ADAM SMITH AND THE MODERN SCIENCE OF ETHICS

James Konow

Economics and Philosophy / Volume 28 / Issue 03 / November 2012, pp 333 - 362
DOI: 10.1017/S0266267112000272, Published online:

Link to this article: http://journals.cambridge.org/abstract_S0266267112000272

How to cite this article:
James Konow (2012). ADAM SMITH AND THE MODERN SCIENCE OF ETHICS. Economics and Philosophy, 28, pp 333-362 doi:10.1017/S0266267112000272

Economics and Philosophy, **28** (2012) 333–362
doi:10.1017/S0266267112000272

ADAM SMITH AND THE MODERN SCIENCE OF ETHICS

JAMES KONOW
Kiel University and Kiel Institute for the World Economy, Germany and Loyola Marymount University, USA
jkonow@lmu.edu

Third-party decision-makers, or *spectators*, have emerged as a useful empirical tool in modern social science research on moral motivation. Spectators of a sort also serve a central role in Adam Smith's moral theory. This paper compares these two types of spectatorship with respect to their goals, methodologies, visions of human nature and emphasis on moral rules. I find important similarities and differences and conclude that this comparison suggests significant opportunities for philosophical ethics to inform empirical and theoretical research on moral preferences and vice versa.

> The general maxims of morality are formed, like all other general maxims, from experience and induction. We observe in a great variety of particular cases what pleases or displeases our moral faculties, what these approve or disapprove of, and, by induction from this experience, we establish those general rules . . . In treating of the rules of morality, in this manner, consists the science which is properly called Ethics. (Adam Smith, 1976 [1759] *The Theory of Moral Sentiments*, VII.iii.2.6, VII.iv.6)

This paper is one product of the 2009 conference celebrating the 250th anniversary of the publication of *The Theory of Moral Sentiments* at the *Centre for the Study of Mind in Nature* in Oslo. I wish to thank two referees and the Editor of this journal, Christian List, for very helpful and constructive comments. I also wish to acknowledge very useful feedback on earlier versions from Martin Binder, Matthew Braham, Maria Carrasco, Thomas Cushman, Sam Fleischacker, Christel Fricke, John O'Neill, Mozaffar Qizilbash, Jonathan Riley, Christian Schubert, Bob Sugden and Viktor Vanberg. Any shortcomings remain, of course, the sole property of the author.

One of the most dramatic developments in economics over the past few decades has been the rapidly increasing willingness of economists to extend their models of human motivation beyond the traditional assumption of narrow self-interest and to incorporate moral and other social preferences. This has been prompted mostly by results from experiments, originally conducted to test predictions of the canonical model but subsequently also designed to inform new or modified theories. Most of this empirical work has involved *stakeholders*, or parties whose personal stakes are affected by their decisions. A fairly recent addition to the toolkit of these researchers, however, is the use of *spectators*, or third parties who make decisions affecting others but not themselves. Spectators of a kind were also at the centre of the moral theory Adam Smith explicated more than 250 years ago in his *The Theory of Moral Sentiments*, or *TMS* (1759). Smith's characterization of ethics as a science and his placement of spectators at the centre of that science differ radically from most mainstream moral philosophy, both in his time and ours. Nevertheless, despite centuries of relative neglect, his moral theory has recently experienced a renaissance with a flurry of high-quality scholarship across many disciplines.[1]

This paper undertakes a comparative analysis of these two concepts of spectatorship stimulated by possibilities for enriching both economic and philosophical research into ethics. Although the two approaches will be fleshed out in greater detail below, brief and simple descriptions at this point should help motivate and clarify the purpose here. Smith's so-called *impartial spectator* is not a literal third party, indeed not a real person at all, but rather what real people, or *agents*, imagine to be the moral judgements of an impartial and well-informed third party. According to Smith, the repeated social interactions of agents, including as real spectators, produce this internalized moral guide.

The empirical spectator studies on which I will focus, on the other hand, involve real people, specifically ones who reveal the moral judgements of their own presumed impartial spectators concerning matters affecting others. I call these *quasi-spectators*, since their judgements typically only approximate those of the ideal impartial spectator. Their views might be elicited in a variety of ways, but the clearest examples involve treatments in which third parties make decisions affecting the material allocations of other subjects but not of themselves, e.g. Konow (2000), Charness and Rabin (2002), Engelmann and Strobel (2004) and Coffman (2011). Of interest here are also hybrid

[1] Among the many recent works that include treatments of Smith's moral theory, see, for example, Brown (1994, 2009), Witzum (1997), Griswold (1999), Verburg (2000), Ashraf *et al*. (2005), Fricke and Schütt (2005), Haakonssen (2006), Rasmussen (2006), Cockfield *et al*. (2007), Göçmen (2007), Parrish (2007), Raphael (2007), Hanley (2008) and Sen (2009).

quasi-spectator/stakeholder experiments in which third parties incur a fixed cost to influence the earnings of others, e.g. Kahneman *et al*. (1986), Fehr and Fischbacher (2004*b*) and Charness *et al*. (2008). Finally, there are studies that encourage participants to reveal their moral views about real or hypothetical situations, but their views do not materially impact themselves or others. These include many survey studies in the social sciences, especially moral psychology, empirical social choice, and in the emerging field of experimental philosophy. They do not, however, include results from the majority of survey studies (e.g. most public opinion research), which do not explicitly address moral questions or consciously promote impartial reasoning. One concern with surveys is that respondents might be insufficiently incentivized to provide thoughtful judgements given the lack of material consequences. Thus, the discussion here of such studies focuses on some recent investigations that purposely target spectator views and provide evidence strongly suggesting sufficient participant motivation.

To be clear, the intent of this study is not to demonstrate an equivalency between Adam Smith's moral theory and a research programme in behavioural economics. Nor is the title of this paper meant to imply an exhaustive treatment of either, which would be beyond the scope of this paper. Rather, I examine issues at their intersection that appear promising for addressing open questions in both research agendas. I argue that both Smith's analysis and modern quasi-spectator investigations explore moral knowledge and its contents, adopt a kind of scientific method, point to a conflict between self and others in which self-deception is often involved, conceive of impartiality in terms of three properties (absence of stakes, information and a common moral sense), and conclude that the moral sense is often characterized by rules. The following section presents Table 1, to which the reader may refer at various stages in the paper as a rough guide to similarities and differences between the two types of spectators.[2]

The discussion is organized around what I claim is a remarkable but tractable area of overlap in the two approaches in terms of their goals and methodologies (section 1), assumptions about human nature (section 2) and spectator attributes and emphasis on moral rules (section 3). I conclude with a discussion of how insights from Smith's impartial spectator can inspire and help improve empirical spectator studies and economic theory, and how, in turn, empirical spectator findings might inform philosophical ethics and normative economics (section 4).

[2] This table is a stylized description of aspects of the two approaches. Although one might object to parts of this characterization, even in its more extensively developed and nuanced form in the main text of the paper, this concise outline is intended only to help guide the reader through the discussion.

Property or quality	Smithian spectator	Quasi-spectator
Primary focus of inquiry	Epistemic exercise	Content of morality
Methods of analysis	Informal observations and introspection	Formal experiments and formal theories
Methods of reasoning	Primarily inductive; secondarily deductive	
Descriptive/ prescriptive	Emphasis on positive; lesser (or debatable) normative intent	
Moral judge	Ideal (or imagined or supposed) spectator	Real spectators (excludes the object of the judge)
Number of judges	Single	Group
Object of judge	Self (agent)	Other agent or agents
Impartiality construct	Third party: no salient material or non-material stakes	
Information	Well informed but not omniscient	
Moral sense	Moral sentiments are first principles of morality	
Spectator shortcomings	Potential bias exerted by agent on spectator	Residual noise of imperfectly measured moral sense
Self-deception	Spectator reduces self-deceit/self-serving bias	
Generality of morality	Pluralism: multiple but general moral rules	

TABLE 1. Comparison of Smithian spectator and quasi-spectator

1. GOALS AND METHODOLOGIES IN SPECTATOR ANALYSES

1.1 Goals

Smith lays out two questions for ethical inquiry: 'First, wherein does virtue consist? Or what is the tone of temper, and tenor of conduct, which constitutes the excellent and praise-worthy character . . .? And, secondly, by what power or faculty in the mind is it, that this character, whatever it be, is recommended to us?' (*TMS* VII.i.2). That is, the second question inquires into the nature of moral judgement, i.e. it is the *epistemic exercise*, which asks how we recognize what is right. The first question examines the results of that inquiry, i.e. it concerns the *contents of moral knowledge*, including its general properties.

The quote at the start of this paper gives direction as to Smith's answers to both questions. We discern virtue using our 'moral faculties' and guided by our moral sentiments. This grows out of the tradition

of moral sense theory that includes Smith's mentor, Francis Hutcheson, and his friend, David Hume – an approach that proceeds from our sense of approval and disapproval of our actions or those of others. Whereas Hutcheson viewed the moral sense as endowed and likened it to the physical senses, Smith argued that it evolves through a process of repeated interactions with others that are initially motivated by our natural desire for their approval. Indeed, Smith eschewed the term 'moral sense' and referred instead to 'moral sentiments', underscoring both the plurality of such sentiments and their putative evolutionary origins. The development of moral sentiments occupies a prominent place in *TMS* (see, especially, *TMS* III. 2 and VI.1), as it does in much of the secondary literature (e.g. Young 1992; Broadie 2006; Weinstein 2007), and it will resurface later in this paper. Nevertheless, my concern is with the realized impartial spectator, rather than its genesis, for a variety of reasons, including reasons of brevity and the lack of empirical quasi-spectator studies on the formation of moral judgement. Thus, I will employ the term 'moral sense' in this paper as a collective noun that encompasses the moral sentiments while refraining from commitment to other aspects of the positions of Hutcheson and Smith on this subject.

Raphael (2007: 9–10) argues that *TMS* focuses mostly on the epistemic exercise and dedicates much less attention to the content of morality. Others, including Fleischacker (1999), provide extensive commentaries on and analysis of Smith's views on morality itself. The most significant fact about Smith's treatment of the content of morality for the current discussion, though, is his claims about moral rules. That moral judgement can be characterized by rules is both his assumption going into the epistemic exercise as well as his conclusion coming out of it. As the opening quote to this paper conveys, we can infer the general rules of morality when our moral sense encounters a variety of different circumstances, and, although this might not be the deliberate exercise of the agent, it is, in Smith's view, the office of the ethicist.

Modern spectator studies[3] have similarly addressed these two questions, although the greater emphasis has clearly been on investigating the content of moral preferences, including altruism (Harbaugh *et al*. 2007), distributive justice (Dickinson and Tiefenthaler 2002; Chavanne *et al*. 2010*a*; Becchetti *et al*. 2012), fair rewards from risk-taking (Huesch and Brady 2010; Cappelen *et al*. forthcoming), fair distribution of losses from risk-taking (Cappelen *et al*. 2011), reciprocity (Charness *et al*. 2008; Croson and Konow 2009; Utikal and Fischbacher 2010), and the effects of moral bias (Traub *et al*. 2005; Croson and Konow 2009; Chavanne *et al*. 2010*b*). As with the larger empirical literature on social preferences that

[3] When it is clear from the context that I am referring to empirical spectator research, I will usually shorten 'quasi-spectator' to 'spectator'.

includes stakeholder studies, this research has usually been conducted under the presumption that general forces are at work and with the goal of contributing to models of those forces (Konow 2000; Charness and Rabin 2002). Thus, this parallels Smith's assumptions and conclusions about moral rules. Although empirical spectator studies have maintained their distance from epistemic and metaethical inquiry, they have also addressed the role of information on spectator judgements (Konow 2009*a*, 2009*b*), possible spectator bias (e.g. Aguiar *et al*. 2010; Chavanne *et al*. 2010*b*), and differences between spectatorship and other concepts of impartiality such as Rawls's (1971) veil of ignorance (Herne and Mård 2008; Aguiar *et al*. 2010). Many of these studies explicitly refer to Smith.

These quasi-spectator studies are social science projects aimed at describing behaviour and its underlying social preferences. Smith, on the other hand, places his objective with *TMS* squarely in the domain of philosophy, a point that Part VII underscores. These differences would usually result in two quite distinct research agendas, viz. with respect to whether the analysis is descriptive or prescriptive, and whether the reasoning is mostly inductive or deductive. Nevertheless, given how they conceptualize and investigate their subject matters, the two approaches are both more distant from much of the mainstream in their respective disciplines and more proximate to one another than one might expect, as discussed in the following sub-section.

1.2 Methods

To understand Smith's take on these questions, we must realize that he adopts an *empiricist*, rather than rationalist, epistemology. That is, he traces the source of moral knowledge to our moral sentiments rather than reason: 'But though reason is undoubtedly the source of the general rules of morality ... it is altogether unintelligible to suppose that the first perceptions of right and wrong can be derived from reason ... These first perceptions, as well as all other experiments upon which any general rules are founded, cannot be the object of reason, but of immediate sense and feeling ... It is by finding in a vast variety of instances that one tenor of conduct constantly pleases in a certain manner, and that another as constantly displeases the mind, that we form the general rules of morality' (VII.iii.2.7). The inductive approach Smith advocates, and even the terminology ('experiments') he employs, could characterize the modern social science research that seeks to identify social preferences empirically and to infer their rules. Smith's impartial spectator embodies that moral sense and represents the preferences targeted by quasi-spectator studies.

Although these two approaches proceed from similar assumptions and share a common disposition to scientific reasoning and language,

there remain important differences in methods. Redman (1993) writes that Smith borrowed from the Newtonian parlance of the day but that the terms 'experiment' and 'observation' were given different meanings by Smith and his Scottish contemporaries in philosophy. Whereas modern science and quasi-spectator research both involve formal methods of observation and analysis, Smith's method was based on introspection and informal psychology, viz. personal and written observations of others. Whereas Smith stresses the grounding of the general moral sense on individual experience, the quasi-spectator agenda is focused on collecting individual experiences of the general moral sense. Smith engages in thought experiments, whereas quasi-spectator researchers employ formal experiments.

As scientific endeavours, both approaches lean heavily on induction but not to the exclusion of deduction. Smith refers to the need both to use reason to infer the general rules of morality but also to apply those lessons in particular cases: 'From reason, therefore, we are very properly said to derive all those general maxims and ideas. It is by these, however, that we regulate the greater part of our moral judgements, which would be extremely uncertain and precarious if they depended altogether upon what is liable to so many variations as immediate sentiment and feeling, which the different states of health and humour are capable of altering so essentially' (VII.iii.2.6). Modern spectator studies have been utilized to infer theories of moral preferences, as in Charness and Rabin (2002) and Konow (2000). With respect to deduction, Smith describes in the passage above the use of general rules to regulate behaviour in specific instances, whereas spectator studies often proceed from general to specific for the purpose of testing existing theories in particular contexts, as in Gaertner *et al*. (2001).

The distinction between induction and deduction in these research agendas is also closely connected to the descriptive and prescriptive intentions of their analyses. Among Smith scholars, the more controversial question of the two concerns whether *TMS* should be understood in prescriptive terms. On the one hand, *TMS* reads like a psychological and sociological treatise on the evolution and practice of social norms. On the other hand, normative words like 'should' and 'ought' occur with great frequency, and the agent's efforts to comply with his moral sentiments and the demands of the spectator are often laid out in the first-person plural, as Smith exhorts the reader to right conduct and character.[4] Raphael (2007: 133–135) sees Smith's usage in psychological,

[4] For example, regarding revenge, he writes 'There is no passion, of which the human mind is capable, concerning whose justness we ought to be so doubtful, concerning whose indulgence we ought so carefully to consult our natural sense of propriety, or so diligently to consider what will be the sentiments of the cool and impartial spectator' (I.ii.3.8).

rather than philosophical, terms. Haakonssen (2006) states that 'morality was, in Smith's eyes, to be approached as a matter of fact', although he points to an indirect normative significance in Smith. Most commentators see some of both purposes in Smith, and a common view is that *TMS* has normative purpose grounded in a descriptive construct, viz. the impartial spectator. Firth argues that Smith presents the impartial spectator not only as a device to explain behaviour but also as an ideal to which people should aspire – indeed, one that is practically attainable for a segment of the population (2007: 118). Fleischacker (1991) writes that 'moral philosophers, he believes, should contribute to moral practice' (p. 255). Griswold (1999) points to the tension between agent and spectator perspectives in contrasting everyday behaviour with moral excellence. He suggests that Smith sees moral theory as a means we *ought* to pursue in order to promote moral practice, pointing, among other things, to Smith's protreptic writing style. Along similar lines, Weinstein (2007) contends that Smith views education chiefly as serving moral growth. Campbell (1971) concludes that the main emphasis in *TMS* is on explanation but that Smith's work should be viewed in the context of his day, when the contrast between science and philosophy was considerably less sharp.

Such close association of descriptive and prescriptive analysis raises the spectre of the 'is–ought fallacy,' which can be traced to Hume's is–ought problem and Moore's naturalistic fallacy. This is often encapsulated in the saying '*ought* cannot be derived from *is*'. That is, one cannot argue from premises that contain only descriptive statements, e.g. most people support capital punishment, to normative conclusions, e.g. capital punishment should be legal. There is some controversy in the philosophical literature about whether this is a fallacy, but the treatment of descriptive and prescriptive statements as equivalent, at a minimum, marks a tautology.

Although quasi-spectator research, by definition, examines the actual values of real people, much of it is motivated by the belief that empirical evidence on impartial moral views is germane to reflection on prescriptive theories. Many authors suggest their results are relevant to reflection on philosophical ethics (Schokkaert *et al*. 2003; Herne and Mård 2008; Amiel *et al*. 2009), normative economics (Konow 2003, 2009*a*, 2009*b*; Gächter and Riedl 2006), and policy (Traub *et al*. 2005; Chavanne *et al*. 2010*a*). The putative normative relevance of these empirical findings is not based on observation of just any actions or views, indeed not even on specifically 'moral' views, but rather on moral views elicited in a particular way. The implicit claim behind this professed relevance might be stated as follows: 'Although *ought* cannot be derived from *is*, *ought* can be derived from a *subset of what is*'. Such a statement is bound to stir controversy, and I have not seen it formulated in this way in the quasi-spectator literature.

But the view that the validity of normative claims depends in some (as yet incompletely specified) manner on their ability to be reconciled with actual impartial moral judgements has parallels to Smith's approach. The spectator serves as the moral measuring rod in passages with value-laden terminology (e.g. see the use of 'ought' in VI.ii.I.22) but is also described as a creation of the agent, who imagines the moral judgements of a well-informed third party. Thus, one interpretation of both agendas is that normatively valid claims are based on a subset of actual views, viz. the moral judgements of informed and impartial third parties, literal ones in the case of quasi-spectators and an imagined one in the case of Smith.

2. HUMAN NATURE: AGENT AND SPECTATOR

This section explores parallels between Smith's vision of human nature and recent approaches in behavioural economics and expands discussions of the two spectator models.[5]

2.1 Split personality

Smith added the following sub-title to *TMS* beginning with edition 4: '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'. This sub-title both underscores the importance of the epistemic exercise to his enterprise and portends the dual nature of people as observers and participants that he develops. In Smith's view, there are two parts to a person, viz., the agent and the spectator, who judges the agent as if from a distance: 'When I endeavour to examine my own conduct ... it is evident that ... I divide myself, as it were, into two persons; and that I, the examiner and the judge, represent a different character from that other I, the person whose conduct is examined into and judged of. The first is the spectator ... The second is the agent' (III.1.6). This is one of many cases of Smith anticipating much later work in the social sciences. In psychology, this resembles the impulsive System 1 and deliberate System 2 accounts of dual process theory (e.g. Evans 2008). In economics, dual-self theories employ far-sighted planners and myopic agents (Thaler and Shefrin 1981; Fudenberg and Levine 2006).

[5] *TMS* is teeming with detailed, rich and trenchant statements about human nature, and the treatment here undeniably neglects alternate renderings that address Smith's theory in more complicated ways. If the presentation here seems too narrow, though, this reflects not only a desire for parsimony but also the aforementioned goal of tractability by remaining close to the intersection of the two approaches.

Dual-self models have addressed mostly problems of intertemporal choice and self-control rather than specifically moral questions. In contrast, most efforts to formalize moral conflict, in both quasi-spectator studies and the more general behavioural economics literature, have been reductionist and included separate terms representing material self-interest and moral preferences within a single utility function, e.g. Levine (1998), Fehr and Schmidt (1999), Bolton and Ockenfels (2000), Konow (2000), Charness and Rabin (2002), Dufwenberg and Kirchsteiger (2004), Engelmann and Strobel (2004) and Falk and Fischbacher (2006). Although moral preferences are variously specified in these models as altruism, reciprocal altruism, equality, equity, maximin, efficiency, or some subset thereof, they incorporate tension between material utility and an internalized motive that includes the interests of others.

However formally modelled in recent work, these internalized and opposing goals, i.e. between one's own interests and those of others, are central in *TMS*; indeed, Smith places them in the opening sentence: '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' (I.1.I.1). On the one hand, individuals are referred to the sympathetic impartial spectator, to the 'tribunal of their own consciences' and the 'great judge and arbiter of their conduct' (III.2.32). On the other hand, self-interest, or *self-love* as Smith expresses it, plays a famously central role in his thinking. Self-love is not to be confused with *prudence*, or the proper care of one's own well-being, including one's health and reputation, which Smith considers a virtue (VI.1). Rather, self-love represents a failure to sympathize appropriately and is associated with the pursuit of 'wealth and greatness (which) are mere trinkets of frivolous utility' (IV.I.8).

2.2 Three properties of spectators

Smith's impartial spectator is not a real person or persons but rather a model conjured in the imagination of the agent, and he frequently refers to the 'imagined' or 'supposed' spectator. This spectator takes neither the position of the agent nor that of a real bystander, but rather the view the agent imagines an impartial observer possesses. Quasi-spectators, by contrast, are real third parties, i.e. individuals lacking material stakes in the matter at hand, who are prompted to volunteer the moral judgements of their own impartial spectators.

Smith fleshes out in some detail the origin of the imagined spectator: he is the product of society (e.g. *TMS* VI.1). People participate in social interactions and, as real spectators (i.e. third parties), judge the behaviour

of others. This leads to a moral sense, eventually manifested as an imagined third party and experienced chiefly in affective terms, which, when turned on the agent himself, becomes conscience. Social approval provides the motivation for, but not the ultimate goal of, moral learning: 'But this desire of the approbation, and the aversion to the disapprobation of his brethren, would not alone have rendered him fit for that society for which he was made. 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; or of being what he himself approves of in other men. The first desire could only have made him wish to appear to be fit for society. The second was necessary in order to render him anxious to be really fit' (III.2.6). Indeed, he describes in eds. 2–5 the process of moral learning in considerable detail, by which we eventually learn to balance or stand above the sometimes conflicting desires of those with whom we deal (III.2). The resulting impartial spectator, as Sugden (2002) writes, 'represents, in idealized form, the *correspondence* of sentiments that is induced by social interaction'.

I will focus on three properties that the impartial spectator and the quasi-spectator have in common: impartiality, information and the moral sense. This is not meant as an exhaustive list of the properties of either, but I believe these categories are the most useful and important ones for this particular comparison. In this section, I describe these properties in the ideal Smithian impartial spectator and then proceed, in following sections, to discuss various practical limitations, both those raised by Smith and those emerging from quasi-spectator research.

First, the spectator is impartial, a modifier that also represents a Smithian addition to the language of spectator theory. That is, as an ideal, the impartial spectator is a disinterested (or *detached*) third party, who has no stakes in the situation or parties being evaluated. This, of course, serves to avoid moral judgements being tainted by self-interest. But there is a potential problem, as Griswold (2006) points out, severing the spectator from self-interest. The spectator is a creation of sympathetic imagination, but individuals are motivated to respond to sympathy, which itself seems self-centred. Quasi-spectator studies and modern behavioural models address this question by drawing the line between self-interest and social preference based on whether the motive or action is directed toward one's own material interests or reflects more general considerations that potentially incorporate the interests of others, respectively. I believe this a productive distinction, in general, for both empirical research and reading Smith: let self-interest refer to motives based on extrinsic reasons, including but not limited to material rewards, e.g. also seeking the approbation, or avoiding the disapprobation, of others, whereas any intrinsic reward or punishment that is derived from or motivated by acting in the interests of others, such as sympathy, is not

characterized as self-interested. As sympathy motivates agents toward right conduct in Smith, casting sympathy as self-interest risks making the latter term tautological. This proposed distinction, by contrast, seems consistent with Smith's treatment of sympathy as praiseworthy and of self-interest as being a different and potentially opposed motive.[6]

Second, impartiality must be coupled with the necessary information conditions. Smith states on several occasions that the morally relevant judgements are those of the 'impartial and well-informed spectator'. Although, at various points, he acknowledges the challenges to this goal, these conditions are not an unattainable abstraction but rather a state that real people can sometimes achieve (III.3.35–37). Importantly, the spectator accesses his own life experiences: 'The man who is conscious to himself that he has exactly observed those measures of conduct which experience informs him are generally agreeable, reflects with satisfaction on the propriety of his own behavior'. Nevertheless, his conduct is not motivated by a desire for social approbation or by false consciousness, since 'he views it in the light in which the impartial spectator would view it . . . and though mankind should never be acquainted with what he has done, he regards himself, not so much according to the light in which they actually regard him, as according to that in which they would regard him if they were better informed' (III.2.5). Moreover, the spectator who reasons with these facts is 'cool' and 'intelligent' (e.g. I.ii.3.8 and VI.3.27). These passages help underscore the importance of these properties of the imagined spectator, which are themes in *TMS*: the spectator is a third party, well informed of the relevant particulars, who processes this information rationally with respect to internalized values.

Third, an informed third party is of no relevance to either Smith's or the modern spectator programme unless, of course, the respective spectator possesses a moral sense. This is related, but not identical, to sympathy in Smith, an often controversial topic among his commentators. Various interpretations distinguish the affective versus cognitive qualities of sympathy and the associated relationship between agent and spectator. Sugden avoids the term sympathy, which he associates with the economic model of altruism, and, instead, refers to 'fellow-feeling', which is a consciousness of another's feelings, where that consciousness itself brings pleasant or unpleasant feelings (2002, 71). Similarly, for Griswold sympathy is a fellow-feeling (of the affective kind) that is not necessarily limited to morals (2006: 25). Raphael (2007) interprets sympathy as the feeling that results from the convergence of spectator and agent

[6] I thank a referee for prompting me to clarify the relationship between self-interest and sympathy and, thereby, to add this discussion to the paper. Of course, this distinction does not rule out the possibility that these motives can be complementary as well as opposed: depending on the context, a selfish act can be morally right or morally wrong.

sentiments. Weinstein describes it as fellow-feeling with any passion of others and as a cognitive process, which subsequently arouses emotion in the spectator (2006: 83; 2007: 132). Others see it in less affective terms. Göçmen (2007) considers sympathy the ability to imagine ourselves in situations of others and distinguishes first-order sympathy, whereby we imagine ourselves in the situation of another, and second-order sympathy, where we imagine ourselves as someone observing us. Forman-Barzilai (2005) makes an even more dramatic break from common interpretations of sympathy as an emotion or a virtue and argues Smithian sympathy is a social practice, i.e. an activity that produces morality in shared physical, affective and historical spaces.

There is merit in each of these readings, I think, but my concern in relating Smith to quasi-spectator research is with the moral sense itself, or, more specifically, with the set of moral judgements and rules implied by the spectator's moral sentiments, and the supportive role of sympathy. To that end, various versions of sympathy will do. I will focus on two functions of sympathy, similar to the distinction Rawls (2000) makes in reference to Hume's spectator. First, sympathy has an *epistemic* role that is of relevance to the spectator: it enlarges his awareness of relevant facts by enabling him to factor in the experiences and feelings of others in coming to moral judgements. In addition, sympathy has a *motivational* function that pertains to the agent and helps him to put aside, or at least to moderate, his own interests relative to those of others and to align his conduct more closely with the judgement of the spectator.

A theme that runs throughout much of *TMS* is the role of sympathy in enabling the moral sense of the spectator and motivating moral action by the agent (e.g. III.3.3). The agent secures the sympathy he desires from others 'by lowering his passion to that pitch, in which spectators are capable of going along with him' (I.i.4.7). And, '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' (I.i.4.8). Perhaps the clearest indication of Smith's belief in the virtues of his model for epistemic purposes emerges in his comparative analysis of moral theories in Part VII. For instance, in a passage added to the 7th edition, he writes of alternative theories that 'None of these systems either give, or even pretend to give, any precise or distinct measure by which this fitness or propriety of affection can be ascertained or judged of. That precise and distinct measure can be found nowhere but in the sympathetic feelings of the impartial and well-informed spectator' (VII.ii.1.49). This assertion not only reinforces his many statements elsewhere about the critical roles of impartiality, information and sympathy but also elevates his claims to a higher level: his theory, unlike others, provides the means to identify right and to distinguish it from wrong.

3. IDEAL VS. REAL SPECTATORS

This section explores Smith's view of human nature in greater detail, examining moral flaws of both agents and spectators. I argue that this exercise produces lessons for empirical research and leads to conclusions about the content of morality.

3.1 Nobody is perfect, not even spectators

Smith frequently refers to the impartial spectator as the 'ideal man within the breast' (e.g. III.3.29), but there are indications of a fallible judge, indicating he uses the modifier *ideal* in the sense of *imagined* rather than *perfect*.[7] Although the spectator encapsulates the moral sense, Smith treats in some detail the 'irregularities of sentiments', and these can cause 'even the impartial spectator (to) feel some indulgence for what may be regarded as the unjust' feelings of agents (II.iii). Smith explicitly notes variability and partiality in the spectator's judgements, such as in the spectator's interpretation of the motives of another person: 'the imagination of the spectator throws upon it either the one colour or the other, according either to his habits of thinking, or to the favour or dislike which he may bear to the person whose conduct he is considering' (III.2.26). Moreover, as evolving moral agents, we initially seek the approbation of all and try to incorporate the sometimes conflicting interests of stakeholders. But 'we find that by pleasing one man, we almost certainly disoblige another' and, therefore, 'soon learn to set up in our own minds a judge between ourselves and those we live with' (III.1, eds. 3–5). Although this is not stated in so many words in *TMS*, the vagaries of spectator impartiality might emerge, therefore, not only from the 'self-love' and shortcomings of the agent who creates him, but rather be inherent to the spectator's balancing act.

Most quasi-spectator studies employ various measures to implement impartiality, e.g. stakeholders and spectators are anonymous, and spectators receive fixed fees unrelated to those of stakeholders. The first such study was the 'dictator game' of Konow (2000). In the standard version of this game, which I call here the 'stakeholder treatment', subjects are anonymously paired, and one subject (the so-called 'dictator') is given a sum of real money to divide between himself and his counterpart (the 'recipient'). In another new 'spectator treatment', a third party (or spectator), divided a sum of money between two anonymous subjects. The dictators in the standard stakeholder treatment demonstrated a bias,

[7] Although some, e.g. Campbell (1971) and Rawls (2000), have read Smith as presenting an ideal observer theory, i.e. as claiming that ethical judgements are those of an omniscient completely impartial observer, the large majority of Smith scholars seems to disagree with this assessment, e.g. see Broadie (2006).

taking for themselves, on average, more than third parties in the spectator treatment gave similarly situated subjects in their treatment.[8]

Most other spectator experiments have adopted a similar method, although I am aware of two studies that deliberately implement potentially partial, as opposed to impartial, spectators with mixed results. In a three-player dictator game, Chavanne *et al*. (2010*b*) vary whether initial stakes are earned or endowed, after which one subject, the spectator, can redistribute earnings between the other two. They find that the spectator's redistribution is unaffected by whether or not he shares with one of the recipients the same experience with respect to the initial stakes, i.e. whether or not initial stakes were earned or endowed. Aguiar *et al*. (2010) come to different conclusions. They report a two-stage dictator game: initial stakes are endowed unequally between three subjects, after which one subject in each triple is paid a fixed fee to allocate as spectator a second sum of money between the other two. They find that spectators whose initial endowments equal those of one of the two stakeholders favour such stakeholders compared with spectators whose initial endowments did not equal those of any stakeholders. Thus, introduction of partial considerations might disturb spectator impartiality, although impartiality does not, thereby, seem to be systematically affected or especially labile.

Smith also identifies shortcomings of spectator sympathy, although in his case not that of sympathizing unequally with two parties but rather of sympathizing insufficiently with one: 'the emotions of the spectator will still be very apt to fall short of the violence of what is felt by the sufferer', since that 'imaginary change of situation, upon which their sympathy is founded, is but momentary' (I.i.4.7). There is an asymmetry in this respect, since, observing a companion, the spectator 'must find it much more difficult to sympathize entirely, and keep perfect time, with his sorrow, than thoroughly to enter into his joy' (I.iii.1.8).[9]

There are relatively few experimental studies of feelings and morality, still fewer involving quasi-spectators. The closest, to my knowledge, is an experiment by Harbaugh *et al*. (2007) using dictators who can allocate money to a charity that serves the poor. Dictators face a series of payoff combinations between themselves and the charities, e.g. respectively (100,0), (100,45), (55,0), (55,45), etc. Both fMRI measures of neural activity

[8] Another difference between this and previous dictator experiments was that subjects first performed a real effort task that generated the joint earnings to be divided between pairs. Spectators in this study then allocated the joint earnings of their assigned pair. Most spectators chose to divide earnings in proportion to the individual productivities of the subjects, which contrasts with the equal splits so often found in other dictator experiments.

[9] In addition, as Nussbaum (1990) points out, there are certain feelings of others, such as bodily desires and romantic love, with which Smith argues we not only cannot, but also should not, sympathize as spectators.

and self-reported subjective satisfaction scores indicate that, holding dictator earnings constant (similar to spectators), they subsequently feel better, when the charity receives more. Using self-reported measures of feelings, the dictator study of Konow (2010) finds a similar, and even stronger, result: dictators feel better, when they transfer more to charities serving the needy, even when dictator earnings are not held constant but rather decrease dollar for dollar with transfers.[10] Greene *et al*. (2001) examine whether feelings affect moral judgement using hypothetical scenarios, including emotionally charged moral dilemmas as well as other non-emotional scenarios. Their fMRI scans confirm that the moral dilemmas cause greater emotional activation in subjects and further that this activation predicts choices in the dilemmas. Thus, these studies find evidence consistent with a relationship between feelings and moral judgements or actions.

Regarding the spectator property of information, as previously noted, sympathy serves a complementary epistemic role: '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 ... and strive to render as perfect as possible, that imaginary change of situation upon which his sympathy is founded' (I.i.4.6). Smith repeatedly makes clear that morally relevant judgement proceeds not from the incomplete and faulty knowledge we often possess but rather from an advantaged informational position that provides access to relevant particulars (e.g. III.2.5–9). Although complete information is the ideal, Smith refers to the 'well-informed', never the 'perfectly' or 'completely' informed, spectator and marks the limits of this goal with comments like 'as much as he can' and 'as perfect as possible'. On several occasions, he stresses the importance of being informed of the causes of passions with which we might sympathize while noting the state of imperfect information in which we often find ourselves (e.g. I.i.1.9, I.ii.3.5).

Contrary to these claims about the benefits of acquiring more information, there are both theoretical arguments and empirical evidence for its deleterious effects. Additional information might conceivably complicate moral reasoning and encourage divergent views, undermining attempts for consensus. Moreover, some research on stakeholders, such as

[10] Nevertheless, compared with a control group of subjects who are given an endowment and no opportunity to transfer, dictators transferring to charities do not feel significantly better, and the less generous subjects in that group actually feel significantly worse compared with the control. In addition, the relationship between generosity and feelings is reversed when student dictators transfer, not to charities, but to fellow students: more generous dictators feel worse than less generous ones and than the control group, and, in fact, average generosity is significantly lower toward students than charities. The theory presented in that study reconciles these patterns based on preferences that depend on context-dependent norms.

that reported by Babcock and Loewenstein (1997), finds that information feeds biases and increases disputes. Two questionnaire studies explore the effects of varying information in quasi-spectator research. A hypothesis in Konow (2009*a*) is that relevant information permits third parties to reason more accurately and, therefore, reach consensus. In fact, the moral judgements of better-informed respondents regarding eight different scenarios exhibit lower variance that cannot be attributed to alternative explanations, such as focal points. This analysis is expanded in Konow (2009*b*) to include different combinations of both relevant and irrelevant information and different types of stakeholders as well as spectators in six different scenarios. Irrelevant information does not contribute to moral consensus, but relevant information creates consensus among both spectators and stakeholders and even significantly reduces stakeholder bias. Thus, quasi-spectator research corroborates the benefits of relevant information for spectators and clarifies the need to distinguish different types of information and the different roles of spectators and stakeholders.

In light of variable impartiality, incomplete sympathy and imperfections of information, one seems justified in speaking of spectators rather than *the* spectator. More often, though, Smith employs the singular, perhaps, in part, because of his focus on the reflexive and personal use of spectator perspective by the agent. Nevertheless, the real conditions Smith frequently illuminates are characterized by various imperfections that colour individual moral judgement. Campbell refers to 'the averaging out of differences between the reactions of spectators', which produces a 'sort of moral consensus' (1971: 138). This characterization of spectatorship is implicit in the approach of quasi-spectator research and is consistent with the findings of the quasi-spectator studies that shed light on such claims, which I will now summarize.

First, minimal evidence of impartiality is that quasi-spectators' judgements often differ from those of stakeholders given the bias of the latter. This has been confirmed in distribution experiments, e.g. Engelmann and Strobel (2004), and dictator games, e.g. the aforementioned Konow (2000) study that identified bias in stakeholders relative to spectators. Stronger evidence of the impartiality of quasi-spectators comes from a vignette study (Konow, 2009*b*) in which stakeholders have opposing interests. Since each is biased in his own (opposite) direction and away from the supposed impartial choice, the mean spectator choices should be intermediate to the mean choices of opposing stakeholders. That further prediction is confirmed in this study.

The above results are about average behaviour. Another type of evidence comes from the dispersion in views. A potential concern with quasi-spectators is that they are insufficiently motivated to acquire or process relevant information about others and instead choose randomly or capriciously. On the other hand, stakeholders might exhibit higher

variance in their judgements than spectators for at least two reasons: first, their interests are often opposed and, second, individual agents differ in the weight they attach to self-interest versus moral norms. Moreover, quasi-spectators should place greater value on the morally right choice; although their views might not align perfectly for various reasons, including those discussed above, there should be a much higher level of consensus in their choices. This latter conjecture is confirmed in studies of distributive preferences (Konow *et al*. 2009) and reciprocal preferences (Croson and Konow 2009), which find significantly lower variance in spectator than stakeholder decisions.

Several other findings bolster confidence in the utility of quasi-spectator research. Stakeholders have been shown to act on the same moral values as quasi-spectators, even when the rules are more complicated, such as proportional rules of equity in Konow (2000), and even when multiple rules compete, as in Engelmann and Strobel (2004) and Cappelen *et al*. (forthcoming). The latter study also reveals that, controlling for individual differences in the weight attached by stakeholders to self-interest, the decisions of stakeholders and spectators contain very similar levels of unexplained variance or 'noise'. The moral preferences of third parties are sufficiently strong that they are willing to incur material costs to enforce them on others, e.g. Fehr and Fischbacher (2004*a*, 2004*b*), Kahneman *et al*. (1986) and Turillo *et al*. (2002), and to incur psychic costs to contemplate them in complex but hypothetical scenarios, e.g. Huesch and Brady (2010) and Konow (2009*a*, 2009*b*).

3.2 The tangled web of self-deception

In the 2009 film 'The Invention of Lying', a world is depicted in which people never lie to themselves or others. Were such honesty combined with unfettered access to morally relevant information, it is unclear what need there would be for impartial spectators, at least for their epistemic function. People sometimes behave in self-interested ways (indeed, often brutally so, in the aforementioned film), but this is not due to a failure to understand what is right, but rather to an unwillingness to act on those morals because of unadulterated self-interest. An explanation for the need for an epistemic spectator role is the possibility of self-deception, specifically, the ability to alter one's beliefs about what is right in the direction of one's own selfish interests, which Smith termed self-deceit and is, in modern scholarship, often called a *self-serving bias*. Indeed, Fleischacker (2011) sees self-deception as critical to understanding not only the impartial spectator's role but Smith's moral philosophy as a whole. Self-deception has been an important topic in philosophy but it has also been prominent in modern social science research beginning with cognitive dissonance theory (Festinger 1957).

Smith writes that 'self-deceit, this fatal weakness of mankind, is the source of half the disorders of human life' (III.4.6). Modern research suggests Smith's estimate might be conservative: even under sterile laboratory conditions, which are least conducive to self-deception, almost two-thirds of unfairness has been traced to such a bias (Konow 2000). The experiment of Di Tella and Pérez-Truglia (2010) also finds a large self-serving bias, specifically, subjects in that study act selfishly towards counterparts by distorting their beliefs about how selfishly the counterparts will behave towards them. Babcock and Loewenstein (1997) describe experimental and field studies showing that the self-serving bias significantly impacts bargaining behaviour, impeding agreements and promoting impasse, such as with contract negotiations and in civil litigation. They also report that this bias is very tenacious, as demonstrated by various experimental attempts to dislodge it.

Is Smith's imagined spectator free of such self-deception? There are both philosophical and empirical grounds for hesitating to answer in the affirmative. This model, as we have discussed, involves looking inward not to one's sentiments but to one's sentiments if one were an imagined other person. Note that this is a kind of second-order introspection, but introspection, just the same. Schwitzgebel (2008) makes an important and compelling argument, in my view, that introspection is generally highly unreliable in at least two ways: it sometimes yields no result and at other times the wrong result. In fact, there is evidence from social science experiments on moral decision-making of the latter: people have systematically biased beliefs about what is right. Since the impartial spectator is conjured by real agents, these considerations should leave us less than sanguine about the objectivity of the derived moral judgements.

Smith claims the impartial spectator is better informed than the agent, since the former considers the interests and perspectives of others. In forming the image, however, the spectator accesses the life experiences of the agent. This seems an improvement over theories of impartiality that rely on constraints on information, such as the Rawlsian veil of ignorance, that require withholding information that could create bias but that might, nonetheless, be necessary to render accurate moral judgements. Nevertheless, the experiences of the agent also limit the imagination of the spectator, as Weinstein (2006) points out. Recondite information might, at a minimum, be a source of error in the spectator's reasoning. Of particular concern, research on self-deception indicates that self-serving biases arise chiefly through the biased collection and recollection of information, e.g. Dunning *et al*. (1989) and Thompson and Loewenstein (1992). Thus, agents, the progenitors of the spectator, might filter the information provided to the latter and affect the imagined correspondence in a biased manner.

Indeed, in ed. 1 of *TMS*, Smith expresses precisely these kinds of doubts and impugns even the spectator for self-deceit: after likening the spectator to a looking glass, he continues 'unfortunately this moral looking-glass is not always a very good one. Common looking-glasses, it is said, are extremely deceitful, and . . . 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' (III.1.5n). This passage was dropped, however, from later editions, which, combined with his treatment of the spectator and of self-deceit as a whole, suggests to me that he viewed the impartial spectator, on balance, as a positive force against self-deception.

In addition to the impartial spectator, Smith mentions the real spectator, i.e. a genuine third party who observes and judges the conduct of others.[11] To be sure, the real spectator is not as prominent a feature in *TMS* as the imagined spectator, and Smith uses this exact term only three times in that book, all in Part III. Nevertheless, he often refers to real spectators in other words (e.g. III.2.15–24 and III.3.22–24), sometimes relating them to the impartial spectator. For example, when writing on gratitude and resentment, he states that feelings 'seem proper and are approved of, when the heart of every impartial spectator entirely sympathizes with them, when every indifferent by-stander entirely enters into, and goes along with them' (II.i.1.7).

Real spectators are explicitly mentioned jointly with the imagined spectator as the voices that are unwisely ignored by a person who lacks self-command (III.3.26). The most detailed discussion of the real spectator, however, is a passage in which Smith acknowledges the kinds of limitations of the imagined spectator discussed above: 'In solitude, we are apt to feel too strongly whatever relates to ourselves: we are apt to over-rate the good offices we may have done, and the injuries we may have suffered: we are apt to be too much elated by our own good, and too much dejected by our own bad fortune. The conversation of a friend brings us to a better, that of a stranger to a still better temper. The man within the breast, the abstract and ideal spectator of our sentiments and conduct, requires often to be awakened and put in mind of his duty, by the presence of the real spectator: and it is always from that spectator . . . that we are likely to learn the most complete lesson of self-command' (III.3.38). Here Smith addresses the effects of the self-serving bias, including the biased processing of information, and the role of the real spectator in correcting

[11] See Brown (1994) for a thoroughgoing discussion of different voices in Smith's writings. In addition, this work presents a novel take on the so-called 'Adam Smith problem', i.e. the ostensible conflict between view of morality in *TMS* and of self-interest in Smith's *The Wealth of Nations* (see also Brown 2009).

that bias and prompting the agent to greater objectivity and purer motives. Moreover, Smith writes that the less personal the relationship between agent and real spectator, the more effective the intervention of the latter. Indeed, Schram and Charness (2011) show experimentally that even the advice of anonymous spectators can influence agent behaviour.

Some commentary on Smith has cast the real spectator in a negative light (e.g. Raphael, 2007). This can be traced to a narrow reading of the first explicit mention of the real spectator in *TMS*, where Smith was attempting in later editions to address a criticism of his theory as presented in the first edition. Specifically, a potential problem of his impartial spectator is this: if the conscience that guides the imagined spectator is a product of society, how is it that the conscience sometimes opposes the judgement of society, as we can observe it sometimes does? Smith's response involves distinguishing the imagined spectator from the morally superior 'all-seeing Judge of the world, whose eye can never be deceived, and whose judgments can never be perverted' (III.2.33).[12] Smith writes at great length in *TMS* on the natural harmony between Nature and the moral values fostered through socialization. In countering this challenge to his theory, however, he devotes two paragraphs to introducing the possibility of occasional tension between these forces. In such cases, the imagined spectator seems to be torn between the real spectators, whose views normally concur with his own, and a higher authority: 'The supposed impartial spectator of our conduct seems to give his opinion in our favour with fear and hesitation; when that of all the real spectators, when that of all those with whose eyes and from whose station he endeavours to consider it, is unanimously and violently against us' (III.2.32). Thus, this problem arises not because of some deficiency specific to real spectators, but rather because of Smith's attempt to explain how both real and even imagined spectators might at times be at odds with the higher moral authority. Indeed, all three passages explicitly mentioning the real spectator imply that the views of imagined and real spectators are typically aligned.

Both real and imagined spectators sometimes err, but Smith views both types of spectators as usually accurate and mutually consistent resources whose moral judgements dominate those of the agent. The largest share of *TMS* treats individual judgement and behaviour, and, therefore, the frequent reflexive use of the imagined spectator might be viewed, at least in part, as a practical matter: that individuals

[12] What Smith considers this higher authority to be (for example, whether he is referring to God) is a contentious point among Smith scholars (e.g. see Hill 2001; Otteson 2002; Clarke 2007). I will not enter into this debate, as this is a wide-ranging question beyond the scope of this paper and one that arises here only in the context of our chief concern with comparing Smith's imagined and real spectators.

would constantly consult literal third parties regarding every action with potential moral implications is blatantly infeasible. The impartial spectator, on the other hand, is a guide that not only can be summoned at almost any moment but also one that does not always require conscious deliberation, when it influences the agent through conscience.

Consider, though, a different purpose for spectators, viz. that of distilling empirically the moral sense and its rules. This is the epistemic exercise in moral philosophy to which Smith refers and is of interest to many modern social scientists as an aid to both descriptive analysis and policy design. What kind of spectator might we want in light of the limitations raised above? I will argue that an answer to this question, which builds on materials Smith provides, is the quasi-spectator approach. The impartial spectator, which the agent imagines for application to his own behaviour, is, as discussed, subject to self-serving biases. Consider, instead, a real spectator modelled on the three aforementioned properties of the imagined spectator but who affects the circumstances of others but not himself. This person has no salient personal (e.g. material or reputational) stake, so that any tendency for self-interest to insinuate itself into his judgement (e.g. through projection of his interests on others) and to cause a self-serving bias is sharply attenuated. In addition, this person has liberal access to information that might be relevant to judging conduct and character. In the absence of personal stakes, the process of sorting relevant information from irrelevant information should not be sullied. Sympathy motivates this spectator's desire to commit resources to the acquisition of information, including about the interests and feelings of affected parties, and to expend mental effort processing the facts with respect to the imagined spectator embedded in him. This person is neither omniscient, due to cognitive limitations and the practical impossibility of acquiring of all relevant information, nor infallible, given potential residual effects of individual interests and insufficient sympathy. Nevertheless, the judgement of this spectator should be superior to that of the imagined one the agent applies to himself: in both cases, the spectator's judgement actually impacts someone, so both are morally motivated, but the difference in personal stakes implies a difference in potential bias. This real spectator is an agent, but one lacking personal stakes in the decision, whereas the imagined spectator is tied to a stakeholding agent, such that, even if we can get past the direct expression of the agent's interests and access his impartial spectator, applying the spectator to oneself creates openings for the aforementioned self-serving biases.

Is there something even better than this type of real spectator? I would say, yes: lots of them. Although the quasi-spectator should dominate the (reflexively applied) imagined spectator in terms of self-serving bias, the former is not perfect, and his judgement is subject to

noise. Thus, there is the straightforward statistical property that increasing the sample size bolsters one's confidence in conclusions based on such observations. There is also considerable empirical research on decisions involving groups that indicates the possibility of more balanced processes and outcomes than with individuals alone, based partly on informational grounds.[13] Of course, a well-known criticism of ideal observer theories is that their informational conditions can never realistically be satisfied, as Zagzebski (2004) points out. But I am not faulting any version of spectators discussed here for failing to be omniscient: neither Smith's impartial spectator nor the quasi-spectator approach, in contrast to ideal observer theory, aspires to perfection. Rather, the claim is that, for the purpose of empirical analysis of moral preferences, the quasi-spectator approach builds and improves on alternative spectator models.

3.3 Moral rules

Recall that one of Smith's two tasks for ethical inquiry concerns the contents of morality, that is, the particular conclusions or general rules that might emerge from the application of the other task, viz. his epistemic exercise. It is striking, though, how brief and even vague the treatment of this topic is in *TMS*, at least relative to other major contributions to moral philosophy. This is likely related to the fact that his theory is grounded in moral sentiments, rather than grand moral principles. As argued in previous sections of this paper, the existence of and search for moral rules are important both in Smith's moral philosophy and in quasi-spectator research. But as Fricke (2011) points out, even concerning the most important moral rules (the rules of justice), 'Smith does not make the exact content of these rules very explicit'.

Although descriptive detail is often sparse, Smith is quite clear about other aspects of these rules. Not only are they grounded in moral sentiments, as discussed thus far, but they are plural and not reducible to a single principle.[14] Smith frequently refers in so many words to virtues, rules, principles and measures of conduct. These include prudence,

[13] One type of evidence comes from research on deliberation, such as that reviewed in the Elster (1998) volume, a research agenda originally inspired by discourse theory (Habermas 1990 [1983]). This represents an alternate line of research into moral preferences, which, unlike the quasi-spectator model described here, involves sharing of information. Although these two approaches differ in various ways, it has been argued that they need not lead to contradictory conclusions and that they, and hybrid versions of them, might even be mutually enhancing (Konow 2009*b*).

[14] In the interests of brevity, I restrict my attention to these more limited claims and do not engage the larger debate about whether Smith was a universalist or a moral relativist. For stimulating and thoughtful perspectives, though, the reader is referred to Griswold (1999), Otteson (2002), Fleischacker (2005), Weinstein (2006), Rasmussen (2008) and Sen (2009).

courage, industry and benevolence (VI) and prohibit killing, stealing and violating the rights of others (II.ii.2.2). Economics experiments using spectators corroborate the pluralism of the moral sentiments, e.g. the studies of Cappelen *et al*. (forthcoming) and Engelmann and Strobel (2004) point to multiple principles of distributive justice. Psychology studies, including those eliciting third-party judgements, also support the descriptive accuracy of such *sentimentalist pluralism* in a variety of contexts, according to Gill and Nichols (2008).

According to Smith, there is an additional benefit of and justification for moral rules: they provide a means to cope with self-deception. Smith recognizes that applying moral rules requires complex interpretative decisions, as they 'require so many modifications, that it is scarce possible to regulate our conduct entirely by regard to them' (III.6.9), a fact that widens the opening for self-serving application of them. Fleischacker (2011) argues that 'rule-following is central to Smith's solution to self-deceit'. Moral rules serve to help overcome this weakness by sanctioning such deviations: 'Those general rules of conduct, when they have been fixed in our mind by habitual reflection, are of great use in correcting the misrepresentations of self-love' (III.4.12). Indeed, the most important rules, viz. the rules of justice, have absolute authority, according to Fricke (2011). Violation of these rules can be so harmful that Smith stresses the importance of habitual compliance with these, even when deviations might cause no harm, in order to avoid a slippery slope (III.6.10).

4. CONCLUSIONS

This paper has examined and contrasted the impartial spectator in Adam Smith's *The Theory of Moral Sentiments* and empirical methods that elicit third-party moral judgements in recent social science research. I have argued that these two approaches have a number of attributes in common. These include the goal of deriving morality from a moral sense, predominant use of inductive methods for description but not to the exclusion of deduction and prescriptive analysis, a tension between self-interest and moral preferences in which self-deception is often involved, grounding the epistemic exercise in the three properties of impartiality, information and a moral sense, a basis in the real and fallible moral judges embedded in real people rather than in views from hypothetical and idealized states, and an interest in deriving moral rules that are assumed to influence the behaviour of agents in many situations. On the other hand, important differences between the two methods were discussed. These include differences in the relative importance of each of the aforementioned attributes in the methods. More fundamentally, though, Smith's impartial spectator is imagined and primarily employed reflexively to guide and motivate individual action, whereas the

quasi-spectator method is based on multiple observations of the judgements or actions of real spectators toward other persons. I conclude with some thoughts about how research in economics and philosophy might benefit from further analysis of Smithian spectatorship and empirical spectator studies.

First, mounting evidence of the importance and complexity of moral motivation for economic activity has made it increasingly difficult for economists to justify their traditional position that economics should be a value-free science (Robbins 1932) or to sequester themselves behind a single normative criterion such as Pareto efficiency. The spectator-based theory of *TMS* offers valuable lessons for descriptive analysis of morals in economics and the other social sciences. For example, we have considered in this paper how Smith's impartial spectator bears similarities to, and sometimes has even exerted explicit influences on, recent empirical spectator methods, analysis of self-deception and theories of dual-selves. I believe there are many other unexploited insights in *TMS* that can inspire and guide empirical spectator studies and descriptive theoretical work in economics. For instance, Smith provides a rich description of the extension of affections to, respectively, other individuals, one's society and the universe of persons (VI.2) that can inform empirical research on moral views about the appropriate reference group, e.g. should fairness apply only to one's own countrymen or also to foreigners? Here quasi-spectators could provide insights into more objective views of how these questions should be resolved. As another example, Smith's treatment of the relationship between moral sentiments and custom and culture (V.2) seems to support a view of universal moral sentiments that permit some variability in their implications for conduct due to a dependence on context, including culture. Despite the recent growth of research into moral preferences and culture (e.g. Henrich *et al*. 2004), there has been almost no quasi-spectator analysis in this area. An exception is the experimental study of Konow *et al.* (2009), which finds the same fairness preferences among US and Japanese quasi-spectators but different behaviour between stakeholders in the two groups due to differences in the willingness to act on fairness. There are many other similar opportunities, including for quasi-spectator studies inspired by Smith's analysis of the relationship between morals and happiness (VI.1) and of the role of moral rules in overcoming pitfalls created by self-deception (III.4).

Second, many claim that descriptive analysis of morality is relevant to philosophical ethics and normative economics (see section 1.2 of this paper). In philosophy, the connection between descriptive and prescriptive ethics is a contentious matter. For instance, deontologists, who include Kantians, oppose judging choices based on their consequences, and most (although not all) are hostile to grounding normative ethics on actual

states, including the state of moral values. Other schools of thought, however, are open to, or even embrace, such a link. This claim is least controversial for empiricist schools such as Smith's (on this, see Campbell 1971; Griswold 1999; Sen 2009). In addition, most of normative economics derives from consequentialism, the philosophical tradition that evaluates the morality of choices based on their consequences. A straightforward consequentialist argument for the study of descriptive ethics is based on the 'ought implies can' claim: actual moral values often constrain choices or mediate their effects on the ultimate good, whatever one holds that to be.[15] For example, suppose a social planner is utilitarian, but workers are egalitarian and willing to strike if pay is sufficiently unequal: the latter fact both alters and potentially constrains the compensation scheme that maximizes the planner's social welfare. Understanding manifested morality, therefore, is not only relevant for normative economics but also for economic policy (see Hausman and McPherson 1993).

Note, however, that any arguments in favour of the relevance of descriptive ethics for prescriptive ethics *in principle* surely depend on the quality of the former *in practice*. In this regard, I have sought to demonstrate that recent empirical research on moral preferences, in particular, that involving quasi-spectators, represents a constructive advance. Evidence has been presented that this method reduces the effects of self-interest and self-deception and facilitates the inference of moral rules. There are multiple paths to establishing a descriptive/prescriptive link, but let me conclude with some thoughts on one strategy that shares many elements with Smith's approach. Gill and Nichols (2008) advocate for what they call sentimentalist pluralism, which holds that commonsense moral judgements are based on moral sentiments and guided by a plurality of moral rules, consistent with the interpretation of Smith presented in this paper. Further, they argue for normative sentimentalism, i.e. the view that the emotions are a proper ground for morality, which they trace to classical sentimentalism. Pointing to recent empirical work on moral judgement, they conclude that alternate approaches leave us 'saddled with normative consequences virtually no one is willing to accept,' whereas sentimentalist pluralism is 'the most initially promising response to these findings.' Nevertheless, they do

[15] Many other approaches to normative ethics are related to descriptive analysis. Moral realism maintains that moral claims are reducible to matters of objective fact, whereas ethical subjectivism holds that such claims are about the attitudes of real people. Much of modern virtue ethics, the school that traces its lineage to Aristotle, is *naturalistic*, i.e. endorses the view that morality is grounded in nature and subject to scientific investigation. The capabilities approach (e.g. see Nussbaum and Sen 1993), a philosophical framework that stresses freedom to pursue what one has reason to value (and relates to the Aristotelian concept of flourishing), was developed with empirical measures of what it claims to be of value.

not work out the meta-ethical foundations for normative sentimentalism or address many of the arguments against it. Thus, future work for philosophers sympathetic to this approach could be to take on these challenges, in part by reference to empirical findings of quasi-spectator studies and the strength of that method in identifying impartial moral judgements.

REFERENCES

Aguiar, F., A. Becker and L. Miller 2010. Whose impartiality? An experimental study of veiled stakeholders, impartial spectators and ideal observers. *Jena Economic Research Papers* #2010-040.

Amiel, Y., F.A. Cowell and W. Gaertner 2009. To be or not to be involved: a questionnaire-experimental view on Harsanyi's utilitarian ethics. *Social Choice and Welfare* 32: 299–316.

Ashraf, N., C. Camerer and G. Loewenstein 2005. Adam Smith, behavioral economist. *Journal of Economic Perspectives* 19: 131–145.

Babcock, L. and G. Loewenstein 1997. Explaining bargaining impasse: the role of self-serving biases. *Journal of Economic Perspectives* 11: 109–126.

Becchetti, L., G. Degli Antoni, S. Ottone and N. Solferino 2012. Whose impartiality? An experimental study of veiled stakeholders, impartial spectators and ideal observers. *Jena Economic Research Papers* #2010-040.

Bolton, G. and A. Ockenfels 2000. ERC: a theory of equity, reciprocity, and competition. *American Economic Review* 90: 166–193.

Broadie, A. 2006. Sympathy and the impartial spectator. In *The Cambridge Companion to Adam Smith*, ed. K. Haakonssen, 158–188. Cambridge: Cambridge University Press.

Brown, V. 1994. *Adam Smith's Discourse: Canonicity, Commerce and Conscience*. London: Routledge.

Brown, V. 2009. Agency and discourse: revisiting the Adam Smith problem. In *Elgar Companion to Adam Smith*, ed. J.T. Young, 52–72. Cheltenham: Edward Elgar.

Campbell, T.D. 1971. *Adam Smith's Science of Morals*. Totowa, NJ: Rowman and Littlefield.

Cappelen, A., R. Luttens, E. Sørensen and B. Tungodden 2011. Fairness in bankruptcy situations: an experimental study. Norwegian School of Economics and Business Administration, mimeo.

Cappelen, A., J. Konow, E. Sørensen and B. Tungodden. Forthcoming. Just luck: an experimental study of fairness and risk taking. *American Economic Review*.

Charness, G. and M. Rabin 2002. Understanding social preferences with simple tests. *Quarterly Journal of Economics* 117: 817–869.

Charness, G., R. Cobo-Reyes and N. Jiménez 2008. An investment game with third-party intervention. *Journal of Economic Behavior and Organization* 68: 18–28.

Chavanne, D., K. McCabe and M.P. Paganelli 2010*a*. Redistributive justice – entitlements and inequality in a third-party dictator game. *SSRN eLibrary* URL=<http://ssrn.com/abstract=1534934>.

Chavanne, D., K. McCabe and M.P. Paganelli 2010*b*. Shared experience and third-party decisions: a laboratory result. *SSRN eLibrary* URL=<http://ssrn.com/abstract=1534942>.

Clarke, P. 2007. Adam Smith, religion and the Scottish Enlightenment. In *New Perspectives on Adam Smith's The Theory of Moral Sentiments*, ed. G. Cockfield, A. Firth and J. Laurent, 47–65. Northampton, MA: Edward Elgar.

Cockfield, G., A. Firth and J. Laurent, eds. 2007. *New Perspectives on Adam Smith's The Theory of Moral Sentiments*. Northampton, MA: Edward Elgar.

Coffman, L.C. 2011. Intermediation reduces punishment (and reward). *American Economic Journal: Microeconomics* 3: 77–106.

Croson, R. and J. Konow 2009. Social preferences and moral biases. *Journal of Economic Behavior and Organization* 69: 201–212.

Dickinson, D.L. and J. Tiefenthaler 2002. What is fair? Experimental evidence. *Southern Economic Journal* 69: 414–428.

Di Tella, R. and R. Pérez-Truglia 2010. Conveniently upset: avoiding altruism by distorting beliefs about others. *NBER Working Paper* URL=<http://www.nber.org/papers/w16645>.

Dufwenberg, M. and G. Kirchsteiger 2004. A theory of sequential reciprocity. *Games and Economic Behavior* 47: 268–298.

Dunning, D., J.A. Meyerowitz and A.D. Holzberg 1989. Ambiguity and self-evaluation: the role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology* 57: 1082–1090.

Elster, J. 1998. *Deliberative Democracy*. Cambridge, MA: Cambridge University Press.

Engelmann, D. and M. Strobel 2004. Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American Economic Review* 94: 857–869.

Falk, A. and U. Fischbacher 2006. A theory of reciprocity. *Games and Economic Behavior* 54: 293–315.

Fehr, E. and U. Fischbacher 2004*a*. Social norms and human cooperation. *Trends in Cognitive Sciences* 8: 1364–1366.

Fehr, E. and U. Fischbacher 2004*b*. Third party punishment and social norms. *Evolution and Human Behavior* 25: 63–87.

Fehr, E. and K.M. Schmidt 1999. A theory of fairness, competition and cooperation. *Quarterly Journal of Economics* 114: 817–868.

Festinger, L. 1957. *A Theory of Cognitive Dissonance*. Stanford: Stanford University Press.

Firth, A. 2007. Adam Smith's moral philosophy as ethical self-formation. In *New Perspectives on Adam Smith's The Theory of Moral Sentiments*, ed. G. Cockfield, A. Firth and J. Laurent, 106–123. Northampton, MA: Edward Elgar.

Fleischacker, S. 1991. Philosophy in moral practice: Kant and Adam Smith. *Kant Studien* 82: 249–269.

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

Fleischacker, S. 2005. Smith und der Kulturrelativismus. In *Adam Smith als Moralphilosoph*, ed. C. Fricke and H.-P. Schütt, 100–127. Berlin: Walter de Gruyter.

Fleischacker, S. 2011. True to ourselves? – Adam Smith on self-deceit. *Adam Smith Review* 6: 75–92.

Forman-Barzalai, F. 2005. Sympathy in space(s). *Political Theory* 33: 189–217.

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

Fricke, C. and H.-P. Schütt 2005. *Adam Smith als Moralphilosoph*. Berlin: Walter de Gruyter.

Fudenberg, D. and D. Levine 2006. A dual-self model of impulse control. *American Economic Review* 96: 1449–1476.

Gächter, S. and A. Riedl 2006. Dividing justly in bargaining problems with claims. *Social Choice and Welfare* 27: 571–594.

Gaertner, W., J. Jungeilges and R. Neck 2001. Cross-cultural equity evaluations: a questionnaire-experimental approach. *European Economic Review* 45: 953–963.

Gill, M.B. and S. Nichols 2008. Sentimentalist pluralism: moral psychology and philosophical ethics. *Philosophical Issues* 18: 143–163.

Göçmen, D. 2007. *The Adam Smith Problem: Human Nature and Society in The Theory of Moral Sentiments and The Wealth of Nations*. London: Tauris Academic Studies.

Greene, J.D., R.B. Sommerville, L.E. Nystrom, J.M. Darley and J.D. Cohen 2001. An fMRI investigation of emotional engagement in moral judgment. *Science* 293: 2105–2108.

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

Griswold, C.L. 2006. Imagination: morals, science, and arts. In *The Cambridge Companion to Adam Smith*, ed. K. Haakonssen, 22–56. Cambridge: Cambridge University Press.

Haakonssen, K. 2006. *The Cambridge Companion to Adam Smith*. Cambridge: Cambridge University Press.

Habermas, J. 1990 [1983]. *Moral consciousness and communicative action*. Trans. Christian Lenhardt and Shierry Weber Nicholson. Cambridge, MA: MIT Press.

Hanley, R.P. 2008. Enlightened nation building; the 'science of the legislator' in Adam Smith and Rousseau. *American Journal of Political Science* 52: 219–234.

Harbaugh, W.T., U. Mayr and D.R. Burghart 2007. Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science* 316: 1622–1625.

Hausman, D.M. and M.S. McPherson 1993. Taking ethics seriously: economics and contemporary moral philosophy. *Journal of Economic Literature* 31: 671–731.

Herne, K. and T. Mård 2008. Three versions of impartiality: an experimental investigation. *Homo Oeconomicus* 25: 27–53.

Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr and H. Gintis 2004. *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-scale Societies*. Oxford: Oxford University Press.

Hill, L. 2001. The hidden theology of Adam Smith. *European Journal of the History of Economic Thought* 8: 1–29.

Huesch, M. and R. Brady 2010. Allowing repeat winners. *Judgment and Decision Making* 5: 374–379.

Kahneman, D., J.L. Knetsch and R.H. Thaler 1986. Fairness and the assumptions of economics. *Journal of Business* 59: S285–S300.

Konow, J. 2000. Fair shares: accountability and cognitive dissonance in allocation decisions. *American Economic Review* 90: 1072–92.

Konow, J. 2003. Which is the fairest one of all? A positive analysis of justice theories. *Journal of Economic Literature* 41: 1186–1237.

Konow, J. 2009*a*. Is fairness in the eye of the beholder? An impartial spectator analysis of justice. *Social Choice and Welfare* 33: 101–127.

Konow, J. 2009*b*. The moral high ground: an experimental study of spectator impartiality. *EconPapers*. Available at <http://EconPapers.repec.org/RePEc:pra:mprapa:18558>.

Konow, J. 2010. Mixed feelings: theories of and evidence on giving. *Journal of Public Economics* 94: 279–297.

Konow, J., T. Saijo and K. Akai 2009. Morals versus mores: experimental evidence on equity and equality. *EconPapers*. Available at <http://EconPapers.repec.org/RePEc:cla:levarc:122247000000002055>.

Levine, D.K. 1998. Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics* 1: 593–622.

Nussbaum, M.C. 1990. *Love's Knowledge: Essays on Philosophy and Literature*. New York, NY: Oxford University Press.

Nussbaum, M.C. and A. Sen, eds. 1993. *The Quality of Life*. New York, NY: Oxford University Press.

Otteson, J.R. 2002. *Adam Smith's Marketplace of Life*. Cambridge: Cambridge University Press.

Parrish, J. 2007. *Paradoxes of Political Ethics: From Dirty Hands to the Invisible Hand*. Cambridge: Cambridge University Press.

Raphael, D.D. 2007. *The Impartial Spectator*. Oxford: Clarendon Press.

Rasmussen, D.C. 2006. Does 'bettering our condition' really make us better off? Adam Smith on progress and happiness. *American Political Science Review* 100: 309–318.

Rasmussen, D.C. 2008. Whose impartiality? Which self-interest? Adam Smith on utility, happiness and cultural relativism. *The Adam Smith Review* 4: 247–253.

Rawls, J. 1971. *A Theory of Justice*. Cambridge, MA: Belknap Press of Harvard University Press.
Rawls, J. 2000. *Lectures on the History of Moral Philosophy*. Cambridge, MA: Harvard University Press.
Redman, D.A. 1993. Adam Smith and Isaac Newton. *Scottish Journal of Political Economy* 40: 210–230.
Robbins, L. 1932. *An Essay on the Nature and Significance of Economic Science*. London: Macmillan.
Schram, A. and G. Charness 2011. Social and moral norms in the laboratory. UCSB manuscript.
Schokkaert, E., B. Capeau and K. Devooght 2003. Responsibility-sensitive fair compensation in different cultures. *Social Choice and Welfare* 21: 207–242.
Schwitzgebel, E. 2008. The unreliability of naive introspection. *Philosophical Review* 117: 245–273.
Sen, A. 2009. *The Idea of Justice*. Cambridge, MA: The Belknap Press.
Smith, A. 1976 [1759]. *The Theory of Moral Sentiments*, ed. D.D. Raphael and A.L. Macfie. Oxford: Oxford University Press.
Sugden, R. 2002. Beyond sympathy and empathy: Adam Smith's concept of fellow-feeling. *Economics and Philosophy* 18: 63–87.
Thaler, R.H. and H.M. Shefrin 1981. An economic theory of self-control. *Journal of Political Economy* 89: 392–406.
Thompson, L. and G. Loewenstein 1992. Egocentric interpretations of fairness and interpersonal conflict. *Organizational Behavior and Human Decision Processes* 51: 176–197.
Traub, S., C. Seidl, U. Schmidt and M.V. Levati 2005. Friedman, Harsanyi, Rawls, Boulding – or somebody else? An experimental investigation of distributive justice. *Social Choice and Welfare* 24: 283–309.
Turillo, C.J., R. Folger, J.J. Lavelle, E.E. Umphress and J.O. Gee 2002. Is virtue its own reward? Self-sacrificial decisions for the sake of fairness. *Organizational Behavior and Human Decision Processes* 89: 839–865.
Utikal, V. and U. Fischbacher. 2010. On the attribution of externalities.TWI Research Paper Series.
Verburg, R. 2000. Adam Smith's growing concern on the issue of distributive justice. *European Journal of the History of Economic Thought* 7: 23–44.
Weinstein, J.R. 2006. Sympathy, difference, and education: social unity in the work of Adam Smith. *Economics and Philosophy* 22: 79–111.
Weinstein, J.R. 2007. Adam Smith's philosophy of education. *The Adam Smith Review* 3: 51–74.
Witztum, A. 1997. Distributive considerations in Smith's conception of economic justice. *Economics and Philosophy* 13: 241–259.
Young, J.T. 1992. Natural morality and the ideal impartial spectator in Adam Smith. *International Journal of Social Economics* 19: 71–82.
Zagzebski, L. 2004. *Divine Motivation Theory*. New York, NY: Cambridge University Press.