practical hurdle presented by infants' lack of language was vaulted with the development of experimental paradigms relying on other clues like the amount of time the infants spend looking at different elements in the experiment. The results of this research indicated that even very young

---

47 As Vaish and Tomasello point out, Piaget was careful to note that he was studying explicit moral judgement, not moral behaviour or sentiments – the more recent research thus complements rather than competes with this tradition (2013, p. 280)

infants make moral judgements. One of the pioneering studies of infant morality found that twelve-month-old babies evaluate helping-behaviour as positive, and hindering-behaviour as negative (Premack & Premack, 1997, pp. 851–852). More recent studies have found evidence that infants as young as six months (Hamlin et al., 2007, p. 558) and three months (Hamlin et al., 2010, pp. 927–930) attribute these behaviours to agents, and that they make corresponding evaluations of the agents as a result: They prefer helpers over hinderers. Reading Karen Wynn's (2008) description of these experiments, it is tempting to label the infants as impartial spectators *in spe*, operating with an innate standard of propriety:

> Their evaluations are made on the basis of witnessed interactions between unfamiliar individuals; the infant, as an unaffected, unrelated—and therefore unbiased—third party, is nonetheless rendering an abstract judgment about the value of a social act. (p. 346)

Although the evidence for moral evaluation in very young infants has been called into question (Salvadori et al., 2015; Scarf, Imuta, Colombo, & Hayne, 2012; see Hamlin, Wynn, & Bloom, 2012 for a reply), the idea that we are either born with a rudimentary moral compass (Hamlin, 2013) or that we develop one early in childhood (Emde, Biringen, Clyman, & Oppenheim, 1991) has support from others sources. There is research showing that toddlers as young as fourteen to eighteen months are 'naturally altruistic', and will tend to help others achieve their goals 'irrespective of any reward from adults' (Warneken & Tomasello, 2009, p. 455), that infants as young as twelve months prefer equal to unequal distribution of goods among third parties and equal over unequal distributors (Geraci & Surian, 2011, pp. 1016–1017), and that in fifteen-month-olds, a sensitivity to fair distribution is correlated with altruistic behaviour (Schmidt & Sommerville, 2011, p. 5). Twenty-one-month-old babies even take into account relative merit in their evaluations of fair distributions (Sloane, Baillargeon, & Premack, 2012, p. 203). Children of ages four to seven are also averse to getting less than others in a distribution of goods (Blake & McAuliffe, 2011, p. 211), and when children three to five years old are themselves responsible for the distribution of goods, they do in fact distribute, that is, they do not simply take everything for themselves even if they could (Rochat et al., 2009, p. 427).

Could the love of praiseworthiness be an original or innate love after all? Other experiments complicate the picture. Wynn (2009) reviews a number of studies on how infants prefer individuals who are like themselves over those who are dissimilar and argues that infants are probably not as indiscriminately altruistic as Warnecken and Tomasello (2009) suppose. Hamlin, Mahajan, Liberman, & Wynn *(2013)* followed up on this by introducing an element of social identification in their experimental setup: Prior to being exposed to a scenario in which a puppet is helped or

hindered, the infant was shown that the puppet either shared or did not share the infant's taste in snacks. If the infant preferred green beans and the puppet preferred Graham crackers, both nine- and fourteen-month-old infants would subsequently want the puppet harmed rather than helped (Hamlin et al., 2013, p. 593). It is not just that infants who love beans are evil; the desire to see the dissimilar other harmed was just as strong for those who chose the crackers. More importantly, the infant's desire to see the other *helped* was just as strong as it had been in the original experiment if the puppet shared the infant's taste in snacks (Hamlin et al., 2013, p. 590). This result is hard to square with the idea of an original love of praiseworthiness 'capable of withstanding and mitigating civilization's corruptions' (Hanley, 2009, p. 140), but it is less than surprising considering the factors in the evolutionary history of humans which are likely responsible for shaping any innate tendencies humans might have, like kin (or even group) selection (Joyce, 2006, pp. 45–47; Lewontin, 1970; Vaish & Tomasello, 2013, pp. 280–282). In contrast, the interpretation that the love of praiseworthiness is a love redirected, especially if this is seen through the lens of the 'conflict within nature' that Brubaker identifies in Smith's writing, is altogether compatible with the finding that infants' innate tendencies are only imperfectly moral. If there is anything like a love of praiseworthiness in us from birth, it is very much the poor relation of the love of praiseworthiness described by Smith.

The rosy image of the fair and altruistic infant is further marred by a second set of results from studies on distributive fairness. At the same age where children object to the inequality of getting less than others, they are not averse to getting significantly more (Blake and McAuliffe 2011, p. 211; Rochat et al. 2009, p. 441). Moreover, five- to six-year-old children will choose to receive fewer goods in absolute terms if they thereby ensure that they get relatively more than an anonymous peer (Sheskin, Bloom, & Wynn, 2014). Sheskin *et al.* suggest that the observed 'anti-equality' in young children might be attributed to the influence of social comparison: children are not concerned to get as much as possible, but to get more than others, even if that means getting less than they otherwise would (2014, p. 155). Babies, it turns out, are not impartial spectators of themselves.

As for when the tendency to impartiality develops, studies of distributive fairness indicate that a change happens around the eight-year mark. Children of this age will prefer an egalitarian over unequal distribution of goods (Fehr, Bernhard, & Rockenbach, 2008, p. 1082), and reject a distribution of candies where they themselves get significantly more than an anonymous other (Blake and McAuliffe 2011, p. 211). Children also gradually develop sensitivity to the importance of impartiality in procedures (Shaw & Olson, 2014, p. 48) and an understanding of its importance in

judges (Mills & Keil, 2008, pp. 544–547). This further supports the conclusion that it takes time to develop the habit of impartial spectatorship.

Finally, and suggestively for the specific question of the development of a love of praiseworthiness, a pair of studies have looked at the disconnect between the recognition that a norm applies equally to others and to oneself and the desire or willingness to follow it. The first study found that children five to six years old are aware of fairness norms, but prefer to act selfishly; children seven to eight years old act fairly, but derive no pleasure therefrom; and children nine to ten years old both act fairly *and* derive pleasure from equal distribution (Kogut, 2012, pp. 235–236). The second study found that children three to eight years old recognise that fairness norms applied equally to themselves and others, but that the younger children failed to follow these norms in practice (C. E. Smith, Blake, & Harris, 2013, p. 7). The failure was not due to weakness of will in the moment. Asked to predict whether they would follow fairness norms in future distributions, children under seven correctly predicted that they would *not*. Children seven to eight years old both predicted that they would follow fairness norms and actually did follow them in practice (Smith et al., 2013, p. 8). The authors hypothesise that the developmental change underlying this effect is not an increased ability to inhibit the impulse to self-satisfaction, but rather an increased regard to the weight of normative considerations in situations where there is tension between norms and selfish desires (Smith et al., 2013, p. 8). In Smithian language, children only gradually develop a love of praiseworthiness that is sufficiently strong to overrule their original self-love.

In sum, even very young infants make judgements about praiseworthiness that appear to be independent of any moral education. However, the moral compass of early life puts them off course through the unchecked influence of self-love and love for similar others. It takes time and effort— about three to five years' worth—to get into the habit of viewing oneself from the perspective of the impartial spectator, and it takes even longer—about eight years—to develop a robust desire to act on what one then sees. Charles Griswold's gloss on Smith's central argument neatly captures the extent to which Smith's eighteenth-century brand of sentimentalism joins up to twenty-first-century moral psychology:

> Smith's argument is … that the fundamental structures and psychology of moral judgment provide us with a means of distinguishing between [praise- and blameworthiness] and also with a natural inclination to do so. It is an inclination to be realized through moral education, such that the impartial spectator's practical reason becomes our own, becomes (as it were) our second nature. (Griswold, 1998, p. 131)

The empirical evidence supports the exegetical conclusion: Love of praiseworthiness is not an inborn, original love, but must be developed from more basic inclinations. How, then, can we develop it?

II.vi. Self-sympathy

The key to providing a Smithian answer to this question lies in Smith's notion of mutual sympathy. Given what I said in section 2 about how the love of praiseworthiness acts as a counterbalance to the desire for harmonious relations with others, this might seem paradoxical. However, the love of praiseworthiness is itself a version of the desire for mutual sympathy.

To see this, consider how approving of your own conduct relates to sympathy. Your ability to judge of the propriety of somebody else's sentiments or conduct is based on your ability or inability to sympathise with them, to go along with them in what they are feeling and doing. Your ability to judge yourself is similarly based on your ability to sympathise with your own sentiments and conduct from the imagined point of view of an impartial spectator. If you, in viewing yourself from this point of view, find yourself to be praiseworthy, you are therefore sympathising with yourself. You are, in effect, in a state of mutual sympathy with yourself, or, more precisely, in a state of mutual sympathy between the agent and spectator aspects of yourself. As Griswold writes,

> The love of virtue is not the love of the approval of some other person, called the 'impartial spectator', but of an aspect of ourselves with which we 'sympathize.' At this level, it is a question of the self's relation to itself … The love of virtue is an outgrowth of sympathy. (Griswold, 1998, p. 133)

The love of praiseworthiness can, therefore, be seen as the desire for mutual sympathy between these two sides. The love of praiseworthiness is the desire for a state in which you, as an impartial spectator, are able to go along with yourself, as an agent, in what you are doing, feeling and thinking. It is, in short, the desire to do what you believe is right.

In her reconstruction of Smith's argument, Carola Freiin von Villiez (2006a, see also 2011) identifies the desire for sympathy as the main motivating principle in *TMS*. She divides the desire for sympathy into three different levels, corresponding to three different stages of the development of moral agency. First, there is the 'instinctive' sympathy of emotional contagion and 'affective communication' with other people, where our 'natural desire' is to be pleasing to those around us (Freiin von Villiez, 2006, pp. 149–50).[48] Second, there is the desire for sympathy with a 'virtual spectator', an internalised representative of conventional morality (Freiin von Villiez, 2006, pp.

---

48   All translations from the German original by the author.

150–51). Third, there is the desire for sympathy with the 'ideal spectator', the mode of moral judgement in which we critically reflect on the conventional norms of society, and possibly reject them as unjustified (Freiin von Villiez 2006, p. 152). The first and second levels of sympathy can be associated with the love of praise, and the second and third levels with the love of praiseworthiness. The reason why the distinction between the two loves does not neatly map onto the three levels identified by Freiin von Villiez is that a properly developed 'virtual spectator'—the target of the desire for sympathy at the second level—is an approximation to an imagined impartial spectator, and thus a source of judgements of praiseworthiness (Freiin von Villiez 2006, p. 151). What changes in terms of motivation when we move from the second to the third level is the renunciation of 'collectively justified approval' (Freiin von Villiez 2006, p. 151), or the expectation of actual, merited praise. In the words of my argument, acting purely from a desire for mutual sympathy with an ideal impartial spectator means acting purely from a love of praiseworthiness.

The love of praise and the love of praiseworthiness are thus really only names for different modes of the desire for mutual sympathy. What changes between them is the spectator whose sympathy you desire: other people or yourself, judging under the perfectionist aspiration to be an impartial spectator. This is also the most precise sense in which love of praiseworthiness is a love redirected: Moral maturation consists, at least in part, in redirecting your desire for mutual sympathy away from other people toward your own conscience.

Fostering a love of praiseworthiness thus means fostering a facility with self-sympathy. Having a facility with self-sympathy should, therefore, be positively correlated with moral autonomy. This is a testable hypothesis. Given the extent to which Smith's *Theory* accords with contemporary moral psychology, there is, I think, good reason to put it to the test.

II.vii. Conclusion

No one is born with the moral character of a Lord Chief Justice. The love of praiseworthiness is not antecedently invested in us by wise Nature, but must be cultivated from more basic natural inclinations. Fortunately, these natural inclinations are neither wholly selfish, nor wholly vain.

Developing the ability to view oneself impartially takes time, and the desire to act in accordance with what we then see even longer. Indeed, most of us will probably not develop this desire sufficiently to never be consciously and unjustifiably partial. Moreover, the love of praiseworthiness has its dark side. Insensitivity to actual praise and blame, which we could call moral arrogance, can lead us astray just as surely as can vanity, and with greater potential for dire consequences. Smith gives clear-headed analyses of this all-to-common corruption of our moral

sentiments in his discussions of false religion (TMS III.5.12, pp. 205-7) and fanaticism (III.3.43, pp. 180-81).

And yet, the two great aims of moral maturation must be the ability to see situations of moral import in the right way and to desire to act on what one then sees—to be appropriately impartial and to love praiseworthiness. According to Smithian sentimentalism, the first aim can only be achieved by continually interrogating the moral assumptions of yourself and those around you, always striving to understand the situation at hand and view it from an impartial point of view. It is to the second aim, without which the first is of little practical value, that this article speaks. The message is that we can only hope to achieve this aim if we leverage our desire for approval from others, combining it with the voice of our conscience so that we find satisfaction in our ability to go along with ourselves in our sentiments and conduct. It is not by recognising one's grim duty, but by learning to take pleasure in doing the right thing that one will develop the moral character of a Lord Chief Justice.

That, at least, is a testable hypothesis.

Bibliography

Agerström, J., & Björklund, F. (2009). Temporal Distance and Moral Concerns: Future Morally
Questionable Behavior is Perceived as More Wrong and Evokes Stronger Prosocial
Intentions. *Basic and Applied Social Psychology*, *31*(1), 49–59.
https://doi.org/10.1080/01973530802659885

Agerström, J., & Björklund, F. (2013a). Temporal Construal and Moral Motivation. In *Handbook of
Moral Motivation* (pp. 181–193). SensePublishers. Retrieved from
http://link.springer.com/chapter/10.1007/978-94-6209-275-4_11

Agerström, J., & Björklund, F. (2013b). Why People With an Eye Toward the Future Are More
Moral: The Role of Abstract Thinking. *Basic and Applied Social Psychology*, *35*(4), 373–
381. https://doi.org/10.1080/01973533.2013.803967

Ames, D. R., & Kammrath, L. K. (2004). Mind-Reading and Metacognition: Narcissism, not Actual
Competence, Predicts Self-Estimated Ability. *Journal of Nonverbal Behavior*, *28*(3), 187–
209. https://doi.org/10.1023/B:JONB.0000039649.20015.0e

Aristoteles. (1995). *The complete works of Aristotle: the revised Oxford translation. 1: ...* Princeton,
NJ: Princeton Univ. Press.

Asch, S. E. (1940). Studies in the Principles of Judgments and Attitudes: II. Determination of
Judgments by Group and by Ego Standards. *The Journal of Social Psychology*, *12*(2), 433–
465. https://doi.org/10.1080/00224545.1940.9921487

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a
unanimous majority. *Psychological Monographs: General and Applied*, *70*(9), 1–70. https://
doi.org/10.1037/h0093718

Batson, C. D. (1987). Prosocial Motivation: Is it ever Truly Altruistic? *Advances in Experimental
Social Psychology*, *20*, 65–122. https://doi.org/10.1016/S0065-2601(08)60412-8

Batson, C. D. (1990). How social an animal? The human capacity for caring. *American
Psychologist*, *45*(3), 336.

Batson, C. D., & Ahmad, N. Y. (2009a). Empathy-induced altruism: A threat to the collective good.
In S. R. Thye & E. J. Lawler (Eds.), *Altruism and Prosocial Behavior in Groups* (Vol. 26,

pp. 1–23). Emerald Group Publishing Limited. https://doi.org/10.1108/S0882-6145(2009)0000026004

Batson, C. D., & Ahmad, N. Y. (2009b). Using Empathy to Improve Intergroup Attitudes and Relations. *Social Issues and Policy Review*, *3*(1), 141–177. https://doi.org/10.1111/j.1751-2409.2009.01013.x

Batson, C. D., Duncan, B. D., Ackerman, P., Buckley, T., & Birch, K. (1981). Is empathic emotion a source of altruistic motivation? *Journal of Personality and Social Psychology*, *40*(2), 290.

Batson, C. D., Lishner, D. A., Carpenter, A., Dulin, L., Harjusola-Webb, S., Stocks, E. L., … Sampat, B. (2003). "... As you Would have Them Do Unto You": Does Imagining Yourself in the Other's Place Stimulate Moral Action? *Personality and Social Psychology Bulletin*, *29*(9), 1190–1201. https://doi.org/10.1177/0146167203254600

Baumeister, R. F., & Newman, L. S. (1994). Self-Regulation of Cognitive Inference and Decision Processes. *Personality and Social Psychology Bulletin*, *20*(1), 3–19. https://doi.org/10.1177/0146167294201001

Bee, M. (2014). The Love of One's Self. *Political Science Working Paper Series*, (60).

Bezrukova, K., Spell, C. S., Perry, J. L., & Jehn, K. A. (2016). A meta-analytical integration of over 40 years of research on diversity training evaluation. *Psychological Bulletin*, *142*(11), 1227–1274. https://doi.org/10.1037/bul0000067

Black, M. (1962). *Models and metaphors: Studies in language and philosophy*. Ithaca, NY: Cornell University Press.

Blacking, J. (1973). *How musical is man?* Seattle: University of Washington Press.

Blacking, J. (1995). *Music, culture, & experience : selected papers of John Blacking*. Chicago: University of Chicago Press.

Blair, R. J. R. (1995). A cognitive developmental approach to morality: investigating the psychopath. *Cognition*, *57*(1), 1–29. https://doi.org/10.1016/0010-0277(95)00676-P

Blake, P. R., & McAuliffe, K. (2011). "I had so much it didn't seem fair": Eight-year-olds reject two forms of inequity. *Cognition*, *120*(2), 215–224.

Broadie, A. (2006). Sympathy and the Impartial Spectator. In K. Haakonssen (Ed.), *The Cambridge Companion to Adam Smith* (pp. 158–188). Cambridge: Cambridge University Press.

Brubaker, L. (2006). Does the 'wisdom of nature' need help? In L. Montes & E. Schliesser (Eds.), *New Voices on Adam Smith* (pp. 168–192). London.

Bruneau, E. G., & Saxe, R. (2012). The power of being heard: The benefits of 'perspective-giving' in the context of intergroup conflict. *Journal of Experimental Social Psychology*, *48*(4), 855–866. https://doi.org/10.1016/j.jesp.2012.02.017

Bruni, T., Mameli, M., & Rini, R. A. (2014). The Science of Morality and its Normative Implications. *Neuroethics*, *7*(2), 159–172. https://doi.org/10.1007/s12152-013-9191-y

Campbell, T. D. (1975). Scientific Explanation and Ethical Justification in the Moral Sentiments. In A. S. Skinner & T. Wilson (Eds.), *Essays on Adam Smith* (pp. 68–82). Clarendon press.

Carrasco, Maria A. (2011). From psychology to moral normativity. *The Adam Smith Review*, *6*, 9–29.

Carrasco, María Alejandra, & Fricke, C. (2016). Adam Smith's Impartial Spectator, *13*(2), 15.

Chambers, C. (2014, July 1). Facebook Fiasco: Was Cornell University's study of 'emotional contagion' a breach of ethics? *The Guardian*. Retrieved from https://www.theguardian.com/science/head-quarters/2014/jul/01/facebook-cornell-study-emotional-contagion-ethics-breach

Chappell, T. (2014a). *Knowing what to do: imagination, virtue, and Platonism in ethics* (First edition). Oxford ; New York: Oxford University Press.

Chappell, T. (2014b). Why ethics is hard. *Journal of Moral Philosophy*, *11*(6), 704–726.

Chappell, T. D. J. (1993). The Virtues of Thrasymachus. *Phronesis*, *38*(1), 1–17.

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences*, *16*(4), 231–239. https://doi.org/10.1016/j.tics.2012.02.007

Conway, P., & Peetz, J. (2012). When Does Feeling Moral Actually Make You a Better Person? Conceptual Abstraction Moderates Whether Past Moral Deeds Motivate Consistency or Compensatory Behavior. *Personality and Social Psychology Bulletin*, *38*(7), 907–919. https://doi.org/10.1177/0146167212442394

Coplan, A. (2011). Will the Real Empathy Please Stand up? A Case for a Narrow Conceptualization. *The Southern Journal of Philosophy*, *49*, 40–65. https://doi.org/10.1111/j.2041-6962.2011.00056.x

Coplan, A., & Goldie, P. (Eds.). (2011). *Empathy: Philosophical and Psychological Perspectives* (1 edition). Oxford ; New York: Oxford University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

Cropsey, J. (1975). Adam Smith and Political Philosophy. In A. S. Skinner & T. Wilson (Eds.), *Essays on Adam Smith* (pp. 132–153). Clarendon press.

Darwall, S. (1999). Sympathetic liberalism: recent work on Adam Smith. *Philosophy & Public Affairs*, *28*(2), 139–164.

Debes, R. (2015). From Einfühlung to Empathy: Sympathy in Early Phenomenology and Psychology. In E. Schliesser (Ed.), *Sympathy: A History*. Oxford University Press. Retrieved from http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199928873.001.0001/acprof-9780199928873-chapter-15

Debes, R. (2016). Adam Smith and the Sympathetic Imagination. In R. P. Hanley (Ed.), *Adam Smith: His Life, Thought, and Legacy*. Princeton: Princeton University Press.

Debes, R. (2017). The Authority of Empathy (Or, How to Ground Sentimentalism). In R. Debes & K. R. Stueber (Eds.), *Ethical Sentimentalism: New Perspectives* (pp. 171–191). Cambridge: Cambridge University Press.

Emde, R. N., Biringen, Z., Clyman, R. B., & Oppenheim, D. (1991). The moral self of infancy: Affective core and procedural knowledge. *Developmental Review*, *11*(3), 251–270.

Emerson, R. W. (2005). *Essays by Ralph Waldo Emerson*. (E. H. L. Turpin, Ed.). Retrieved from https://www.gutenberg.org/ebooks/16643

Engelmann, J. M., Herrmann, E., Rapp, D. J., & Tomasello, M. (2016). Young children (sometimes) do the right thing even when their peers do not. *Cognitive Development*, *39*, 86–92. https://doi.org/10.1016/j.cogdev.2016.04.004

Englich, B., Mussweiler, T., & Strack, F. (2005). The Last Word in Court--A Hidden Disadvantage for the Defense. *Law and Human Behavior*, *29*(6), 705–722. https://doi.org/10.1007/s10979-005-8380-7

Epley, N., & Caruso, E. M. (2014). Perspective Taking: Misstepping Into Others' Shoes. In *Handbook of Imagination and Mental Simulation*. Routledge. https://doi.org/10.4324/9780203809846.ch20

Eyal, T., & Liberman, N. (2012). Morality and psychological distance: A construal level theory perspective. In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 185–202). Washington, DC, US: American Psychological Association. https://doi.org/10.1037/13091-010

Eyal, T., Liberman, N., & Trope, Y. (2008). Judging near and distant virtue and vice. *Journal of Experimental Social Psychology*, *44*(4), 1204–1209. https://doi.org/10.1016/j.jesp.2008.03.012

Eyal, T., Liberman, N., & Trope, Y. (2014). Thinking of why a transgression occurred may drawn attention to extenuating circumstances: A comment on Žeželj & Jokić replication. *Social Psychology*, *45*(4), 329–331.

Fehr, E., Bernhard, H., & Rockenbach, B. (2008). Egalitarianism in young children. *Nature*, *454*(7208), 1079–1083. https://doi.org/10.1038/nature07155

FFU. (2017, November 7). Veiledende retningslinjer for artikkelbasert avhandling. Det humanistiske fakultet. Retrieved from https://www.uib.no/hf/23852/skjema-reglement-og-retningslinjer#regler-og-retningslinjer

Findlay, L. C., Girardi, A., & Coplan, R. J. (2006). Links between empathy, social behavior, and social understanding in early childhood. *Early Childhood Research Quarterly*, *21*(3), 347–359. https://doi.org/10.1016/j.ecresq.2006.07.009

Firth, R. (1952). Ethical Absolutism and the Ideal Observer. *Philosophy and Phenomenological Research*, *12*(3), 317–345. https://doi.org/10.2307/2103988

Fleischacker, S. (2011). Adam Smith and cultural relativism. *Erasmus Journal for Philosophy and Economics*, *4*(2), 20–41.

Fleischacker, S. (2016). Adam Smith's Impartial Spectator: Symposium Remarks. *Econ Journal Watch*, *13*(2), 273–283.

Fleischacker, S. (2017). Adam Smith's Moral and Political Philosophy. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2017). Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/spr2017/entries/smith-moral-political/

Forman-Barzilai, F. (2010). *Adam Smith and the Circles of Sympathy: Cosmopolitanism and Moral Theory*. Cambridge: Cambridge University Press.

Frazer, M. L. (2010). *The Enlightenment of Sympathy: Justice and the Moral Sentiments in the Eighteenth Century and Today* (Kindle edition). Oxford University Press, USA.

Frazer, M. L. (2016). Natural and Artificial Impartiality. *Econ Journal Watch*, *13*(2), 284–297.

Freiin von Villiez, C. (2006a). *Dimensionen der Unparteilichkeit: Adam Smith auf der Suche nach dem moralischen Standpunkt* (Habilitationsschrift). Universität Bremen.

Freiin von Villiez, C. (2006b). Double standard - Naturally! Smith and Rawls: a comparison of methods. In L. Montes & E. Schliesser (Eds.), *New Voices on Adam Smith* (pp. 115–139). London: Routledge.

Freiin von Villiez, C. (2011). Adam Smith's Story of Moral Progress. *The Adam Smith Review*, *6*, 30–45.

Fricke, C. (2011a). Adam Smith and 'the most sacred rules of justice'. *The Adam Smith Review*, *6*, 46–74.

Fricke, C. (2011b). Adam Smith and 'the Most Sacred Rules of Justice'. In F. Forman-Barzilai (Ed.), *The Adam Smith Review* (Vol. 6, pp. 46–74).

Galinsky, A. D., Ku, G., & Wang, C. S. (2005). Perspective-Taking and Self-Other Overlap: Fostering Social Bonds and Facilitating Social Coordination. *Group Processes & Intergroup Relations*, *8*(2), 109–124. https://doi.org/10.1177/1368430205051060

Galinsky, A. D., & Moskowitz, G. B. (2000). Perspective-taking: Decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Personality and Social Psychology*, *78*(4), 708–724. https://doi.org/10.1037//0022-3514.78.4.708

Geraci, A., & Surian, L. (2011). The developmental roots of fairness: infants' reactions to equal and unequal distributions of resources. *Developmental Science*, *14*(5), 1012–1020. https://doi.org/10.1111/j.1467-7687.2011.01048.x

Gilbert, D. T., & Wilson, T. D. (2007). Prospection: Experiencing the Future. *Science*, *317*(5843), 1351–1354. https://doi.org/10.1126/science.1144161

Gilbert, Daniel T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, *117*(1), 21–38. https://doi.org/10.1037/0033-2909.117.1.21

Gino, F., & Galinsky, A. D. (2012). Vicarious dishonesty: When psychological closeness creates distance from one?s moral compass. *Organizational Behavior and Human Decision Processes*, *119*(1), 15–26. https://doi.org/10.1016/j.obhdp.2012.03.011

Gong, H., & Medin, D. L. (2012). Construal levels and moral judgment: Some complications. *Judgment and Decision Making*, *7*(5), 628.

Griswold, C. L. (1998). *Adam Smith and the virtues of enlightenment* (Kindle edition). Cambridge: Cambridge University Press.

Griswold, C. L. (2010). Smith and Rousseau in Dialogue. In V. Brown & S. Fleischacker (Eds.), *The Adam Smith Review: The Philosophy of Adam Smith: Essays commemorating the 250th anniversary of The Theory of Moral Sentiments* (Vol. 5, pp. 59–84). Abingdon, Oxon: Routledge.

Guevara, B. B. de. (2016). Journeys to the limits of first-hand knowledge: politicians' on-site visits in zones of conflict and intervention. *Journal of Intervention and Statebuilding*, *10*(1), 56–76. https://doi.org/10.1080/17502977.2015.1137394

Haakonssen, K. (2002). Introduction. In *The Theory of Moral Sentiments (Cambridge Texts in the History of Philosophy)* (pp. vii–xxiv). Cambridge: Cambridge University Press.

Hahn, U., & Harris, A. J. L. (2014). What Does It Mean to be Biased. In *Psychology of Learning and Motivation* (Vol. 61, pp. 41–102). Elsevier. https://doi.org/10.1016/B978-0-12-800283-4.00002-2

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814–834.

Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, *65*(4), 613.

Hamlin, J. K. (2013). Moral Judgment and Action in Preverbal Infants and Toddlers Evidence for an Innate Moral Core. *Current Directions in Psychological Science*, *22*(3), 186–193. https://doi.org/10.1177/0963721412470687

Hamlin, J. K., Mahajan, N., Liberman, Z., & Wynn, K. (2013). Not Like Me = Bad Infants Prefer Those Who Harm Dissimilar Others. *Psychological Science*, *24*(4), 589–594. https://doi.org/10.1177/0956797612457785

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*(7169), 557–559. https://doi.org/10.1038/nature06288

Hamlin, J. K., Wynn, K., & Bloom, P. (2010). 3-month-olds show a negativity bias in their social evaluations. *Developmental Science*, *13*(6), 923–929. https://doi.org/10.1111/j.1467-7687.2010.00951.x

Hamlin, J. K., Wynn, K., & Bloom, P. (2012). Reply to Scarf et al.: Nuanced social evaluation: Association doesn't compute. *Proceedings of the National Academy of Sciences*, *109*(22), E1427–E1427. https://doi.org/10.1073/pnas.1204712109

Hanley, R. P. (2009). *Adam Smith and the character of virtue* (Kindle Edition). Cambridge: Cambridge University Press.

Hansen, K., Gerbasi, M., Todorov, A., Kruse, E., & Pronin, E. (2014). People Claim Objectivity After Knowingly Using Biased Strategies. *Personality and Social Psychology Bulletin*, *40*(6), 691–699. https://doi.org/10.1177/0146167214523476

Haraway, D. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, *14*(3), 575–599. https://doi.org/10.2307/3178066

Hare, R. M. (1981). *Moral Thinking: Its Levels, Method, and Point*. Oxford, UNITED KINGDOM: Oxford University Press USA - OSO. Retrieved from http://ebookcentral.proquest.com/lib/bergen-ebooks/detail.action?docID=728892

Harman, G. (1999). Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error. *Proceedings of the Aristotelian Society*, *99*, 315–331. https://doi.org/10.2307/4545312

Harvey, N. (1997). Confidence in judgment. *Trends in Cognitive Sciences*, *1*(2), 78–82.

Hastorf, A. H., & Cantril, H. (1954). They saw a game; a case study. *The Journal of Abnormal and Social Psychology*, *49*(1), 129–134. https://doi.org/10.1037/h0057880

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1993). Emotional Contagion. *Current Directions in Psychological Science*, *2*(3), 96–100. https://doi.org/10.1111/1467-8721.ep10770953

Haynes, B., & Cooke, P. (2015, December 4). Pitch. Retrieved 12 April 2015, from http://www.oxfordmusiconline.com/subscriber/article/grove/music/40883

Hepach, R., Vaish, A., & Tomasello, M. (2013). Young children sympathize less in response to unjustified emotional distress. *Developmental Psychology*, *49*(6), 1132. https://doi.org/10.1037/a0029501

Hepach, R., & Westermann, G. (2013). Infants' sensitivity to the congruence of others' emotions and actions. *Journal of Experimental Child Psychology*, *115*(1), 16–29. https://doi.org/10.1016/j.jecp.2012.12.013

Hesse, M. B. (1966). *Models and analogies in science*. University of Notre Dame Press.

Hobbes, T. (2009). *Leviathan: The Matter, Forme, & Power of a Common-Wealth Ecclesiastical and Civill*. The Floating Press.

Horowitz, I. A., & Kirkpatrick, L. C. (1996). A concept in search of a definition: The effects of reasonable doubt instructions on certainty of guilt standards and jury verdicts. *Law and Human Behavior*, *20*(6), 655–670. https://doi.org/10.1007/BF01499236

Hume, D. (2007). *A treatise of human nature: a critical edition*. (D. F. Norton & M. J. Norton, Eds.). Oxford : New York: Clarendon Press ; Oxford University Press.

Ichheiser, G. (1949a). Analysis and Typology of Personality Misinterpretations. *American Journal of Sociology*, *55*(2), 26–56.

Ichheiser, G. (1949b). The Image of the Other Man. *American Journal of Sociology*, *55*(2), 5–11.

Jahoda, G. (2005). Theodor Lipps and the shift from "sympathy" to "empathy". *Journal of the History of the Behavioral Sciences*, *41*(2), 151–163. https://doi.org/10.1002/jhbs.20080

Johnson, M. (2014). *Moral Imagination: Implications of Cognitive Science for Ethics* (1 edition). University of Chicago Press.

Joyce, R. (2006). *The Evolution of Morality*. Cambridge, Mass.: The MIT Press.

Kawall, J. (2006). On the Moral Epistemology of Ideal Observer Theories. *Ethical Theory and Moral Practice*, *9*(3), 359–374. https://doi.org/10.1007/s10677-006-9016-8

Kennedy, G. (2009). Adam Smith and the invisible hand: From metaphor to myth. *Econ Journal Watch*, *6*(2), 239–263.

Kennedy, K. A., & Pronin, E. (2008). When Disagreement Gets Ugly: Perceptions of Bias and the Escalation of Conflict. *Personality and Social Psychology Bulletin*, *34*(6), 833–848. https://doi.org/10.1177/0146167208315158

Klein, D. B., & Clark, M. (2011). The Music of Social Intercourse: Synchrony in Adam Smith. *The Independent Review*, *15*(3), 413–420.

Kogut, T. (2012). Knowing what I should, doing what I want: From selfishness to inequity aversion in young children's sharing behavior. *Journal of Economic Psychology*, *33*(1), 226–236. https://doi.org/10.1016/j.joep.2011.10.003

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*(24), 8788–8790. https://doi.org/10.1073/pnas.1320040111

Krueger, J. I., & Funder, D. C. (2004). Towards a balanced social psychology: Causes, consequences, and cures for the problem-seeking approach to social behavior and cognition. *Behavioral and Brain Sciences*, *27*(3), 313–327. https://doi.org/10.1017/S0140525X04000081

Kruger, J., & Evans, M. (2004). If you don't want to be late, enumerate: Unpacking reduces the planning fallacy. *Journal of Experimental Social Psychology*, *40*(5), 586–598. https://doi.org/10.1016/j.jesp.2003.11.001

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498.

Levine, L. J., Lench, H. C., & Safer, M. A. (2009). Functions of remembering and misremembering emotion. *Applied Cognitive Psychology*, *23*(8), 1059–1075. https://doi.org/10.1002/acp.1610

Lewontin, R. C. (1970). The Units of Selection. *Annual Review of Ecology and Systematics*, *1*, 1–18.

Liberman, N., Trope, Y., McCrea, S. M., & Sherman, S. J. (2007). The effect of level of construal on the temporal distance of activity enactment. *Journal of Experimental Social Psychology*, *43*(1), 143–149. https://doi.org/10.1016/j.jesp.2005.12.009

Lilienfeld, S. O., Ammirati, R., & Landfield, K. (2009). Giving Debiasing Away: Can Psychological Research on Correcting Cognitive Errors Promote Human Welfare? *Perspectives on Psychological Science*, *4*(4), 390–398. https://doi.org/10.1111/j.1745-6924.2009.01144.x

Long, E. C., & Andrews, D. W. (1991). Perspective taking as a predictor of marital adjustment. *Journal of Personality and Social Psychology*, *59*(1), 126. https://doi.org/10.1037/0022-3514.59.1.126

Luguri, J. B., Napier, J. L., & Dovidio, J. F. (2012). Reconstruing Intolerance: Abstract Thinking Reduces Conservatives' Prejudice Against Nonnormative Groups. *Psychological Science*, *23*(7), 756–763. https://doi.org/10.1177/0956797611433877

Mackie, J. L. (1977). *Ethics: inventing right and wrong*. Harmondsworth: Penguin books.

Maibom, H. L. (2008). The Mad, the Bad, and the Psychopath. *Neuroethics*, *1*(3), 167–184. https://doi.org/10.1007/s12152-008-9013-9

Maibom, H. L. (Ed.). (2017). *The Routledge handbook of philosophy of empathy*. London ; New York: Routledge, Taylor & Francis Group.

Malle, B. F. (2006). The actor-observer asymmetry in attribution: A (surprising) meta-analysis. *Psychological Bulletin*, *132*(6), 895–919. https://doi.org/10.1037/0033-2909.132.6.895

Malloch, S., & Trevarthen, C. (2009). *Communicative musicality : exploring the basis of human companionship*. Oxford University Press.

Mandeville, B. (2011). *The Fable of the Bees or Private Vices, Publick Benefits* (Vol. 1). Liberty Fund, Inc.

Mårtensson, E. (2017). Construal Level Theory and Moral Judgments: How Thinking Abstractly Modifies Morality. *Journal of European Psychology Students*, *8*(1), 30–40. https://doi.org/10.5334/jeps.413

McHugh, J. W. (2011). Relaxing a Tension in Adam Smith's Account of Sympathy. *Journal of Scottish Philosophy*, *9*(2), 189–204. https://doi.org/10.3366/jsp.2011.0015

McPherson Frantz, C. (2006). I AM Being Fair: The Bias Blind Spot as a Stumbling Block to Seeing Both Sides. *Basic and Applied Social Psychology*, *28*(2), 157–167. https://doi.org/10.1207/s15324834basp2802_5

McPherson Frantz, C., & Janoff-Bulman, R. (2000). Considering Both Sides: The Limits of Perspective Taking. *Basic and Applied Social Psychology*, *22*(1), 31–42. https://doi.org/10.1207/S15324834BASP2201_4

Mesoudi, A., Whiten, A., & Dunbar, R. (2006). A bias for social information in human cultural transmission. *British Journal of Psychology*, *97*(3), 405–423. https://doi.org/10.1348/000712605X85871

Messervey, D. L. (2008). *Time and time again: cultural differences in construal levels* (PhD Thesis).

Meyer, M. L., Taylor, S. E., & Lieberman, M. D. (2015). Social working memory and its distinctive link to social cognitive ability: an fMRI study. *Social Cognitive and Affective Neuroscience*, *10*(10), 1338–1347. https://doi.org/10.1093/scan/nsv065

Meyers, D. T. (2017). A Modest Feminist Sentimentalism. In R. Debes & K. R. Stueber (Eds.), *Ethical Sentimentalism: New Perspectives* (pp. 210–229). Cambridge: Cambridge University Press.

Milgram, S. (1963). Behavioral Study of obedience. *The Journal of Abnormal and Social Psychology*, *67*(4), 371–378. https://doi.org/10.1037/h0040525

Miller, J. G. (1984). Culture and the development of everyday social explanation. *Journal of Personality and Social Psychology*, 961–978.

Mills, C. M., & Keil, F. C. (2008). Children's developing notions of (im)partiality. *Cognition*, *107*(2), 528–551. https://doi.org/10.1016/j.cognition.2007.11.003

Moody-Adams, M. M. (2002). *Fieldwork in familiar places : morality, culture, and philosophy* (2nd ed.). Cambridge, Mass.: Harvard University Press.

Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal theories of emotion: State of the art and future development. *Emotion Review*, *5*(2), 119–124.

Nahemow, L., & Lawton, M. P. (1975). Similarity and Propinquity in Friendship Formation. *Journal of Personality and Social Psychology*, *32*(2), 205–213. https://doi.org/10.1037/0022-3514.32.2.205

Nanay, B. (2010). Adam Smith's concept of sympathy and its contemporary interpretations. In V. Brown & S. Fleischacker (Eds.), *The Philosophy of Adam Smith: The Adam Smith Review, Volume 5: Essays Commemorating the 250th Anniversary of The Theory of Moral Sentiments* (pp. 85–105). Abingdon, Oxon: Routledge.

Nersessian, N. J. (1999). Model-Based Reasoning in Conceptual Change. In L. Magnani, N. J. Nersessian, & P. Thagard (Eds.), *Model-Based Reasoning in Scientific Discovery* (pp. 5–22). Springer US. https://doi.org/10.1007/978-1-4615-4813-3_1

Newcomb, T. M. (1956). The Prediction of Interpersonal Attraction. *The American Psychologist*, *11*(11), 575–586. http://dx.doi.org/10.1037/h0046141

Nussbaum, S., Liberman, N., & Trope, Y. (2006). Predicting the near and distant future. *Journal of Experimental Psychology: General*, *135*(2), 152–161. https://doi.org/10.1037/0096-3445.135.2.152

Obama, B. (2006). Obama to Graduates: Cultivate Empathy: Northwestern University News. Retrieved 6 February 2019, from https://perma.cc/4WBN-MTP6

Osgood, C. E. (1959). Suggestions for winning the real war with communism. *Journal of Conflict Resolution*, *3*(4), 295–325.

Pashler, H., Coburn, N., & Harris, C. R. (2012). Priming of Social Distance? Failure to Replicate Effects on Social and Food Judgments. *PLoS ONE*, *7*(8), e42510. https://doi.org/10.1371/journal.pone.0042510

Persson, I., & Savulescu, J. (2018). The Moral Importance of Reflective Empathy. *Neuroethics*, *11*(2), 183–193. https://doi.org/10.1007/s12152-017-9350-7

Peters, K., Luck, L., Hutchinson, M., Wilkes, L., Andrew, S., & Jackson, D. (2011). The emotional sequelae of whistleblowing: findings from a qualitative study. *Journal of Clinical Nursing*, *20*(19–20), 2907–2914. https://doi.org/10.1111/j.1365-2702.2011.03718.x

Petty, R. E., Briñol, P., Tormala, Z. L., & Wegener, D. T. (2007). The Role of Meta-Cognition in Social Judgment. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: handbook of basic principles* (2nd ed.). New York: Guilford Press.

Plato. (2000). *Plato: 'The Republic'*. (G. R. F. Ferrari, Ed., T. Griffith, Trans.) (1 edition). Cambridge University Press.

Premack, D., & Premack, A. J. (1997). Infants Attribute Value± to the Goal-Directed Actions of Self-propelled Objects. *Journal of Cognitive Neuroscience*, *9*(6), 848–856. https://doi.org/10.1162/jocn.1997.9.6.848

Prinz, J. J. (2007). *The Emotional Construction of Morals*. Oxford, New York: Oxford University Press.

Prinz, J. J. (2011). Is Empathy Necessary for Morality? In A. Coplan & P. Goldie (Eds.), *Empathy: Philosophical and Psychological Perspectives*. Oxford University Press. Retrieved from http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199539956.001.0001/ acprof-9780199539956-chapter-14

Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the Eye of the Beholder: Divergent Perceptions of Bias in Self Versus Others. *Psychological Review*, *111*(3), 781–799. https://doi.org/10.1037/0033-295X.111.3.781

Pronin, E., Lin, D. Y., & Ross, L. (2002). The Bias Blind Spot: Perceptions of Bias in Self Versus Others. *Personality and Social Psychology Bulletin*, *28*(3), 369–381. https://doi.org/10.1177/0146167202286008

Pulp. (1995). *Common People*. Island Records.

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

Raphael, D. D. (2007). *The impartial spectator: Adam Smith's moral philosophy*. Oxford : Oxford ; New York: Clarendon Press ; Oxford University Press.

Rasmussen, D. C. (2017). *The Infidel and the Professor: David Hume, Adam Smith, and the Friendship That Shaped Modern Thought* (Reprint edition). Princeton University Press.

Rawls, J. (1999). *A theory of justice* (2nd ed.). Cambridge, Mass: The Belknap Press of Harvard University Press.

Redman, D. A. (1993). Adam Smith and Isaac Newton. *Scottish Journal of Political Economy*, *40*(2), 210–230. https://doi.org/10.1111/j.1467-9485.1993.tb00651.x

Rochat, P., Dias, M. D., Liping, G., Broesch, T., Passos-Ferreira, C., Winning, A., & Berg, B. (2009). Fairness in distributive justice by 3-and 5-year-olds across seven cultures. *Journal of Cross-Cultural Psychology*, *40*(3), 416–442.

Ross, L. (1977). The Intuitive Psychologist And His Shortcomings: Distortions in the Attribution Process. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 10, pp. 173–220). Academic Press. https://doi.org/10.1016/S0065-2601(08)60357-3

Ross, L., & Ward, A. (1995). Psychological Barriers to Dispute Resolution. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 27, pp. 255–304). Academic Press. https://doi.org/10.1016/S0065-2601(08)60407-4

Rozin, P., & Royzman, E. B. (2001). Negativity Bias, Negativity Dominance, and Contagion. *Personality and Social Psychology Review*, *5*(4), 296–320. https://doi.org/10.1207/S15327957PSPR0504_2

Ruffman, T., Lorimer, B., & Scarf, D. (2017). Do Infants Really Experience Emotional Contagion? *Child Development Perspectives*, *11*(4), 270–274. https://doi.org/10.1111/cdep.12244

Sabini, J., Siepmann, M., & Stein, J. (2001). The Really Fundamental Attribution Error in Social Psychological Research. *Psychological Inquiry*, *12*(1), 1–15. https://doi.org/10.1207/S15327965PLI1201_01

Salvadori, E., Blazsekova, T., Volein, A., Karap, Z., Tatone, D., Mascaro, O., & Csibra, G. (2015). Probing the Strength of Infants' Preference for Helpers over Hinderers: Two Replication Attempts of Hamlin and Wynn (2011). *PLOS ONE*, *10*(11), e0140570. https://doi.org/10.1371/journal.pone.0140570

Sayre-McCord, G. (1994). On Why Hume's "General Point of View" Isn't Ideal–and Shouldn't Be. *Social Philosophy and Policy*, *11*(01), 202–228. https://doi.org/10.1017/S0265052500004350

Sayre-McCord, G. (2010). Sentiments and spectators: Adam Smith's theory of moral judgment. In V. Brown & S. Fleischacker (Eds.), *The Philosophy of Adam Smith: The Adam Smith*

Review, Volume 5: Essays Commemorating the 250th Anniversary of The Theory of Moral Sentiments (pp. 123–144). Abingdon, Oxon: Routledge.

Sayre-McCord, G. (2013). Hume and Smith on Sympathy, Approbation, and Moral Judgement. *Social Philosophy and Policy*, *30*(1–2), 208–236. https://doi.org/10.1017/S0265052513000101

Scarf, D., Imuta, K., Colombo, M., & Hayne, H. (2012). Social Evaluation or Simple Association? Simple Associations May Explain Moral Reasoning in Infants. *PLOS ONE*, *7*(8), e42698. https://doi.org/10.1371/journal.pone.0042698

Schliesser, E. (2008). Hume's Newtonianism and Anti-Newtonianism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2008). Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/win2008/entries/hume-newton/

Schmidt, M. F. H., & Sommerville, J. A. (2011). Fairness Expectations and Altruistic Sharing in 15-Month-Old Human Infants. *PLoS ONE*, *6*(10), e23223. https://doi.org/10.1371/journal.pone.0023223

Semi, M. (2012). *Music as a Science of Mankind in Eighteenth-Century Britain*. (T. Keates, Trans.). Farnham: Ashgate Publishing Ltd.

Shaw, A., & Olson, K. (2014). Fairness as partiality aversion: The development of procedural justice. *Journal of Experimental Child Psychology*, *119*, 40–53. https://doi.org/10.1016/j.jecp.2013.10.007

Sheskin, M., Bloom, P., & Wynn, K. (2014). Anti-equality: Social comparison in young children. *Cognition*, *130*(2), 152–156. https://doi.org/10.1016/j.cognition.2013.10.008

Sloane, S., Baillargeon, R., & Premack, D. (2012). Do infants have a sense of fairness? *Psychological Science*, *23*(2), 196–204. https://doi.org/10.1177/0956797611422072

Small, C. (1998). *Musicking: The meanings of performing and listening*. Wesleyan University Press.

Smith, A. (1982). *Essays on philosophical subjects*. (W. P. D. Wightman, J. C. Bryce, & I. S. Ross, Eds.). Indianapolis: Liberty Classics.

Smith, A. (1985). *Lectures on rhetoric and belles lettres*. (J. C. Bryce, Ed.). Indianapolis: Liberty Classics.

Smith, A. (2002). *The Theory of Moral Sentiments (Cambridge Texts in the History of Philosophy)*. (K. Haakonssen, Ed.). Cambridge: Cambridge University Press.

Smith, C. E., Blake, P. R., & Harris, P. L. (2013). I Should but I Won't: Why Young Children Endorse Norms of Fair Sharing but Do Not Follow Them. *PLOS ONE*, *8*(3), e59510. https://doi.org/10.1371/journal.pone.0059510

Stewart, D. (1861). Account of the life and writings of Adam Smith. In *The theory of moral sentiments; or, an essay towards an analysis of the principles by which men naturally judge concerning the conduct and character, first of their neighbors, and afterwards of themselves, to which is added a dissertation on the origin of languages (new ed.)* (pp. xi–lxix). London, Great Britain: Henry G Bohn.

Stiller, J., & Dunbar, R. I. M. (2007). Perspective-taking and memory capacity predict social network size. *Social Networks*, *29*(1), 93–104. https://doi.org/10.1016/j.socnet.2006.04.001

Stinson, L., & Ickes, W. (1992). Empathic accuracy in the interactions of male friends versus male strangers. *Journal of Personality and Social Psychology*, *62*(5), 787. https://doi.org/10.1037/0022-3514.62.5.787

Stravinsky, I. (1947). *Poetics of Music*. Cambridge, Mass.: Harvard University Press.

Strawson, P. F. (1987). *Scepticism and Naturalism: Some Varieties*. Abingdon, UNITED KINGDOM: Routledge. Retrieved from http://ebookcentral.proquest.com/lib/bergen-ebooks/detail.action?docID=235356

Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences*, *30*(3), 299–313. https://doi.org/10.1017/S0140525X07001975

Thagard, P., & Beam, C. (2004). Epistemological Metaphors and the Nature of Philosophy. *Metaphilosophy*, *35*(4), 504–516.

Thorstensen, E. (2011). Mennesket, moralen og forståelsen. Maksimen «tout comprendre, c'est tout pardonner» i folkemordsforskningen sett i lys av dens historie. *Etter Lemkin*, *3*. Retrieved from https://www.academia.edu/6028729/Mennesket_moralen_og_forst%C3%A5elsen._Maksimen_tout_comprendre_cest_tout_pardonner_i_folkemordsforskningen_sett_i_lys_av_dens_historie

Trope, Y., & Liberman, N. (2010). Construal-Level Theory of Psychological Distance. *Psychological Review*, *117*(2), 440–463. https://doi.org/10.1037/a0018963

Ugazio, G., Majdandzic, J., & Lamm, C. (2014). Are empathy and morality linked? Insights from moral psychology, social and decision neuroscience, and philosophy. *Empathy in Morality*, 155–171.

Uhlmann, E. L., & Cohen, G. L. (2007). "I think it, therefore it's true": Effects of self-perceived objectivity on hiring discrimination. *Organizational Behavior and Human Decision Processes*, *104*(2), 207–223. https://doi.org/10.1016/j.obhdp.2007.07.001

Vaish, A., & Tomasello, M. (2013). The Early Ontogeny of Human Cooperation and Morality. In M. Killen & J. G. Smetana (Eds.), *Handbook of Moral Development* (pp. 279–298). Psychology Press.

Ward, A., Ross, L., Reed, E., Turiel, E., & Brown, T. (1997). Naive realism in everyday life: Implications for social conflict and misunderstanding. *Values and Knowledge*, 103–135.

Warneken, F., & Tomasello, M. (2009). The roots of human altruism. *British Journal of Psychology*, *100*(3), 455–471. https://doi.org/10.1348/000712608X379061

Watts, J. (2018, December 4). U-turn on eco-tax rise gives President Macron fuel for thought. *The Guardian*. Retrieved from https://www.theguardian.com/environment/2018/dec/04/macron-u-turn-on-eco-tax-rise-gives-green-lobby-fuel-for-thought

Weinstein, J. R. (2016). My Understanding of Adam Smith's Impartial Spectator. *Econ Journal Watch*, *351–358*(2), 8.

Williams, B. (2006). *Ethics and the Limits of Philosophy*. Routledge. https://doi.org/10.4324/9780203969847

Williams, L. E., & Bargh, J. A. (2008). Keeping One's Distance: The Influence of Spatial Distance Cues on Affect and Evaluation. *Psychological Science*, *19*(3), 302–308. https://doi.org/10.1111/j.1467-9280.2008.02084.x

Williams, L. E., Stein, R., & Galguera, L. (2014). The Distinct Affective Consequences of Psychological Distance and Construal Level. *Journal of Consumer Research*, *40*(6), 1123–1138. https://doi.org/10.1086/674212

Wilson, J. M., Boyer O'Leary, M., Metiu, A., & Jett, Q. R. (2008). Perceived Proximity in Virtual Work: Explaining the Paradox of Far-but-Close. *Organization Studies*, *29*(7), 979–1002. https://doi.org/10.1177/0170840607083105

Wong, V. C., & Wyer, R. S. (2016). Mental traveling along psychological distances: The effects of cultural syndromes, perspective flexibility, and construal level. *Journal of Personality and Social Psychology*, *111*(1), 17–33. https://doi.org/10.1037/pspa0000048

Wynn, K. (2008). Some innate foundations of social and moral cognition. *The Innate Mind: Foundations and the Future*, 330–347.

Wynn, K. (2009). Constraints on natural altruism. *British Journal of Psychology*, *100*(3), 481–485. https://doi.org/10.1348/000712609X441312

Yates, R. (2011). *Revolutionary road*. London: Vintage Classic.

Zaki, J. (2015, October 20). Choosing Empathy: A Conversation With Jamil Zaki (Edge.com). Retrieved 27 February 2019, from https://perma.cc/M624-X2Y5

Zaki, J., & Ochsner, K. N. (2012). The neuroscience of empathy: progress, pitfalls and promise. *Nature Neuroscience*, *15*(5), 675–680. https://doi.org/10.1038/nn.3085

Žeželj, I. L., & Jokić, B. R. (2014). Replication of Experiments Evaluating Impact of Psychological Distance on Moral Judgment. *Social Psychology*, *45*(3), 223–231. https://doi.org/10.1027/1864-9335/a000188

uib.no