

Adam Smith and the Modern Science of Ethics

By

JAMES KONOW*
DEPARTMENT OF ECONOMICS
LOYOLA MARYMOUNT UNIVERSITY
ONE LMU DRIVE, SUITE 4200
LOS ANGELES, CA 90045-2659

Abstract

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 important opportunities for philosophical ethics to inform empirical research and vice versa.

*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.

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, *The Theory of Moral Sentiments*, VII.iii.2.6, VII.iv.6.

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 (potentially) 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 concerning others but not themselves. Spectators of a kind were also at the center 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 center of that science differs 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 in the paper, 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 judgments 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

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

people, specifically, ones who reveal the moral judgments of their own presumed impartial spectators concerning matters affecting others. I call these *quasi-spectators*, since their judgments 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., Charness and Rabin (2002), Coffman (forthcoming), Engelmann and Strobel (2004) and Konow (2000). These differ from most economics experiments on morals, which have been based on stakeholders whose decisions potentially affect their own earnings. Of interest here are also experiments in which third parties incur a fixed cost to influence the earnings of others, e.g., Charness, Cobo-Reyes and Jiménez (2008), Fehr and Fischbacher (2004b) and Kahneman, Knetsch and Thaler (1986). These are hybrids of quasi-spectator and stakeholder studies, since the decision-makers are third parties, but they incur a personal cost to affect others. 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. These 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 judgments given the lack of material consequences. Thus, the discussion here of such studies focuses on some recent investigations that are among those that purposely target spectator views and provide evidence strongly suggesting their participants were appropriately motivated.

To be clear, the intent of this study is not to demonstrate an equivalency between Adam Smith's moral theory and a research program in behavioral economics. Nor is the title of this paper meant to imply an exhaustive treatment of both. Rather, this is an examination of issues at their intersection that might prove useful for addressing open questions in both research agendas. Attention to the commonalities should not be taken as a lack of appreciation of their dissimilarities. Rather, even many differences, I argue, suggest ways in which the two might complement one another. Ultimately, insights from Smith's impartial spectator can inspire and help improve empirical spectator studies and, consequently, economic theory² and, in turn,

² Indeed, some recent empirical spectator studies discussed later in the paper make explicit reference to Smith, e.g.,

empirical spectator findings might also inform philosophical ethics. But the immediate goal is to demonstrate the potential value of such a program through a comparative analysis. To be sure, Smith's impartial spectator differs in important ways from quasi-spectator studies and is so rich a concept that a thorough treatment of it is beyond the scope of this paper. Thus, the analysis remains close to common boundaries of the two spectator concepts while limiting the treatment of differences that are often important but further afield.

The discussion is organized around what I claim is a remarkable but tractable area of overlap in the two approaches in terms of their respective goals, methodologies, assumptions about human nature, spectator attributes, and emphasis on moral rules. Specifically, 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 toward a conflict between self and others in which self-deception is often implicated, conceive of impartiality in terms of three properties (absence of stakes, information and a common moral sense), and conclude that the moral sense can often be characterized by rules. The paper proceeds in this same order from parallel discussions of *TMS* and quasi-spectator research beginning with their respective goals and methodologies before proceeding to a more detailed treatment of their assumptions, particular models and conclusions.

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 tenour 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 judgment, 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" or moral sense. This grows out of the tradition of moral sense theory that includes Smith's mentor, Francis Hutcheson, and his friend,

Aguiar, Becker and Miller (2010), Chavanne, McCabe and Paganelli (2010a), Croson and Konow (2009) and Konow (2009a,b).

David Hume, an approach that proceeds from our sense of approval and disapproval of our actions or those of others. According to Smith, this moral sense develops through a process of repeated interactions with others that are initially motivated by our natural desire for their approval. This subject 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., Broadie 2006, Weinstein 2007, Young 1992), and it will resurface later in this paper. Nevertheless, the main topic here is the realized impartial spectator, rather than its genesis, for variety of reasons, including for the sake of brevity and in light of the lack of empirical quasi-spectator studies on the formation of moral judgment.

Raphael (2007: 9-10) argues that *TMS* focuses mostly on the epistemic exercise and dedicates considerably 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 judgment can be characterized by rules is both his assumption going into the epistemic exercise as well as his conclusion coming out of it. More specifically, 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 studies³ have similarly addressed these two questions, although the greater emphasis has clearly been on investigating the content of moral preferences, including altruism (Harbaugh, Mayr and Burghart, 2007), distributive justice (e.g., Chavanne, McCabe and Paganelli, 2010a, Dickinson and Tiefenthaler, 2002), fair rewards from risk-taking (Cappelen, Konow, Sørensen and Tungodden, 2011, Huesch and Brady, 2010), fair distribution of losses from risk-taking (Cappelen, Luttens, Sørensen and Tungodden, 2011), reciprocity (Charness, Cobo-Reyes and Jiménez, 2008, Croson and Konow, 2009), and the effects of moral bias (e.g., Chavanne, McCabe and Paganelli, 2010b, Croson and Konow, 2009, Traub, Seidl, Schmidt and Levati, 2005). As with the larger empirical literature on social preferences that includes stakeholder studies, this research has usually been approached with the expectation that general forces are at work and with the goal of contributing to models of those forces (e.g., Charness and

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

Rabin, 2002, Konow, 2000). 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 judgments (Konow, 2009a,b), possible spectator bias (e.g., Aguiar, Becker and Miller, 2010, Chavanne, McCabe and Paganelli, 2010b), and differences between spectatorship and other concepts of impartiality such as Rawls's (Herne and Mård, 2008, Aguiar, Becker and Miller, 2010).

These quasi-spectator studies have been social science research projects aimed at describing social behavior and the preferences that underlie it. Smith, on the other hand, places his objective with *TMS* squarely in the domain of philosophy, a point that Part VII of *TMS* 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 their particular methods and conceptualizations of 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 otherwise 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 sense 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 observation of the words and behavior 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 endeavors, 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 judgments, 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 here the use of general rules to regulate behavior 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, Jungeilges and Neck (2001).

The distinction between induction and deduction in these research agendas is also closely connected to the descriptive and prescriptive intentions of their analyses. 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, for example, in Gächter and Riedl (2006). Indeed, some empirical studies are explicitly designed to test normative theories, as is the case, for example, with many empirical social choice investigations such as those of Amiel, Cowell and Gaertner (2009) and Schokkaert, Capeau and Devooght (2003).

There is wide agreement among Smith scholars of the descriptive intent in *TMS*, indeed, the more controversial claims concern the extent to which it 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. 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). Most commentators see some of both purposes in Smith, although to varying degrees. 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. Firth argues that “Smith wants his readers and students to dedicate themselves to the attainment of inner moral purity” (2007, pg. 120 in CFL). Fleischacker (1991) writes that “moral philosophers, he believes, should contribute to moral practice” (pg. 255). 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 specter 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.

Quasi-spectator studies, as defined here, are descriptive undertakings and are not promoted as prescriptive analysis. Nevertheless, many authors do suggest their results are relevant to reflection on philosophical ethics (Amiel, Cowell and Gaertner, 2009, Herne and Mård, 2008), normative economics (e.g., Gächter and Riedl, 2006, Konow, 2003, 2009a,b), and policy (e.g., Chavanne, McCabe and Paganelli, 2010a, Traub et al., 2005). The potential 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 (“ought”) views elicited from a particular group in a particular way. The implicit claim motivating 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 exactly in this manner in the quasi-spectator literature. But the subject matter and statements of some of these contributions suggest the view that the validity of normative claims depends in some (as yet incompletely specified) manner on their ability to be reconciled with the actual moral judgments of impartial parties elicited under certain conditions.

Whatever one’s take on the aforementioned debate regarding positive vs. normative intent in *TMS*, Smith explicitly employs both prescriptive and descriptive terms in advancing the impartial spectator. 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 judgments of a well-informed third party. That is, the spectator is embedded in the agent. To be sure, Smith describes sometimes fierce conflicts between the spectator and the agent’s other desires and goals, but the impartial spectator is the correspondence of a subset of the agent’s views. This parallels the version of how “is” and “ought” can be connected in quasi-spectator research above: normatively valid claims are based on a subset of actual views, viz., the moral judgments of informed and impartial third parties, literal ones in the case of quasi-spectators and an imagined one in the case of Smith.

Smith scholars have interpreted the is-ought relationship in *TMS* in different ways. Raphael (2007: 133-135) sees Smith’s usage in psychological, rather than philosophical, terms. Griswold (1999) points to the tension between agent and spectator perspectives in contrasting everyday behavior 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. Firth argues that Smith presents the impartial spectator not only as a device to explain behavior but also an ideal to which people should aspire, indeed, one that is practically attainable for a segment of the population (2007: 118). Thus, whatever their differences, a common view of commentators is that *TMS* has normative purpose grounded in a descriptive construct, viz., the impartial spectator.

2. HUMAN NATURE: AGENT AND SPECTATOR

This section explores some parallels between Smith’s vision of human nature and recent behavioral approaches in economics, including expanded discussions of the respective spectator

models in both.⁴

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 *TMS*. In Smith’s view, there are two parts to an individual: 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 but one of many instances of Smith anticipating much later work in the social sciences. Some recent research in psychology and economics revolves around dual decision-makers, and, in the latter discipline, theoretical accounts of dual-selves mirror, in many respects, the roles of spectator and agent in Smith, e.g., Fudenberg and Levine (2006), and Shefrin and Thaler (1981). In *TMS*, as in modern research, these dual-selves are often seen as being in conflict.

Despite the attraction of dual-self models, most contributions to formalizing this moral conflict, in both quasi-spectator studies and the more general behavioral economic literature, have adopted a simpler approach that involves a single decision-maker trading off material self-interest with social preferences, e.g., Andreoni and Miller (2002), Bolton and Ockenfels (2000), Charness and Rabin (2002), Dufwenberg and Kirchsteiger (2004), Engelmann and Strobel (2004), Falk and Fischbacher (2006), Fehr and Schmidt (1999), Konow (2000) and Levine (1998). Although social preferences are variously specified in these models as altruism, reciprocal altruism, equality, equity, maximin, efficiency, or some subset thereof, they share the feature of portraying a tension between material utility and an internalized motive that includes the interests of others.

These internalized and opposing goals, i.e., between one’s own interests and those of

⁴ *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.

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 judgments 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 behavior of others. This leads to a moral sense, eventually manifested as an imagined third party and experienced chiefly in affective terms, which then 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 perhaps better stated, *detached*) third party, who has no stakes in the situation or parties who are being evaluated. This, of course, serves to avoid moral judgments being tainted by self-interest. But there is a potential problem, as Griswold (2006) points out, in disassociating 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-centered. Quasi-spectator studies and modern behavioral 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 alone or reflects more general considerations that potentially incorporate the interests of others, respectively. A generalization of this practice provides, in my view, a productive distinction for both empirical research and for 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 other people, whereas any intrinsic reward or punishment that is derived from thinking of or 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 void of meaning. This proposed distinction, by contrast, seems consistent with Smith’s treatment of sympathy as praiseworthy and of self-interest as being a different and generally opposed motive.⁵

Second, impartiality must be coupled with the necessary information conditions. Smith states on several occasions that the morally relevant judgments are those of the “impartial and

⁵ I thank a referee for prompting me to clarify the relationship between self-interest and sympathy and, thereby, to add this distinction to the paper.

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 program unless, of course, the respective spectator is endowed with a moral sense. This is related, but not identical, to sympathy in Smith, a complex and 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 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 the situation 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-Barzilay (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, i.e., the set of moral judgments and rules embraced by the spectator, 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 judgments. 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 judgment 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, exploring 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, suggesting he uses the modifier *ideal* in the sense of *imagined* rather than *perfect*.⁶ 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 judgments, 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 limitations of the agent who creates him, but rather be inherent to the spectator’s balancing act.

Most quasi-spectator studies employ a number of 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” (or dictator experiment) of Konow (2000). In the standard version of this experiment, 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 quasi-spectator (which I will call here “spectator” for short), divided a sum of money between two anonymous subjects.⁷ The dictators in the standard stakeholder treatment demonstrated a bias, taking for themselves, on average, more than third parties in the spectator treatment gave similarly situated subjects in their treatment.

⁶ Although some, e.g., Campbell (1971) and Rawls (2000), have read Smith as presenting an ideal observer theory, i.e., as claiming that ethical judgments 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).

⁷ 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.

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 and produce mixed results. In a three player dictator game, Chavanne, McCabe and Paganelli (2010b) 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, suggesting that even spectators having an association with one party can remain impartial. Aguiar, Becker and Miller (2010) 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 spectators whose initial endowments equal those of one of the two stakeholders favor such stakeholders compared to spectators whose initial endowments did not equal those of any stakeholders. Thus, it appears that deliberate introduction of partial considerations can disturb spectator impartiality, although this 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).⁸

There are relatively few carefully controlled empirical studies of feelings and morality, still fewer involving quasi-spectators. The closest, to my knowledge, is an experiment by Harbaugh, Mayr and Burghart (2007) using dictators who can allocate money to a charity, specifically a local food bank 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 and self-reported subjective satisfaction scores indicate that, holding dictator earnings constant (similar to spectators), they subsequently feel better,

⁸ 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.

when the charity receives more. The dictator study of Konow (2010) uses before and after self-reported measures of feelings and 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.⁹ Greene et al. (2001) examine whether feelings affect moral judgment 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, on the one hand, and moral judgments and actions, on the other.

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 judgment 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 in practical contexts. Additional information might conceivably complicate moral reasoning and encourage divergent views, undermining attempts for consensus. Moreover, some research on stakeholders, such as 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-

⁹ Nevertheless, compared to a control group of subjects who were given an endowment and no opportunity to transfer, dictators transferring to charities did not feel significantly better, and the less generous subjects in that group actually felt significantly worse compared to 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.

spectator research. A hypothesis in Konow (2009a) is that relevant information permits third parties to reason more accurately and, therefore, reach consensus. In fact, the moral judgments 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 (2009b) 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 the putative variability of impartiality, incompleteness of 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 color individual moral judgment. 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 the judgments of quasi-spectators should 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 on the impartiality of quasi-spectators comes from the vignette study (Konow, 2009b), where stakeholders have different and 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 behavior. Another type of evidence comes from the

dispersion in views. A potential concern is that quasi-spectators might be insufficiently motivated to acquire or process relevant information about others and might instead choose randomly or capriciously. On the other hand, if the quasi-spectator method succeeds, stakeholders should be expected to exhibit higher variance in their judgments 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 conduct. Quasi-spectators, on the other hand, 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 proposition is confirmed in studies of distributive preferences (Konow, Saijo and Akai, 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 Englemann and Strobel (2004) and Cappelen, Konow, Sørensen and Tungodden (2011). 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 (2004a,b), Kahneman, Knetsch and Thaler (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 (2009a, 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 and are always truthful with themselves and 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 (forthcoming) 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 (e.g., 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 behavior, 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. He argues it is unreliable in 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 judgments.

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. If one accepts the possibility of real impartiality and a

common moral sense, this seems an improvement over theories of impartiality that rely on constraints on information, such as the Rawlsian veil of ignorance, which might require withholding information that could create bias but that might, nonetheless, be necessary to render accurate moral judgments. 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, Meyerowitz and Holzberg (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.

The impartial spectator is a mental image, but another type of spectator Smith discusses is the real spectator, i.e., a genuine third party who observes and judges the conduct of others.¹⁰ To be sure, the real spectator is not as prominent 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 these 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

¹⁰ 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 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 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 behavior.

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*, in which 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 judgment 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).¹¹ Smith writes at great length in *TMS* on the natural harmony between Nature and the moral values fostered through socialization. In countering this particular 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:

¹¹ 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 Clarke, 2007, Hill, 2001, and Otteson, 2002). 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.

“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 the imagined and real spectators are typically aligned.

Both real and imagined spectators, therefore, sometimes err, the former due to occasional deviations from the highest moral authority and the latter also because of self-serving biases of the agent’s imagination. Nevertheless, Smith views both types of spectators as usually accurate and mutually consistent resources whose judgments dominate, in moral terms, those of the agent. The largest share of *TMS* treats individual judgment and behavior, and, consequently, the frequent reflexive use of the imagined spectator. This use might be viewed, at least in part, as a practical matter: that individuals 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. This relates to the epistemic exercise in moral philosophy to which Smith refers and is of interest to many modern social scientists, including moral psychologists, political theorists and behavioral economists, as an aid to both descriptive analysis and policy design. What kind of spectator might we want? I will argue that an answer to this question that takes account of the considerations raised above and that builds on materials Smith provides is the quasi-spectator approach described previously. The impartial spectator, which the agent imagines for application to his own behavior, is, as discussed, subject to self-serving biases. Consider, instead, a real spectator modeled 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 judgment (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 imbedded in him. This person is neither omniscient, because the context does not usually provide or permit the acquisition of all relevant information, nor infallible, given potential residual effects of individual interests and insufficient sympathy. Nevertheless, the judgment of this spectator should be superior to that of the imagined one the agent applies to himself: in both cases, the spectator's judgment actually impacts others, so both should be appropriately motivated, but the difference in personal stakes implies in 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, the latter is vulnerable to the aforementioned self-serving biases.

Is there something even better than this type of real spectator (i.e., quasi-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 judgment is subject to noise. Thus, there is the straightforward statistical property that increasing the sample size bolters 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.¹² 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 specific purpose of empirical analysis of moral preferences, the quasi-spectator approach builds and improves on alternative spectator

¹² 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, 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, 2009b).

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 (forthcoming) 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.¹³ Smith frequently refers in so many words to virtues, rules, principles and measures of conduct. These include prudence, courage, industry, and benevolence (VI) and prescribe 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, Konow, Sørensen and Tungodden (2011) and Engelmann and Strobel (2004) point to multiple principles of distributive justice. Psychology studies, including those eliciting third party judgments, also support the descriptive accuracy of such *sentimentalist pluralism* in a variety of contexts, according to Gill and Nichols (2008).

The importance of moral rules to the impartial spectator and quasi-spectator approaches has been discussed and defended throughout this paper. According to Smith, there is an additional benefit of and justification for such 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 (forthcoming) argues that "rule-following is central to Smith's solution to self-

¹³ 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 Fleischacker (2005), Griswold (1999), Otteson (2002), Rasmussen (2008), Sen (2009), and Weinstein (2006).

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 (forthcoming). 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 judgments 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 often to influence the behavior of agents. 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 is primarily employed reflexively to guide and motivate individual action, whereas the quasi-spectator method is based on multiple observations of the judgments or actions of real spectators concerning other persons.

The goal of this exercise was less to close the book on certain subjects than to open further avenues of research suggested by this comparison. Thus, I would consider it a sign of success if more questions were raised than answered. One important area, in my view, concerns analysis of possible links between descriptive and prescriptive ethics. What are the philosophical foundations for turning attention to empirical findings about impartial moral judgments when analyzing, criticizing and even formulating normative statements and theories? What are the limits of these arguments and practices? What are the implications of empirical research for ethical theories and moral epistemology? These questions have a long philosophical tradition and

are central to current debates, especially in empiricist schools such as Smith's (see, for example, Campbell 1971, Griswold 1999, and Sen 2009). But any arguments in favor 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 here that recent empirical research on moral preferences, in particular, that involving quasi-spectators, represents an important advance.

Similarly, the work of economists and other social scientists who conduct empirical research on moral preferences would, in my view, benefit from greater engagement with the philosophical arguments concerning the intersection of descriptive and prescriptive analysis. Given the mounting, and now substantial, evidence on the importance and richness of moral motivation for economic activity, it has become increasingly difficult for economists to justify their traditional position that economics be a value-free science (e.g., Robbins, 1932) or to sequester themselves to some single normative standard such as Pareto efficiency. This is particularly so for research that might, intentionally or not, influence policy, since policy is ultimately built on normative foundations (even when those foundations are not explicit). Thus, lessons from the philosophical discourse on the relationship between descriptive and prescriptive analysis could prove useful for the design of empirical research. For example, arguments in favor of reflected moral judgment, rather than naïve intuition, might imply the elicitation of certain types of views or those of certain types of subjects in quasi-spectator studies or the introduction of communication among subjects, as in empirical research on deliberative democracy.

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.
- Andreoni, J. and J. Miller 2002. Giving according to GARP: an experimental test of the consistency of preferences for altruism. *Econometrica* 70 (2): 737-753.
- Ashraf, Nava, Camerer, Colin, and Loewenstein, George (2005). "Adam Smith, Behavioral Economist." *Journal of Economic Perspectives* 11 (Summer): 109-26.
- Babcock, L. and G. Loewenstein 1997. Explaining bargaining impasse: the role of self-serving biases. *Journal of Economic Perspectives* 11: 109-26.
- Bolton, G. and A. Ockenfels 2000. ERC: a theory of equity, reciprocity, and competition. *American Economic Review* 90 (1): 166-193.
- Broadie, A. 2006. Sympathy and the impartial spectator. In *The Cambridge Companion to Adam Smith*, ed. K. Haakonssen, 158-188. Cambridge, UK: 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, UK: Edward Elgar.
- Campbell, T.D. 1971. *Adam Smith's Science of Morals*. Totowa, NJ: Rowman and Littlefield.
- Cappelen, A., J. Konow, E. Sørensen and B. Tungodden 2011. Just luck: an experimental study of fairness and risk taking. Resubmitted to the *American Economic Review*.
- 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.
- Charness, G., R. Cobo-Reyes and N Jiménez 2008. An investment game with third-party intervention. *Journal of Economic Behavior and Organization* 68 (1): 18-28.
- Charness, G. and M. Rabin 2002. Understanding social preferences with simple tests. *Quarterly Journal of Economics* 117 (3): 817-869.
- Chavanne, D., K. McCabe and M.P. Paganelli 2010a. Redistributive justice – entitlements and inequality in a third-party dictator game. *SSRN eLibrary* <http://ssrn.com/abstract=1534934>.
- Chavanne, D., K. McCabe and M.P. Paganelli 2010b. Shared experience and third-party decisions: a laboratory result. *SSRN eLibrary* <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*, eds. Geoff Cockfield, Ann Firth, and John Laurent, 47-65. Northampton, MA: Edward Elgar.

- Cockfield, Geoff, Ann Firth, and John Laurent (eds.) 2007. *New Perspectives on Adam Smith's The Theory of Moral Sentiments*. Northampton, MA: Edward Elgar.
- Coffman, L.C. (forthcoming). Intermediation reduces punishment (and reward). *American Economic Journal: Microeconomics*.
- Croson, R. and J. Konow 2009. Social preferences and moral biases. *Journal of Economic Behavior and Organization* 69 (3): 201-212.
- Dickenson, 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* <http://www.nber.org/papers/w16645>.
- Dufwenberg, M. and G. Kirchsteiger 2004. A theory of sequential reciprocity. *Games and Economic Behavior* 47(2): 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 (4): 857-869.
- Falk, A. and U. Fischbacher 2006. A theory of reciprocity. *Games and Economic Behavior* 54 (2): 293-315.
- Fehr, E. and U. Fischbacher 2004a. Social norms and human cooperation. *Trends in Cognitive Sciences* 8 (4): 1364-1366.
- Fehr, E. and U. Fischbacher 2004b. 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 (3): 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*, eds. Geoff Cockfield, Ann Firth, and John 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: Princeton University Press.
- Fleischacker, S. 2005. Smith und der Kulturrelativismus. In *Adam Smith Als Moralphilosoph* eds. Fricke,

- Christel, and Schütt, Hans-Peter, 100-127. Berlin, New York: Walter de Gruyter, Inc.
- Fleischacker, S. (forthcoming). True to ourselves? – Adam Smith on self-deceit. *Adam Smith Review*, vol. VII.
- Forman-Barzalai, F. 2005. Sympathy in space(s). *Political Theory* 33 (2): 189-217.
- Fricke, C. (forthcoming). Adam Smith and ‘the most sacred rules of justice.’ *Adam Smith Review*, vol. VII.
- Fricke, Christel, and Schütt, Hans-Peter (2005). *Adam Smith Als Moralphilosoph*. Berlin, New York: Walter de Gruyter, Inc.
- Fudenberg, D. and D. Levine 2006. A dual-self model of impulse control. *American Economic Review* 96 (5): 1449-1476.
- Gächter, S. and A. Riedl 2006. Dividing justly in bargaining problems with claims. *Social Choice and Welfare* 27 (3): 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, New York: Tauris Academic Studies.
- Greene, J.D., R.B. Sommerville, L.E. Nystrom, J.M. Darley and J.D. Cohen. 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, UK: 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, UK: Cambridge University Press.
- Haakonssen, K. 2006. *The Cambridge Companion to Adam Smith*. Cambridge, UK: Cambridge University Press.
- Habermas, J. 1983 [1990]. *Moral Consciousness and Communicative Action*. Trans. Christian Lenhardt and Shierry Weber Nicholson. Cambridge, MA: MIT press.
- Hanley, Ryan Patrick (2008). "Enlightened Nation Building; the ‘Science of the Legislator’ in Adam Smith and Rousseau." *American Journal of Political Science* 52 (April): 219-34.
- 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.
- Herne, K. and T. Mård 2008. Three versions of impartiality: an experimental investigation. *Homo Oeconomicus* 25 (1): 27-53.

- Hill, L. 2001. The hidden theology of Adam Smith. *European Journal of the History of Economic Thought* 8 (1): 1-29.
- Huesch, M. and R. Brady 2010. Allowing repeat winners. *Judgment and Decision Making* 5 (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 (4): 1072-92.
- Konow J. 2003. Which is the fairest one of all?: a positive analysis of justice theories.” *Journal of Economic Literature* 41(4): 1186-1237.
- Konow, J. 2009a. Is fairness in the eye of the beholder?: An impartial spectator analysis of justice. *Social Choice and Welfare* 33 (1): 101-127.
- Konow, J. 2009b. The moral high ground: an experimental study of spectator impartiality. *EconPapers* <http://EconPapers.repec.org/RePEc:prz: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* <http://EconPapers.repec.org/RePEc:cla:levarc:122247000000002055>.
- Levine, D.K. 1998. Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics* 1 (3): 593-622.
- Nussbaum, M.C. 1990. *Love's Knowledge: Essays on Philosophy and Literature*. New York: Oxford University Press.
- Otteson, J.R. 2002. *Adam Smith's Marketplace of Life*. Cambridge, UK: Cambridge University Press.
- Parrish, J. 2007. *Paradoxes of Political Ethics: From Dirty Hands to the Invisible Hand*. Cambridge, UK; New York: 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 (3): 309-18.
- 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, John (1971). *A Theory of Justice*. Cambridge: Belknap Press of Harvard University Press.
- Rawls, J. 2000. *Lectures on the History of Moral Philosophy*. Cambridge, MA; London: Harvard University Press.
- Redman, D.A. 1993. Adam Smith and Isaac Newton. *Scottish Journal of Political Economy* 40 (2): 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 (2): 245-73.
- Sen, Amartya (2009). *The Idea of Justice*. Cambridge: The Belknap Press.
- Smith, Adam. 1759 (1976). *The Theory of Moral Sentiments*. D.D. Raphael and A.L. Macfie (eds), Oxford: Clarendon Press; reprinted in Indianapolis, IN: Liberty Fund (1984).
- Sugden, R. 2002. Beyond sympathy and empathy: Adam Smith's concept of fellow-feeling. *Economics and Philosophy* 18 (1): 63-87.
- Thaler, R.H. and H.M. Shefrin 1981. An economic theory of self-control. *Journal of Political Economy* 89 (21): 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.
- 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 (1): 79-111.
- Weinstein, J.R. 2007. Adam Smith's philosophy of education. *The Adam Smith Review* 3: 51-74.
- Weinstein, J.R. 2008. Review of *The Impartial Spectator*, D.D. Raphael. In *Economics and Philosophy* 24 (1): 129-137.
- Witztum, A. 1997. Distributive considerations in Smith's conception of economic justice. *Economics and Philosophy* 13 (2): 241-259.
- Young, J.T. 1992. Natural morality and the ideal impartial spectator in Adam Smith. *International Journal of Social Economics* 19 (10/11/12): 71-82.
- Zagzebski, L. 2004. *Divine Motivation Theory*. New York: Cambridge University Press.