Adam Smith and Moral Knowledge

Konow, James

Loyola Marymount University

September 2009

COMMENTS WELCOME September 2009

Adam Smith and Moral Knowledge

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

Abstract

This paper examines the contribution of *The Theory of Moral Sentiments* to the study of how we acquire moral knowledge. In Smith, this is associated with the moral judgment of an impartial spectator, a hypothetical ideal conjured in the imagination of an agent. This imagined spectator has the properties of impartiality, information and sympathy. I argue Smith develops this construct in the context of *personal ethics*, i.e., as a guide to moral conduct in personal relationships. There are limitations, however, to this model for personal ethics, as acknowledged by Smith himself and suggested by subsequent social science findings. Moreover, this model does not necessarily extend to *social ethics*, i.e., to moral judgment in less personal economic and social interactions, such as firms, industries and governments. Hence, I propose modifying the spectator model in light of modern social science methods and of Smith's own insights to address its limitations for personal ethics and to provide it with a foundation for social ethics. The proposed approach is based on a *quasi-spectator*, i.e., the empirical analysis of the moral views of real spectators whose properties approximate those of the ideal spectator. A review of quasi-spectator studies suggests this as a promising method for informing both descriptive and prescriptive ethics.

*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 Maria Carrasco, Thomas Cushman, Sam Fleischacker, Christel Fricke and John O'Neill for very helpful and constructive comments. Any shortcomings remain, of course, the sole property of the author.

I. Introduction

Adam Smith's first book, *The Theory of Moral Sentiments* (hereafter TMS), was generally well received and acknowledged as an important contribution to moral theory and practice soon after its publication in 1759. Nevertheless, Smith's second major work, *The Wealth of Nations* (hereafter WN), soon eclipsed TMS and established his reputation as the father of modern economics. Smith's first book, however, has experienced a significant renaissance in recent years, prompting a considerable volume of research across the humanities and the social sciences. Both the depth and the breadth of recent Smith scholarship have been quite impressive and include Brown (1994), Fleischacker (1999), Fricke and Schütt (2005) and Weinstein (2006) in philosophy, Ashraf, Camerer and Loewenstein (2005), Rothschild and Sen (2006), Sugden (2002) and Verburg (2000) in economics, and Hanley (2008), Parrish (2007) and Rasmussen (2006) in political science, among many other notable contributions.

The Theory of Moral Sentiments is usually seen as explicating Smith's moral theory. I will focus in this paper on various questions addressed in TMS about the acquisition of moral knowledge. What are the conditions for acquiring moral knowledge? What was the social context in which Smith developed his model? What are the limitations of his method? Can his approach be productively extended to new domains? I believe these questions, and especially their answers, are of great importance not only for philosophy but also for the social sciences. Clarification of the conditions for the acquisition of moral knowledge can guide descriptive analysis of moral learning and moral development. In addition, such a model could inform the conduct of social science research and policy itself. That is, social scientists might adopt methods inspired by such models to infer empirically the underpinnings of moral conduct and moral rules. The lessons produced from these methods could, in turn, serve two purposes. First, they could improve the descriptive analysis of the impact of morals on behavior by helping to isolate their effects from those of others forces, e.g., self-interest. Second, they could be utilized for prescriptive analysis, either to test the consistency of behavior with normative claims or even to inform prescriptive analysis. The latter function is not only important for prescriptive theoretical work but is particularly critical to social scientists who are called upon to provide policy advice and who require some first principles for formulating, evaluating and defending such proposals.

Section II of this paper describes Smith's approach to moral judgment, which involves each agent constructing in his or her imagination an impartial spectator, who provides a model

for evaluating and directing right conduct. Section III summarizes Smith's ethics, i.e., some of the conclusions at which Smith arrives using this model. Section IV presents limitations of Smith's model and a new Adam Smith Problem that concerns the acquisition of moral knowledge in TMS and WN. Section V suggests an approach to moral knowledge based on an empirical method as a means to address these limitations and bridge the gaps between TMS and WN. This involves building on the original model using modern social science methods and Smith's own insights. Section VI concludes.

II. Smith's Epistemic Model

TMS is replete with conceptual originality, keen social observation, and philosophical insight, including about the acquisition of moral knowledge. Given this, one might expect a treatment in it of the broader questions of moral epistemology, e.g., how is moral knowledge possible? Nevertheless, as Fleischacker (2004) notes, "Smith has surprisingly little to say, directly, about epistemology" (pg. 27), and relatively little has been written about Smith in this regard: Harrison (1995) and Kawall (2006) are among few works that even skirt this topic. Although Smith did not emphasize epistemology in general terms in his work, TMS is teeming with rich, detailed and trenchant statements about moral learning, moral knowledge and moral judgment. This paper focuses on the ideal conditions for moral judgment, or what I will call the *epistemic model*, of Adam Smith.[1] I believe his thinking in this area represents one of the most original, important and exciting of his contributions.

That Smith's subject was, in large part, the acquisition of knowledge about right conduct is suggested by the sub-title added with edition 4 of *TMS*: "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 also portends the dual nature of people as both observers and participants, which is developed in TMS. In Smith's view, a person is divided into two parts: 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

[1] An important, but distinct, topic in TMS is the stages of moral learning. Although I will touch on this, the central subject matter of this paper concerns the conditions for acquiring moral knowledge by a *fully realized* moral judge.

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. Recent research in psychology and economics points to dual-decision-makers, and, in the latter discipline, theoretical accounts of dual-self models often refer to a principal and an agent that mirror, in many respects, the roles of spectator and agent, respectively, in Smith (e.g., Shefrin and Thaler, 1981). In TMS, as in the later empirical research, these dual-selves are often seen as being in conflict.

In broad terms, Smith is a successor to earlier ideal observer theories, including those of his friend, David Hume, and his mentor, Francis Hutcheson, which deem an action or a rule morally right if an ideal observer would approve of it. Whereas spectators were often portrayed by his predecessors as real individuals, however, Smith develops a more complex and subtle theory. His spectator is not a real person or persons but rather a model conjured in the imagination of the agent, and Smith 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 ideal observer possesses.

Smith fleshes out in some detail the origin of the imagined spectator: he is the product of society (see, especially, TMS VI.1). People participate in social interactions and, as real spectators (i.e., third parties), judge the behavior of others. This produces the moral sense or conscience, which experiences morality chiefly in affective terms. 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."

Smith's impartial spectator represents a rich and sublime construct, but I believe the core characteristics can be distilled into three fundamental properties: impartiality, information and

sympathy. First, the spectator is impartial, a modifier that also represents a Smithian addition to the language of spectator theory. That is, 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, ensures that moral judgments are not tainted by any trace of self-interest.

Second, impartiality must be coupled with the necessary information conditions. Smith states that the morally relevant judgments are those of the "impartial and well-informed spectator." Although, at various points, he acknowledges the challenges to ideal impartiality, these conditions are not an unattainable abstraction but rather a state that real people can sometimes achieve. 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 behaviour." 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." These passages help underscore the importance of these two properties of the imagined spectator. First, the spectator is not and never will be implicated in a situation being evaluated, that is, he has no perceived stake that might bias his judgments of right and wrong. Second, the spectator is fully informed of the relevant particulars and processes this information rationally with respect to internalized values.

Third, an important ingredient to spectatorship is sympathy (or *fellow-feeling*, in the terminology of Sugden, 2002). Although the emphasis, both in Smith's writings and in those of many of his commentators, is on sympathy as the mutual sharing of the feelings of others, I believe this includes adopting the positions of others quite generally, i.e., both cognitively and affectively. I see this kind of sympathy as serving two functions in Smith, similar to the distinction Rawls (2000) makes in reference to Hume's spectator. On the one hand, 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. On the other hand, 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.

For the goals of this paper alone, one might argue that sympathy can be subsumed under the second property of information where it serves a purely epistemic function. The first two properties, impartiality and information, should certainly assume a prominent position in our thinking, as they mirror the dual purposes of TMS more generally. On the one hand, it reads like a psychological and sociological treatise on the development and practice of social norms that might emerge from empiricist philosophy. Both the impact of the spectator on the agent as well as the spectator's search for right conduct are products of unambiguous, if casual, social science analysis. On the other hand, the treatment of morals is not solely descriptive but also prescriptive: normative words like "should" and "ought" occur with great frequency. 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). That is, TMS is also a guide for acquiring and applying moral knowledge in order to promote moral behavior.

Perhaps the clearest indication of Smith's belief in the strength 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 all others, provides the means to identify right and to distinguish it from wrong.

To summarize, Smith's account of human nature in TMS presents two sides: the imagined spectator and the agent. The imagined spectator is a product of socialization, who develops a conscience that is consonant with the values of society and God. The agent attempts to apply reflexively the lessons of the former, although Smith distinguishes degrees of success from the imperfections of a child or person of weak character to the self-command of the ideal

imagined spectator (III.3). The central epistemic claims are that the appropriate moral judge is an imagined spectator, who is impartial, informed and sympathetic, and that this ideal spectator can be summoned by real people.

III. Smith's Ethics

Up to this point, we have considered Smith's conditions for moral judgment, or his epistemic model, but had nothing to say about the content of such judgments. In other words, we have not addressed Smith's ethics, that is, the particular conclusions or general rules that might emerge from his epistemic exercise. Indeed, it is striking 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 surely related to the fact that his theory is grounded in moral sentiments, rather than grand moral principles, but I believe it also reflects his primary aim of analyzing moral judgment itself rather than fleshing out its content. Nonetheless, a discussion of Smith's ethics will aid in establishing this claim and in laying the groundwork for arguments that follow in this paper.

I begin with the broadest question of whether Smith views his epistemic model as leading to general conclusions and, if so, to what extent. It seems particularly necessary to address this issue in light of the claims of some commentators that Smith is a moral relativist (see Rasmussen, 2008, for a clear and concise discussion). This interpretation follows from the derivation of Smith's spectator from the values of society coupled with the claim that moral values are heterogeneous across societies. I believe the answer to the first part of this question regarding the existence of moral rules is unambiguous: Smith frequently refers in so many words to the virtues, rules, principles and measures of conduct that emerge from the aforementioned socialization, including "the most sacred rules of morality" (II.ii.3.8). Clearly, Smith maintains that moral rules accompany moral learning and moral knowledge.

The second part of the question concerns the level of generality of such rules. Smith's system might still be accused of moral relativism, if there exist rules but they are *context specific*, e.g., moral rules hold within societies but differ across them. As Weinstein (2006) discusses, related questions arise about reconciling differences within sub-groups of society. Smith, in fact, concedes that even slavery and infanticide have, at various times and places, been regarded as acceptable. By designing the spectator as the correspondence of society, Smith appears vulnerable on this point. But the reader should note that my question above was not whether his

moral theory does necessarily lead to general rules, but rather whether Smith sees general conclusions as emerging from his epistemic model. In fact, Smith resounding condemns practices such as slavery and infanticide and sees them as aberrations on a path otherwise characterized by moral progress. In such cases, Smith demurs and subordinates the supposed spectator to a higher authority, as discussed in Section V of this paper.

Despite occasional ambiguity, the weight of Smith's writing suggests a universalist approach to ethics. Sen (2009), for example, concludes that Smith's device is one of "open impartiality," which incorporates the moral views of other groups and societies, as opposed to "closed impartiality," which only draws on members of a given group or society (e.g., Rawls). Smith does view the relative weight assigned to individual virtues as varying across cultures but rules out substantive differences in those virtues: "Though the manners of different nations require different degrees of the same quality, in the character which they think worthy of esteem, yet the worst that can be said to happen even here, is that the duties of one virtue are sometimes extended so as to encroach a little upon the precincts of some other" (V.2.13).

Thus far, the emphasis on general conclusions has been on the existence and extent of moral rules. But one can think of a different type of generality that pertains not to rules but to judgments about specific cases. For instance, suppose each case were judged the same by all persons in all societies, even if their judgments could not be characterized otherwise according to any rules. Then moral judgments would be universal even if all they were all sui generis. In the absence of complete consensus, the degree of universalism might be thought of as the extent of moral agreement about some set of specific cases. Alternately, one can think of a world characterized by a combination of rules and exceptions. Universalism could be then conceptualized in terms of the extent of agreement about moral rules and moral exceptions. In fact, Smith's words often suggest not only universalism but a version that stresses the broad application of moral rules to particular cases: "When these general rules … are universally acknowledged and established, … we frequently appeal to them as the standard of judgment, in debating concerning the degree of praise and blame that is due to certain actions of a complicated and dubious nature" (III.4.11). One benefit of and justification for such rules is as a means to cope with self-deception, as Fleischacker (2009) points out. For Smith, "self-deceit … is the source of half the disorders of human life" (II.4.6). In his divided self model, this involves the agent deceiving itself about merit of its conduct when that conduct departs from the dictates of

the spectator. 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).

On the other hand, Smith also emphasizes the importance of particular moral judgments and exceptions to rules. He acknowledges that the "general rule(s) of almost all of the virtues … admit of many exceptions, and require so many modifications, that it is scarce possible to regulate our conduct entirely by a regard to them," and even the most precise rule will "upon the most superficial examination … appear to be in the highest degree loose and inaccurate, and to admit of ten thousand exceptions" (III.6.9). These statements follow from two claims. First, moral rules describe necessary but not sufficient conditions for right conduct and fail to incorporate, for example, the virtues of grace and mercy (III.6.1-5). Second, these rules do not provide an exhaustive moral guide because of the many interpretative decisions that must precede action.

Although I find these claims substantively persuasive, I do not believe that they necessarily compromise the generality of moral rules but rather might only reflect Smith's particular usage of the terms "rule" and "exception." On the first claim, Smith equates rules with moral norms, which connote rules that are socially sanctioned. But a more inclusive characterization of moral rules might integrate non-sanctioned virtues such as grace and mercy. On the second claim, I argue that the question is not one of exceptions to rules but rather of the practical challenges to interpreting and applying general rules to particular cases. That is, the approach Smith describes is not *context-specific*, since it spans different contexts (e.g., peoples and cultures), but it is *context-dependent*, i.e., the rules are general and apply across contexts but their practical implementation is sometimes subject to vagaries due to the existence of alternative interpretations of their application, whereby these interpretations depend on the context. As an example, Smith discusses the necessity of expressing measured gratitude for a benefactor's services but the challenges to determining the appropriate amount, metric, timing and duration of that expression given contextual factors, such as the individual circumstances of the affected parties (III.6.9).

Thus, I conclude that Smith is a universalist: he believes that his epistemic model leads to general moral conclusions. How much more detailed can one be about the content of moral conclusions? For this purpose, I believe it is helpful to distinguish Smith's *personal ethics*, or his

treatment of morality in personal relationships, from his *social ethics*, or his normative statements about more impersonal social interactions, such as with firms, industries and governments. His statements about personal ethics are mostly uncontroversial and include admonitions against deceit, murder, theft and adultery. Thus, I will focus on Smith's social ethics, in particular, on his discussion of justice, the aspect of social ethics that has received the largest share of attention by both Smith and his commentators. Indeed, Witztum (1997) argues that "Smith does not have a theory of justice which is clearly distinguishable from his general ethics."

The epistemic model for justice is, for Smith, the same as with his personal ethics: justice is grounded in the approval and disapproval of the impartial spectator (see, for example, II.2.3. and VII.2.1). Beyond that, however, the reader is struck by the negative tone and limited scope of most of Smith's writings about justice. The rules of justice are precise and exacting and are often associated with juridical contexts. They concern the potential for harm to others, often involving the violation of property rights, and the appropriate punishment for that harm. In fact, Smith's emphasis on commutative justice has contributed to doubts about whether he has a distinct theory of distributive justice. Verburg (2000), however, argues that, in Smith's view, commutative justice, which involves the observation of property rights, typically complements the aims of distributive justice. Smith sees the observation of rights in commercial society as promoting economic growth, which ultimately best serves to raise the least advantaged in society, viz., the workers, from poverty to subsistence. Nevertheless, Verburg concludes that Smith came increasingly to doubt the path of progress toward this goal and began to conceptualize justice in a broader sense.

Smith explicitly relates justice to the satisfaction of basic needs and, in turn, to the happiness of society: "No society can surely be flourishing and happy, of which the far greater part of the members are poor and miserable. It is but equity, besides, that they who feed, clothe, and lodge the whole body of the people, should have such a share of the produce of their own labour as to be themselves tolerably well fed, clothed, and lodged" (WN, Book I, Chapter 8). In the "other" invisible hand passage in TMS, Smith writes that the rich "are led by an invisible hand to make nearly the same distribution of the necessaries of life, which would have been made, had the earth been divided into equal portions among all its inhabitants" such that "in what constitutes the real happiness of life, (the poor) are in no respect inferior to those who would

seem so much above them" (IV.1.10). Indeed, Rasmussen (2006) argues that Smith sees the commercial society that drives economic progress as undermining happiness but, nevertheless, advocates for it because this effect is more than offset by the disadvantages of overcoming poverty.

On basic needs and happiness, Smith foretells the conclusions of one school of thought in the modern social science research on happiness. The "absolutist" school argues that, at low levels of income, improving materials standards help satisfy basic needs and increase happiness, but once basic needs are fulfilled additional income has little or no effect on happiness (e.g., Veenhoven, 1991). The "relativist" school argues instead that it is relative, not absolute, income that matters for happiness (e.g., Easterlin, 1995). Either conclusion, however, presents important challenges to welfare economics, which Fleurbaey, Schokkaert and Decancq (2009) discuss and attempt to address.

In summary, Smith espouses a universalist view of ethics with a prominent role for moral rules, partly in order to solve problems of self-deception. I have characterized Smith's approach as context dependent: right conduct obeys general rules that require interpretation in particular contexts. He derives both his personal and social ethics from the same epistemic model, i.e., from the impartial spectator. His social ethics stress justice, whereby distributive justice is associated with the satisfaction of basic needs and, in turn, happiness, representing a precursor of important modern social science research.

IV. Ein anderes Adam Smith Problem

In this section, I consider possible limitations and shortcomings of Smith's epistemic model. I begin with an analysis of problems associated with using this model, which was developed for personal ethics, to acquire knowledge of social ethics. Then I address more general limitations of this model, even as applied to personal ethics.

The picture of moral motivation and its wide-ranging impact on behavior that is presented and praised in fine detail in TMS contrasts with the portrayal in WN of the benefits of self-interest. Moreover, the ostensibly conflicting accounts prompted attempts to reconcile them, which became known as "das Adam Smith Problem," after its origins with the German economist Bruno Hildebrand in the mid-nineteenth Century. Vivienne Brown (2008) nicely summarizes the current consensus that there is, in fact, no inconsistency between the two works.

Raphael and Macfie (1976) present one line of thought leading to this conclusion: Smith holds both self-interest and virtue to be two important motives in both works, although TMS focuses more on the former and WN on the latter. This view garners support from a consideration of the differing social contexts that dominate each book. The specifics in TMS deal disproportionately with personal, small-scale interactions, where moral norms are more prominent and moral reasoning more likely to be activated. The subject matter emphasized in WN, by contrast, is larger scale and less personal market and social interactions, such as with firms, industries, and governments, where self-interest will more likely arise as the primary motive that affects outcomes. Of course, there are exceptions, e.g., Brown (1994) analyzes the expanded discussion of statesmen and legislators, which emerged when Part VI was added with ed. 6 of TMS. Nevertheless, these stylized facts about personal vs. impersonal domains arguably account for the much of the differing emphasis on motives in the two works.

As previously noted, TMS contains both descriptive and prescriptive statements, and the same is true of WN. In his first work, though, Smith transitions almost seamlessly between the two modes, whereas there is a more pronounced change in tone about mid-way in WN from empirical observation and analysis to polemic. The fact that Smith advocates for morality in TMS and free markets in WN merely serves to accentuate the appearance of conflict. The model of agent and spectator outlined in TMS can remove concerns about contradictory assumptions of motives. But a potential conflict, or rather gap, still remains: TMS provides a derivation for the virtues it espouses, whereas WN implicitly assumes the normative value of the outcomes promoted by self-interest.

Let me put this more broadly in terms of the epistemic issues that are the topic of this paper. TMS presents a clear, if not entirely uncontroversial, statement of the conditions for the acquisition of moral knowledge. WN, on the other hand, does not provide an analogous basis for desiring the ends it advocates, despite elegant arguments about how the pursuit of self-interest can promote economic growth and justice. This differs from the original Adam Smith Problem and resembles more recent "Adam Smith problems," like the one Brown (2008) identifies, which represent differences rather than inconsistencies between TMS and WN. But it does seem an odd asymmetry that the two works both describe and espouse but that only the first develops a distinct epistemic foundation for its advocacy. Certainly, the impartial spectator theory of TMS plays no explicit role in WN. The word spectator is used only twice in WN, and not in any

clearly moral sense, but rather in the sense of an actual third party. David Levy devotes a chapter of his 2001 book, "How the Dismal Science Got Its Name," to the claim that the actual spectator plays a central role in WN, but it seems very difficult to support the claim that impartial spectator theory forms the normative foundation for that work.

Of course, this incongruity might simply be yet another example of how Smith's work presaged later developments, in this case, methodological differences across disciplines. Whereas the acquisition and content of moral knowledge falls squarely in the domain of moral philosophy, the mainstream in economics tends to maintain its distance from moral values and, even more, from reflection on how to acquire knowledge about them. Although growing numbers of economists take interest in these questions, traditionally most have held a range of alternate or fairly narrow views about morality. These include a Weberian belief in value free science, skepticism about the sincerity of moral motivation, and the implicit or explicit endorsement of efficiency alone, i.e., the maximization of economic value as the single goal. This final view has dominated most economics, and Smith was likely wearing his economist's hat when advocating for free markets. But, as just noted, Smith did not, in fact, shun normative statements in WN. Moreover, the problem raised here does not concern the differences in such statements across the two works but rather the absence of an epistemic model for them in WN.

One might counter that, although moral judgments are not derived from an epistemic model in WN, many of the values promoted in WN, such as justice, were developed from the impartial spectator model in TMS. This point is quite correct and helps to clarify the precise problem I wish to address. The crux of the problem is the lack of a thoroughly developed and distinct epistemic model for social ethics in either TMS or WN. Alternately, it is the failure to provide a compelling case that the epistemic model for personal ethics so thoroughly developed in TMS logically extends to social ethics. Despite its discussion of social ethics, the focus of TMS, particularly on matters of the acquisition of moral knowledge, is on personal relationships. The impartial spectator is born and nurtured in repeated, personal contacts. In sociology, people are often thought to make moral distinctions between their personal and impersonal relationships, which Tönnies (1887) called Gemeinschaft and Gesellschaft, respectively. Moreover, as discussed below, empirical evidence suggests that these relationships differ not only in the strength of the role for self-interest but sometimes also qualitatively in terms of the moral rules that are activated (Konow, Saijo and Akai, 2009). Thus, Smith commits the fallacy

of composition in seeking to apply his epistemic model for personal ethics to social ethics.

At any rate, I raise this "problem" not because I think it is such an inherently interesting puzzle in the history of philosophy. Rather, I maintain that the subject matter of this gap is of great philosophical, social and economic significance. But I also argue that Smith's own writings on moral philosophy contain the elements for a promising approach to filling this gap. The question of how we identify right conduct, or the rules that govern right conduct, in impersonal relationships, such as in markets or in relations between countries, is a terribly important one for both philosophy and the social sciences. Rawls (1971), for instance, does not disregard it when he addresses the distribution of broadly defined goods at the social level with a clearly specified thought experiment in the social contract tradition. Social scientists are called upon to recommend and design policies that have wide-ranging impacts on society, including on income and wealth distribution, education, poverty, health care and the environment. But the optimal recommendations depend on the presumed moral values, which themselves must be ascertained and justified according to some criteria.

In addition to doubts about the validity of extending the impartial spectator model to social ethics, this model is vulnerable on several counts as a basis for personal ethics (making it, a fortiori, a dubious foundation for social ethics). So, I will now attempt to punch a few holes in impartial spectator theory as such. I will also argue, however, that Smith expressed an awareness of vulnerabilities of his theory. Indeed, although his 18th century station denied him access to modern social science tools, he directs us towards the elements of such a solution. I will structure this discussion around the three previously discussed properties of the imagined spectator. I begin with sympathy, which, as interpreted here, strikes me as the least controversial and vulnerable component of the theory. Then we will turn to impartiality and information, respectively, which relate more closely to the main topic of this paper.

Although Smith's sympathy is often interpreted in terms of feelings alone, I have argued here that Smith's tone and words imply sympathy is also a cognitive exercise. This interpretation finds support, I believe, based both on a priori reasoning and empirical findings. Cognitive sympathy enables the spectator to enlarge his set of relevant information and to balance competing interests and perspectives. Affective sympathy operates chiefly on the agent to provide motivation to right action and to provide direction when the spectator is not fully engaged or in control, as a kind of auto-pilot. In addition, if sympathy is to help provide a true

moral compass, empirical observation suggests it must carefully balance the two mental processes. For example, drawing on wide ranging research in psychology and economics, Loewenstein and Small (2007) argue that generosity is a function of both affective and cognitive mental processes and that either motive alone is insufficient to generate wise decisions in the area of generosity. In particular, they state that affectively motivated generosity alone is erratic and immature, responding to irrelevant context specific factors.

Regarding impartiality, there are both philosophical and empirical grounds for questioning the objectivity of Smith's imagined spectator. This model, as we have discussed, involves one looking inward not to one's feelings but to one's feelings if one were an imagined other person. Note that this is a kind of second order introspection, but introspection, just the same. Eric Schwitzgebel (2008), in a recent contribution to *Philosophical Review*, 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. Consistent with cognitive dissonance theory (e.g., Festinger, 1957), people are prone to engage in self-deception, altering their beliefs about what is right in the direction of their own selfish interests, which Smith termed self-deceit and is, in modern scholarship, often called a *self-serving bias*. Konow (2000) finds, for example, that even under sterile laboratory conditions that are least conducive to self-deception, almost two-thirds of unfairness is still due to such a bias. Moreover, Babcock and Loewenstein (1997) describe experimental and field studies demonstrating that the self-serving bias significantly impacts behavior, impeding agreements and promoting impasse in bargaining, such as with contract negotiations and in civil litigation. They also report that this bias is very tenacious, as various experimental attempts to dislodge it demonstrate. 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 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. Thus, agents, the progenitors of the spectator, are inclined to provide the latter with information that exhibits a self-serving bias.

In addition, much empirical research on decisions of groups rather than individuals indicates the possibility of more balanced processes and outcomes with groups. An example of this is research on deliberation, such as that reviewed in the Elster (1998) volume, a research agenda originally inspired by Habermasian (1983) discourse theory. This also emerges in many studies of procedural justice. The experimental study of Sulkin and Simon (2001) indicates that opportunities to share information and opinions produce both fairer outcomes and more favorable perceptions of the outcome itself. Even without these empirical findings, there are a priori reasons for being suspicious of the informational content behind single actor judgments. To take a simple example, and perhaps an all too frequent one for some of us, suppose I wish to maximize the satisfaction my Aunt Olga takes from my birthday gift to her (and it is also right that it should do so). Neither I nor any imagined version of myself knows her tastes and what would please her. Certainly, I could do more good with informational input from others, perhaps her friends. Of course, there is the well known criticism of spectator theories that their informational conditions can never realistically be satisfied. As Zagzebski (2004) points out, no spectator can be omniscient. But that is not the point toward which I am driving: I am not criticizing spectator theory because it lacks perfection, but rather the claim being developed is that, both on a priori grounds and from empirical evidence, the informational conditions of the impartial spectator can be improved upon, i.e., that they are dominated by alternative approaches.

In this section, I have argued that Smith's impartial spectator theory is missing a distinct epistemic foundation for social ethics. In addition, philosophical and social science research raises questions about the spectator approach even for personal ethics and casts doubts on the impartiality and information assumptions of his spectator. The following section makes a suggestion for addressing these questions, justifying it based on empirical evidence and motivating it with elements in Smith's own writings.

V. From Thought Experiment to Formal Experiment

As discussed, Smith's spectator is a mental image. Another type of spectator Smith mentions is a real spectator, i.e., a genuine third party who observes and judges the conduct of others. Just as the impartial spectator is an idealized version of the imagination, let us take this a step further and construct an idealized version of the real spectator that contains the three main properties of the imagined spectator. This person has no personal stake whatsoever, whether material or reputational, for example. This fact guarantees that self-interest cannot insinuate itself into his judgment and cause a self-serving bias. In addition, this person has the greatest feasible access to information that might be relevant to judging conduct and character. Given the absence of personal stakes, the process of sorting relevant information from irrelevant information is not sullied. Sympathy motivates the real spectator's desire to commit the resources to the optimal acquisition of information, including information about the interests and feelings of all affected parties, and to expending the appropriate mental effort to process the facts with respect to spectator's moral sense. Although ideal in the sense just stated, this spectator is not necessarily omniscient or infallible in his judgments, because the context and reigning moral standards might not require or even permit as a possibility the acquisition of all information. That is, this spectator might himself be subject to moral standards that balance the expected moral benefit of incremental information against the expected marginal cost in resources of acquiring that information. Nevertheless, this spectator could not be improved upon in light of these considerations.

To be sure, the real spectator is not as prominent feature of 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).

On one occasion, 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 that we discussed in the section 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, from whom we can expect the least sympathy and indulgence, 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.

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). Smith writes at great length in TMS on the natural harmony between God (or 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: "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 sometimes 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.

So far, I have still not proposed how one might fill the gap discussed in the previous section. Before beginning, however, let me be unequivocal in trying to anticipate and address two criticisms that might arise in this connection. First, the foregoing discussion should not be construed as an attempt to attribute this proposal to Smith. I have not consulted a Ouija board that provides me special access to his ghost. Instead, the intent above was to show that Smith was aware of some of the potential shortcomings of the imagined spectator and entertained elements of a possible solution, which also inspire this proposal. Second, it is not a goal of this paper to provide a philosophical justification for the proposal. That is a much larger undertaking for another project. Instead, the goal is to demonstrate that, if one buys into the core philosophical and empirical claims in TMS, then this proposal represents a reasonable epistemic model for acquiring knowledge of personal and social ethics, which employs modern social science methods to improve on the original impartial spectator model. Thus, the argument proceeds from Smith's premises, including the universalist view of a common moral sense and the position that the judgments of the imagined and real spectators are mostly useful but also fallible in the previously noted ways. I seek to invoke the weakest possible assumptions, however, and do not base this approach on any further assumptions. Specifically, it is agnostic regarding the origin of the moral sense and whether it is formed by society, evolutionary biology or a deity; we require only that it be commonly shared by and accessible to real people, albeit with error.

This suggestion is, in a sense, a modification and extension, in the spirit of empiricism, of Smith's original epistemic model. The impartial spectator that guides Smith's personal ethics is the product of a thought experiment in which the agent imagines an ideal moral judge. The current proposal borrows elements of this theory and outlines a so-called "quasi-spectator" who is a real person, or rather a set of real persons, who approximates the ideal real spectator. This quasi-spectator describes an empirical method that is employed to induce the moral sense, i.e., the claim is that the quasi-spectator has value precisely because of the ability to draw on the moral sense of the imagined spectator. Thus, it extends Smith's thought experiment and opens the door to a formal experiment. I argue below that a recent and growing empirical research

agenda along these lines has already contributed not only to descriptive analysis but also to social ethics, at least as practiced in the social sciences.

All people are, at different times and to varying degrees, spectators, that is, they observe and morally judge situations in which they are not, in any significant sense, stakeholders. It is dubious, however, that real people ever actually achieve the state of pure objectivity and optimal information of the ideal real spectator. Two questions follow. First, can we approximate the conditions of such a spectator? The quasi-spectator method seeks to answer this by eliciting empirically the moral judgments of individuals about situations in which their real or imagined stakes are minimal and about which they are amply informed. On the face of it, the answer to this question is clearly affirmative: social scientists have studied moral judgments for decades using a variety of methods, including surveys, laboratory experiments and field data, and some of this research draws on studies of quasi-spectators. The second question is whether we can, using these methods, identify with some degree of certainty the views of the ideal real spectator. That is, what evidence is there that these findings shed light on the moral sense of ideal spectators? This second question is really the critical one to defending on normative grounds why one should care about the results of such studies.

A first step in answering this is to acknowledge that we might, at best, observe impartial spectator judgments with error. On the one hand, there are the shortcomings inherent to spectators, real or imagined: information is incomplete, cognitive abilities are limited, people might project their own stakes on those they are judging, and even views of otherwise objective persons can be affected by unrepresentative experiences and beliefs. On the other hand, error is inherent to all empirical research, including errors in measurement, data entry, sampling, etc. Empirical science approaches these problems by collecting multiple data points and, when possible, controlling for such effects, and, when not, treating them as random. The quasi-spectator method, therefore, not only strives for conditions of impartiality and high information but also involves sampling numerous real spectators rather than just one in isolation. The earlier discussion of deliberation highlighted the possibilities for more informed and less biased decisions. But skeptics of this deliberation argue that deliberative groups, such as juries, are often prone through a group dynamic toward extreme views, which Sunstein (2002) calls *group polarization*. Be that as it may, deliberation concerns *juries*, whereas the quasi-spectator method involves *judges*. That is, the former involves groups whose members communicate with one

another and reason jointly, whereas the latter consists of informed individuals who proffer moral judgments individually and independently and are not, therefore, subject to group dynamics of deliberation. There are potential strengths and weaknesses of both methods, but group polarization cannot be a shortcoming of a spectator approach.

The social science research on moral values is copious, and much involves elicitation of impartial and/or informed views. For example, this can be found in public opinion research in political science, attitude surveys in psychology and vignette studies in sociology. To my knowledge, however, these studies have not been consciously designed with an eye toward testing our second question concerning evidence on the quasi-spectator. For this reason, and in the interests of brevity, therefore, I will focus on several studies that were designed with this specific intent.

First, minimal evidence of impartiality is that the judgments of quasi-spectators should differ from those of agents given the bias of the latter. This has been confirmed in a dictator experiment (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 divided a sum of money between two anonymous subjects. The third party, or spectator, was paid a fixed fee for this decision unrelated to her decision. The dictators in the standard stakeholder experiment demonstrated a bias, taking for themselves, on average, more than third parties in the spectator treatment gave similarly situated subjects in their treatment. Stronger evidence on the impartiality of quasi-spectators comes from a subsequent vignette study of fairness (Konow, 2009b). There stakeholders have different and opposing interests. Since each is biased in his own (opposite) direction and away from the right choice, the spectator choices should be a mean of the opposing stakeholder choices. 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. Stakeholders are expected to exhibit high variance in their judgments 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. Spectators, on the other hand, should value only the morally right choice; given their common moral sense, there should, therefore, be a much higher level of consensus in their views. These propositions are confirmed in studies of

justice (Cappelen, Konow, Sørensen and Tungodden, 2009, Konow, 2000, Konow, 2009a, and Konow, Saijo and Akai, 2009) and in a study that includes reciprocal preferences (Croson and Konow, 2009).

A potential criticism of the empirical application of spectator theory is that it essentially puts morality up for a vote. One might expect this to lead to some capricious and even objectionable conclusions. Consider, for example, Smith's references to the acceptability of infanticide or slavery in some societies. This view rightly cautions us to avoid over-simplified interpretations of or naïve implementations based on this method. There are several parts to my response.

First, the mere fact that people sometimes accept certain practices does not necessarily imply that they also morally condone them. There is empirical evidence that the views of people about what is right sometime differ from what they accept or even prefer (e.g., Ordóñez and Mellers, 1993). Thus, Smith declares even infanticide as justified in the early development of societies but as abhorrent and indefensible in the later stages, and tolerance of it beyond its time is a result of custom and an aberration that is overturned in civilized societies (V.2.15-16).

Second, some exemplars advanced in opposition to consideration of folk moral intuition are not really counter-examples but really "counter-thought experiments." That is, statements about the moral sense are often based on the speaker's own beliefs about that sense rather than actual findings about it. For example, empirical studies of fairness concerns find little support for the premises of Nozick or Rawls about moral intuition that form the foundations for their theories (see Konow, 2003). This fact reinforces, rather than diminishes, the importance of conducting empirical research on moral views.

Third, the power of these ostensible counter-examples derives precisely from and is in proportion to our shared moral intuition. Thus, their ability to persuade relates directly to their accuracy as a description of our actual values, a fact that itself suggests an empirical study. If our Smithian premise of a shared moral intuition is correct, then we should find that we either must revise our view of others' moral beliefs or admit that our own intuition was incorrect and the exemplars are not true counter-examples. Furthermore, we cannot rule out the value of empirical analysis, even if we abandon the assumption of a shared moral sense. A pluralist, for example, strengthens his or her case by demonstrating empirically that group A, in fact, has a different moral sense that group B. Alternately, if only a subset of society is seen as capable of meaningful

moral reflection, say philosophers, then a given philosopher's claim establishes its power to persuade by being able to demonstrate it is shared with other philosophers.

Fourth, just as introspection alone cannot lead to moral certitude, one cannot expect easy answers from an empirically informed method. If, however, our assumption of a shared moral sense is justified, then counter-intuitive findings might still result from a deficit in the particular attempt to access that sense rather than a failure of program as a whole. Indeed, as Weinstein (2006) points out, Smith viewed slavery as a departure from virtue and traced its existence to failings of pride, self-deceit and lack of sympathy, i.e., to failures of the agent to act on the judgment of the impartial spectator, rather than deficiencies of the spectator per se. As Sen (2009) writes, "The claim that people would agree on a particular proposition if they were to reason in an open and impartial way does not, of course, assume that people are already so engaged, or even that they are eager to be so" (pg. xix). Indeed, the quasi-spectator method proposes means for determining whether such residual biases have contaminated the results. Moreover, as in all intellectual endeavors, it is wise to apply not only inductive but deductive reasoning. Citing again Sen, "There is no irreducible conflict here, I argue, between reason and emotion, and there are very good reasons for making room for the relevance of emotions" (pg. xvii). Thus, it is the work of both moral philosophers and social scientists studying ethics to consider carefully derived empirical findings about the moral sense, but also to apply reasoning in interpreting and drawing conclusions from such results.

Finally, I conclude this section with some findings about the potential value of studying both stakeholders and (quasi-)spectators empirically that has implications for descriptive and prescriptive analysis. A dictator experiment conducted in the US and Japan (Konow, Saijo and Akai, 2009) reveals that spectators and stakeholders apply distinct rules of justice. Spectators in both countries allocate very close to the equity rule that requires rewarding others in proportion to their merit. Stakeholders, of course, are biased toward their own interests, but to the extent they deviate from self-interest and allocate fairly, they are significantly more likely to divide their earnings equally rather than proportionately. The level of the personal relationship between stakeholders is varied across treatments, e.g., in one treatment stakeholders are anonymous and in another they know one another, and we find the more personal the relationship, the more equally they allocate. One interpretation of this is the following: spectators are following a moral principle, i.e., a rule that applies in impersonal relationships, such as markets, whereas

stakeholders, who are by definition in a personal relationship, are affected not only by equity but also by a moral norm of equality. That is, one can distinguish moral principles, which apply generally and are the sole forces in more impersonal relationships, such as work, markets and government, from moral norms that are unique to personal relationships, such as friends, family and acquaintances.[2] An interesting implication of this conjecture is that moral rules differ based on personal versus social ethics, a conclusion that was not obvious when beginning this project of going from the impartial spectator theory thought experiment to a quasi-spectator laboratory experiment. Moreover, this finding corroborates the necessity of distinguishing personal ethics from social ethics. A further implication is that there is sometimes value in adjusting the empirical method to the type of ethical question, personal or social, being targeted.

VI. Conclusion

Casual empiricism suggests that spectatorship is not merely a philosophical ideal or a valid empirical method but also a standard for impartiality valued by society at large. We can recognize its contours in a variety of social institutions and practices. In the justice system, the assignment of judges and juries and the rules of evidence are designed to approach third party impartiality while liberally providing relevant information. Politicians are subject to conflict of interest provisions, and an important impetus to campaign reform is to sever legislators from stakes that might bias how they represent the interests of their constituencies. Although its efficacy is debatable, government regulation, at least as an ideal, provides informed, third party oversight of industries. Sometimes even private organizations attempt to bolster their impartiality credentials by supporting informed (ostensibly independent) verification of their practices, such as the Better Business Bureau or ClimateCounts.org, a privately supported organization that reports the performance of businesses to limit their greenhouse gas emissions.

This paper has focused on spectatorship as a means for acquiring moral knowledge and has underscored its importance for philosophical reflection and for social science research and social policy. We reviewed Adam Smith's contribution to designing a model for the acquisition of moral knowledge, arguing that ideal moral judgment, in his view, is passed by an imagined

[2] A related example can be found in the work of the social psychologist Mikula (1980), who found a generosity bias: in choosing between equity or equality, his dictator subjects, who consisted of soldiers in the same unit, chose the rule that worked to their disadvantage, i.e., high performers chose equality and low performers chose proportionality. That is, the personal relationships of these stakeholders prompted more generous behavior.

spectator who is impartial, informed and sympathetic. I considered problems with extending his theory from its original domain of personal ethics to social ethics, as well as some vulnerabilities of the theory, even within its original context. Then I proposed an epistemic model that draws on Smith's own insights and modern empirical social science methods. The proposed quasi-spectator method analyzes the moral views of real spectators who approximate the properties of impartiality, information and sympathy. Recent empirical studies provide promising results on this method.

REFERENCES

Ashraf, Nava, Camerer, Colin, and Loewenstein, George (2005). "Adam Smith, Behavioral Economist." *Journal of Economic Perspectives* 11 (Summer): 109-26.

Babcock, Linda, and Loewenstein, George (1997). "Explaining Bargaining Impasse: The Role of Self-Serving Biases." *Journal of Economic Perspectives* 11 (Winter): 109-26.

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

Brown, Vivienne (2008). "Agency and Discourse: Revisiting the Adam Smith Problem," discussion paper 72, The Open University.

Cappelen, Alexander, Konow, James, Sørensen, Erik, and Tungodden, Bertil (2009). "Just Luck: An Experimental Study of Fairness and Risk Taking." Manuscript.

Croson, Rachael, and Konow, James (2009). "Social Preferences and Moral Biases." *Journal of Economic Behavior and Organization* vol. 69 (3): 201-212.

Easterlin, Richard A. (1995). "Will Raising the Incomes of All Increase the Happiness of All?." *Journal of Economic Behavior and Organization* vol. 27 (1): 35-47.

Elster, Jon (1998). Deliberative democracy. Cambridge University Press, Cambridge, Massachusetts.

Festinger, Leon (1957). *A Theory of Cognitive Dissonance*. Stanford: Stanford University Press.

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

Fleischacker, Samuel (2004). *On Adam Smith's Wealth of Nations: A Philosophical Companion*. Princeton, New Jersey: Princeton University Press.

Fleischacker, Samuel (2009). "True to Ourselves? – Adam Smith on Self-Deceit," manuscript, Philosophy Department, University of Illinois, Chicago.

Fleurbaey, Marc, Schokkaert, Eric, and Decancq, Koen (2009). "What Good is Happiness?" ECORE Discussion paper 2009/38, Université catholique de Louvain.

Fricke, Christel, and Schütt, Hans-Peter (2005). *Adam Smith Als Moralphilosoph*. Berlin, New York: Walter de Gruyter, Inc.

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.

Haberrnas, Jürgen (1983)[1990]. *Moral Consciousness and Communicative Action.* Trans. Christian Lenhardt and Shierry Weber Nicholson. Cambridge, Mass: MIT press.

Harrison, John (1995). "Imagination and Aesthetics in Adam Smith's Epistemology and Moral Philosophy." *Contributions to Political Economy* 14: 91-112.

Kawall, John G. (2006). "On Moral Epistemology of Ideal Observer Theories." *Ethical Theory and Moral Practice* vol. 9 (3): 359-374.

Konow, James (2000). "Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions." *American Economic Review* 90 (September): 1072-92.

Konow James (2003). "Which is the fairest one of all?: a positive analysis of justice theories." *Journal of Economic Literature* 41(4): 1186-1237.

Konow, James (2009a). Is Fairness in the Eye of the Beholder?: An Impartial Spectator Analysis of Justice." *Social Choice and Welfare*, 33(1): 101-127.

Konow, James (2009b). "The Moral High Ground: An Experimental Study of Spectator Impartiality." Manuscript.

Konow, James, Saijo, Tatsuyoshi, and Akai, Kenju (2009). "Morals Versus Mores: Experimental Evidence on Equity and Equality." Manuscript.

Loewenstein, George, and Deborah A. Small (2007). "The Scarecrow and the Tin Man: The Vicissitudes of Human Sympathy and Caring." *Review of General Psychology* 11.2 (June): 1-15.

Mikula, Gerold, Editor (1980). *Justice and Social Interaction: Experimental and Theoretical Contributions from Psychological Research*. New York: Springer-Verlag.

Ordóñez, Lisa D., and Barbara A. Mellers (1993). "Trade-Offs in Fairness and Preference Judgments." In *Psychological Perspectives on Justice*, edited by Barbara A. Mellers and Jonathan Baron. Cambridge, UK: Cambridge University Press, 138-54.

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

Rasmussen, Dennis C. (2006). "Does 'Bettering Our Condition' Really Make Us Better Off? Adam Smith on Progress and Happiness." *American Political Science Review* 100.3 (August): 309-18.

Rasmussen, Dennis 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, John (2000). *Lectures on the History of Moral Philosophy*. Cambridge, London: Harvard

University Press.

Rothschild, Emma, and Sen, Amartya (2006). “Adam Smith's Economics,” in *The Cambridge Companion to Adam Smith.* Ed. K. Haakonssen, Cambridge University Press, New York: 319-365.

Schwitzgebel, Eric (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)[1809]. *The Theory of Moral Sentiments*. Glascow: R. Chapman.

Sugden, Robert (2002). "Beyond Sympathy and Empathy: Adam Smith’s Concept of Fellow-Feeling." *Economics and Philosophy* 18.1 (April): 63-87.

Sulkin, Tracy, and Simon, Adam F. (2001). "Habermas in the Lab: A Study of Deliberation in an Experimental Setting." *Political Psychology* 22.4 (December): 809-26.

Sunstein, Cass R. (2002). "The Law of Group Polarization." *Journal of Political Philosophy* 10.2: 175-195.

Thaler, Richard H., and Shefrin, H.M. (1981). "An Economic Theory of Self-Control." *Journal of Political Economy* 89.21 (April): 392-406.

Verburg, Rudi (2000). “Adam Smith’s Growing Concern on the Issue of Distributive Justice” *European Journal of the History of Economic Thought* 7: 23-44.

Veenhoven, Ruut (1991). “Is Happiness Relative?” *Social Indicators Research* vol. 24: 1-34.

Weinstein, Jack Russell (2006). “Sympathy, Difference, and Education: Social Unity in the Work of Adam Smith,” *Economics and Philosophy* 22.1: 79-111.

Witztum, Amos (1997). “Distributive Considerations in Smith’s Conception of Economic Justice,” *Economics and Philosophy* 13.2: 241-259.

Zagzebski, Linda (2004). *Divine Motivation Theory*. New York: Cambridge University Press.