ADAM SMITH'S CONCEPT OF SELF-COMMAND AS A SOLUTION TO DYNAMIC INCONSISTENCY AND THE COMMITMENT PROBLEM

ELIAS L. KHALIL*

How is rationality related to morality and the emotions? In response to Hume, Smith argues that sympathy is about the attenuation, rather than the escalation, of original emotions because sympathy involves judgment. Sympathy means that the spectator understands an emotion felt by the principal by placing him or herself in the principal's shoes. Such understanding would not take place unless the principal's emotion is proper in that the principal has attenuated the pitch of the emotion via self-command, that is, via rational choice. Smith's notion of sympathy solves the commitment problem: agents command their emotions, which include the temptation to cheat their future selves and others, in order to receive approval. (*JEL* B12, D01, D64)

I. INTRODUCTION

In his famous letter of July 28, 1759 to Adam Smith, David Hume objects to Smith's conception of sympathy. Hume argues that Smith fails to explain apparently contradictory possible pathways of fellow feeling.[1] The first pathway, as Hume exposed in his early writings (Hume 1896; Sutrop 2007), shows that fellow feeling is contagious. As a result, fellow feeling escalates the original emotion. The second pathway, emphasized by Smith, shows that fellow feeling is noncontagious. As a result, fellow feeling mitigates or attenuates the original emotion. Hume therefore asks in his letter how "intimate Friendship" with someone in pain could *attenuate* the pain and even engender a counter movement of joy—when fellow feeling in "ordinary Cases," such as with an "ill-humord Fellow," *escalates* the original emotion and engenders "Damp on Company":

> I am told that you are preparing a new Edition [2nd edition of *Theory of Moral Sentiments*] and propose to make some Additions and Alterations, in order to obviate Objections. I shall use the Freedom to propose one, which, if it appears to be of any Weight, you may have in your Eye. I wish you had more particularly and fully prov'd, that all kinds of Sympathy are necessarily Agreeable. This is the Hinge of your System, and yet you only mention the Matter cursorily in p. 20. Now it wou'd appear that there is a disagreeable Sympathy, as well as an agreeable: And indeed, as the Sympathetic Passion is a reflex Image of the principal, it must partake of its Qualities, and be painful where that is so. Indeed, *when we converse with a man with whom we can entirely sympathize*, that is, where there is a warm and intimate Friendship, the cordial openness of such a Commerce overpowers the Pain of a disagreeable Sympathy, and renders the whole Movement agreeable. But in ordinary Cases, this cannot have place. An ill–humord Fellow; a man tir'd and disgusted with every thing, always *ennuié*; sickly, complaining, embarass'd; such a one throws an evident Damp on Company, which I suppose wou'd be accounted for by Sympathy; and yet is disagreeable. (Hume in Smith 1977, 43)

Put differently, Hume is asking how could fellow feeling with a parent in pain over the sickness of a child *mitigate* the parent's pain,

ABBREVIATIONS

MNS: Mirror-Neuron System
TMS: The Theory of Moral Sentiments
VCM: Voluntary Contribution Mechanism

*The article benefited from the support of the Konrad Lorenz Institute for Evolution and Cognition Research (Altenberg, Austria) and the Faculty Reseach Grants (2006, 2007) of Monash University. The article benefited from the comments of Aldo Rustichini, Stephen Miller, Harold Demsetz, Taiki Takahashi, Eric Schliesser, Leonidas Montes, Ian McDonald, James Buchanan, Robin E. Pope, and three anonymous referees. It also benefited from the assistance of Michael Dunstan. The usual caveat applies.

Khalil: Associate Professor, Department of Economics, Monash University, Clayton VIC 3800, Australia. Phone +61-3-9905-2407, Fax +761-3-9905-5476, E-mail elias.khalil@buseco.monash.edu.au

1. Eric Schliesser alerted me to Hume's letter to Smith. Levy and Peart (2004) and Sugden (2002) also discuss the letter.

Economic Inquiry
(ISSN 0095-2583)
Vol. 48, No. 1, January 2010, 177–191

doi:10.1111/j.1465-7295.2008.00182.x
Online Early publication November 28, 2008

while fellow feeling with people involved in mob activity *escalate* their emotion.

Smith was well aware of Hume's view of fellow feeling. In response to Hume's letter, Smith (1977, 48) claims that he answered Hume's objection and thinks that he "entirely discomfitted him." What is Smith's response? Smith argues that sympathy is about understanding and, as a result of Smith's definition of understanding, it inextricably entails moral judgment. Such judgment, by definition, mitigates, rather than escalates, the emotion of the principal—where the term "principal" is reserved for the person whose emotion/action is being judged. If so, moral judgment, which is totally based on the emotions, amounts to the exercise of self-command in that the agent restrains or commands his original emotion by far-sighted emotion. With this notion of sympathy, or its sister of self-command, Smith solves the problem of making rational decisions in the face of temptations, known also as the commitment problem or dynamic inconsistency: How can agents command their temptations and cooperate voluntarily when, in light of instantaneous reward, it is so enticing to cheat and indulge?

The purpose of this article is to examine Smith's solution and its ramifications. To appreciate Smith's solution, we need to understand Hume's objection. Section II therefore analyzes Hume's objection, called here the "fellow-feeling paradox." Section III then discusses Smith's response, which is the core of his theory of self-command. Section IV shows that Smith's notion of sympathy amounts to rational decision making that is used to solve the commitment problem (dynamic inconsistency). (The commitment problem concerns succumbing to weakness of will, which is defined as choosing an option that undermines the interest of future self to an extent judged as suboptimal.)

Section V examines some ramifications of Smith's response. One ramification is the distinguishing of self-control (i.e., precommitment) from self-command. While self-control is an extrinsic punishment adopted to solve the commitment problem, self-command is another method that does not rely totally on extrinsic punishment. Another ramification is that Smith's formulation allows us to view the commitment problem more broadly. The problem is not only about fighting the temptation to smoke or overeat, which would hurt the future self. It is also about justice, viz., fighting the temptation to cheat others, which would hurt the future self as in the case of the production of public goods. Smith's solution can therefore throw light on how to explain voluntary contribution mechanism (VCM) games (Isaac, McCue, and Plott 1985), as in public goods and Dictator games, which are contrary to the predictions of subgame perfect equilibrium.

A further ramification is related to the recent discovery of the mirror-neuron system (MNS), spearheaded by Rizzolatti. The discovery highlighted the role of understanding and, hence, has strengthened the case for Smith's notion of sympathy (Rizzolatti and Craighero 2004, 2005; Rustichini 2005).

Finally, Smith's notion of sympathy amounts to a bridging of rationality with the emotions. In *The Theory of Moral Sentiments* (TMS), Smith shows how understanding entails good reason and, in turn, good reason attenuates the emotions and prevents mob mentality, according to the principles of rationality. So, rationality, for Smith, is not borrowed from a set of abstract axioms or from mental constructs that are alien to the emotions or to morality. Rather, rationality and morality spring, amazingly, from the same emotions. The major contribution of TMS is thus the rejection of the dichotomy between affective processes (emotions) and cognitive processes (rational deliberation) that is usually called Cartesian dualism (Khalil 2009). Modern neuroscientists have just started to distance themselves from Cartesian dualism (e.g., Damasio 1994) and, hence, would also benefit from reading TMS.

Following Smith's rejection of Cartesian dualism, this article uses the terms "emotion," "passion," "choice," and "action" interchangeably. When one is angry, one is also choosing something or acting, even when anger does not lead to any bodily motion. As long as one expresses anger, even when expressed by silence, it is an action. Also, any body movement, such as hitting someone, is prompted by anger and such movement can be seen as a higher degree of anger than simple verbal expression. This means that Smith's discussion of emotions and sentiments is a discussion about behavior and vice versa.

II. HUME'S OBJECTION

It seems that key commentators have misunderstood Hume's objection in the letter

quoted in the introduction to this article above.[2] This appears to be because Hume uses the terms "agreeable" and "disagreeable" at *two* different levels of analysis. The first level concerns the principal's *original* emotion, which can be either agreeable or disagreeable. The second level, on the other hand, is about the principal's *derived* emotion, which can also be either agreeable or disagreeable. But at the second level, the derived emotion can take two different paths of fellow feeling—*attenuation* of the original emotion or its *escalation*. These two different paths seemed puzzling to Hume, which is the core of what is called here the "fellow-feeling paradox": How can the principal's derived emotion take either one of two different paths for no apparent reason?

To analyze the paradox, let us begin by examining the first level of analysis. The principal's *original* emotion can be either disagreeable or agreeable. The emotion is disagreeable when the principal experiences "Pain of a disagreeable Sympathy" such as, for example, when a father hears news concerning the sickness of his child. Conversely, the emotion is agreeable when the principal experiences joy, such as the case of a father upon hearing the recovery of his child. Therefore, according to this definition, in what Hume considers the "ordinary Cases," the emotion of an "ill–humord Fellow; a man tir'd and disgusted with every thing" would be disagreeable.

At second level of analysis, the principal's *derived* emotion can be either disagreeable or agreeable *as well*. However, the principal's derived emotion can take either one of two different paths for no apparent reason (at least for Hume). According to the first path, which consists of the "ordinary Cases," the principal's derived disagreeable or agreeable emotion *escalates* the principal's original (disagreeable or agreeable) emotion. The original emotion is disagreeable when, as described in the above paragraph, the principal acts as "ill–humord Fellow; a man tir'd and disgusted with every thing." Such a principal, who is in a bad mood, is a "Damp on Company": the principal annoys the spectator and makes him or her also ill humored and tired. The emotion of the spectator, in return, makes the principal even more ill humored and tired. The same escalation path takes place if the original emotion is agreeable, that is, when the principal is good humored and pleased with everything.

Hume regarded such escalation of the original emotion to be the "ordinary Case" because he defined sympathy as what can be called "sympathy-as-reflexivity": "[T]he Sympathetic Passion is a reflex Image of the principal, it must partake of its Qualities, and be painful where that is so."

According to the second path of the derived emotion, which will be labeled here the "non-ordinary case," however, the principal's derived disagreeable or agreeable emotion *attenuates* the principal's original (disagreeable or agreeable) emotion. The original emotion is again disagreeable when, for example, the principal is sad about the sickness of a child. Here, "intimate Friendship, the cordial openness of such a Commerce overpowers the Pain of a disagreeable Sympathy, and renders the whole Movement agreeable." That is, the fellow feeling of a spectator toward the principal in pain makes the principal feel a bit joyful.[3] Employing the same reasoning, which is actually used extensively by Smith, the same attenuation path takes place if the original emotion is agreeable, that is, when the principal feels overjoyed over the recovery of a child. The sympathy of a friend, as shown below in Smith's analysis, would force the principal to clip or lower the pitch of the original joy. That is, the derived emotion attenuates, rather than escalates, the original emotion.

Hume definitely thought that the attenuating character of emotion, which Smith

2. In Smith's letter to Gilbert Elliott on October 10, 1759, Smith claims that he answered Hume's objection and thinks that he "entirely discomfitted him" (Smith 1977, 48). In a note to the letter, the editors of *The Correspondence of Adam Smith* (Smith 1977), E. C. Mossner and I. S. Moss, state that Smith did actually answer Hume in paragraph "I.iii.1.9" of TMS (Smith 1976, 45–46). In a note to that paragraph in TMS, the editors, D. D. Raphael and A. L. Macfie, refer to the letter of Hume and to Smith's assertion to Elliott. However, the said paragraph discusses why people resist crying in a theatre when watching a tragedy. It is part of Smith's (1976, 43–50) chapter on what is called today "loss aversion," that is, the theory that grief that comes with tragedy is more acute than the rejoicing that comes with elation. But Hume's objection is not about loss aversion, as presumed by the editors.

3. This phenomenon of sympathy with someone in grief giving a sense of joy has also fascinated the mystic philosopher Stein (2002) in her analysis of how the suffering associated with the Christian cross affords a sense of joy.

stresses, must be either unimportant or inconsistent with the more usual escalatory character of emotion. In the following paragraph of the same letter to Smith discussed above, Hume sarcastically states that if fellow feeling with grief from tragedy attenuates the grief and gives rise to its opposite, that is, "agreeable" joy, hospitals would be more of entertaining places than ballrooms:

> It is always thought a difficult Problem to account for the Pleasure, receiv'd from the Tears and Grief and Sympathy of Tragedy; which wou'd not be the Case, if all Sympathy was agreeable. An Hospital wou'd be a more entertaining Place than a Ball. (Hume in Smith 1977, 43)

However, Hume confusingly uses the terms agreeable/disagreeable in two different senses. First, he uses terms in the usual sense: "agreeable" denotes joy and "disagreeable" denotes grief. Second, he uses the terms in the *pathway* sense: "agreeable" denotes the *attenuation* pathway of grief, which would be joyful and hence agreeable, and "disagreeable" denotes the *escalation* pathway of sadness, which would, in the cases when the original emotion is sadness, engender more sadness and hence be disagreeable.

III. SMITH'S RESPONSE

There is no direct evidence on how Smith understood Hume's objection. Nonetheless, as shown below, commencing with the sixth paragraph of TMS, Smith challenges the idea, without mentioning Hume, that sympathy in "ordinary Cases" is about reflexivity of emotions and, hence, escalation.

A few commentators have noted that Smith's concept of sympathy can easily explain the attenuation pathway of fellow feeling (Haakonssen 2002, xiv; Levy and Peart 2004, 334). However, such a concept poses a paradox for Hume as noted above (Fontaine 1997; Gordon 1995). To repeat, Hume defined sympathy as emotional reflexivity or contagion, which leads to escalation, usually facilitated by projecting one's emotion onto others.[4] This view of sympathy as contagion leading to escalation cannot handle the attenuation pathway.

4. Fontaine seems aware of the problem of defining sympathy as emotional contagion. In contrast, Gordon is uncritical of the definition.

For Smith, however, sympathy affords attenuation of original emotions. In fact, it can also lead to an emotion counter to the original emotion, such as the joy that arises from the sympathy with a parent who is distraught. Smith provides a single answer that remarkably explains both the attenuation of the original emotion and the counter movement that affords joy.

A. Attenuation of the Principal's Emotion

Starting with the sixth paragraph of TMS (Smith 1976, 11), Smith betrays his keen awareness of Hume's paradox or at least as reconstructed above. Smith defines "sympathy" as a *judgment* of an emotion vis-à-vis its cause, that is, the spectator is not simply imitating the emotions of the principal. The spectator might at first imitate the principal's emotion but only upon "some occasions":

> Upon some occasions sympathy may seem to arise merely from the view of a certain emotion in another person. The passions, upon some occasions, may seem to be transfused from one man to another, instantaneously, and antecedent to any knowledge of what excited them in the person principally concerned. Grief and joy, for example, strongly expressed in the look and gestures of any one, at once affect the spectator with some degree of a like painful or agreeable emotion. A smiling face is, to every body that sees it, a cheerful object; as a sorrowful countenance, on the other hand, is a melancholy one. (Smith 1976, 11)

However, Smith continues and argues, contrary to Hume, that some fellow feeling can only arise after the spectator has examined closely the cause behind the principal's passion. Thus, reflexivity does not hold universally or "with regard to every passion":

> This, however, does not hold universally, or with regard to every passion. There are some passions of which the expressions excite no sort of sympathy, but before we are acquainted with what gave occasion to them, serve rather to disgust and provoke us against them. The furious behaviour of an angry man is more likely to exasperate us against himself than against his enemies. As we are unacquainted with his provocation, we cannot bring his case home to ourselves, nor conceive any thing like the passions which it excites. But we plainly see what is the situation of those with whom he is angry, and to what violence they may be exposed from so enraged an adversary. We readily, therefore, sympathize with their fear or resentment, and are immediately disposed to take part against the man from whom they appear to be in so much danger. (Smith 1976, 11)

Smith's critical innovation is that sympathy is not about reflexivity. It is not about mirroring the principal's emotion, as Hume supposed, without any critical understanding of the cause or incentive behind the emotion. For Smith, sympathy rather arises from examining the emotion in relation to the incentive, which occasioned it: "Sympathy, therefore, does not arise so much from the view of the passion, as from that of the situation which excites it" (Smith 1976, 12). While the term "empathy" was not known in Smith's time,[5] sympathy differs from the modern sense of empathy. While empathy entails understanding of the emotion in relation to its cause without any judgment, sympathy denotes for Smith understanding of the emotion in relation to its cause *with judgment*. So, sympathy denotes understanding with the judgment that the emotion is justified given the cause. In reflexivity, however, we do not have judgment and, hence, it is closer to the modern sense of empathy. Reflexivity amounts to mimicking of the emotion—without examining whether the emotion is justified in relation to the cause of the emotion.

Smith makes clear this definition of sympathy in the second paragraph of TMS. For the spectator to sympathize, he or she cannot simply stay in his or her own station, and just imagine what the observed emotion would feel like, without actually relating to the station of the principal. Thus, the failure to "transport the self" to the station of the principal, and simply just fancy the observed emotion as happening to one's own station, amounts merely to self-centered sensation, while a successful transportation to the station of the principal amounts to sensing the emotion as actually experienced by the principal and to approve the emotion as justified.

That is, for the spectator to "occupy the shoes" of the principal, he or she must also judge that the emotion of the principal is efficient, that is, justified as proportional to the cause. This is Smith's central concept of sympathy, which is set distinctly apart from the modern meaning of the term "understanding" or "empathy." For Smith, understanding (empathy) necessarily *also* entails judgment of the propriety (i.e., efficiency or the propriety or proportionality of the emotion to the cause) of the emotion/action, while the modern concept of understanding (empathy) does not. Thus, to be clear, Smith's "sympathy" involves the modern notion of understanding/empathy *in conjunction with* judgment because, for Smith, empathy cannot take place without judgment.

But why does empathy for Smith (i.e., sympathy) entail the approbation of propriety? To state it tersely (Khalil 1990, 2005), the principal, in the heat of the moment, usually feels the "violent emotions" (Smith 1976, 157). Such violence has to be tuned down, via self-command, if the principal wants the spectator to sympathize with him or her. Self-command is related to propriety: while self-command is the action, propriety is the outcome. If the outcome is "improper," that is, the principal exhibits disproportional emotions, the spectator just simply cannot transport him or herself to the principal's station, that is, cannot sympathize. Given that the spectator does not directly experience the incentive, his or her excitement is muted, according to Smith. That is, the spectator is always remote from the immediate incentives and, hence, the spectator's passion is usually more circumspect than the principal's. So, for the spectator to experience the passion of the principal, the principal's passion has to be low—or as low as the spectator's passion. So, sympathy takes place when the principal exercises self-command or, in other words, exhibits behavior consistent with the optimal decision.

According to how this article interprets Smith's observation, the spectator's passion is usually more circumspect than the principal's because the spectator examines the situation from a "distance." The geographical distance is a metaphor for the spectator's station that allows the spectator to examine the cause facing the principal in relation to the interest of the principal's future self. As the spectator examines the cause, the spectator would not want the principal's current self to react in an excessive manner that undermines the interest of the principal's future self. When the spectator is less infuriated or excited by the cause, it is because the spectator is viewing the cause in relation to other concerns, namely, the interest of the principal's future self. Such

5. According to Gladstein (1984, 40), the term "empathy" was coined in 1909 as a translation of the German *einfühlung* [from the words *ein* ("in") and *fühlung* ("feeling")]. The German word was coined in 1858 by German philosopher Rudolf Lotze from the Greek *empatheia* ("passion"), which itself is derived from *en-* ("in") and *pathos* ("feeling").

distancing from the cause amounts to demanding propriety, which is the stipulation that the principal respects the rights of his or her future self. For instance, when the spectator demands a restraint on anger, the spectator is worried that anger could blind the principal and cause a rage that would hurt the principal's future self. The spectator—or, in other words, the nonmyopic principal—does not want the impetuous principal to indulge in the passion of the moment that might be detrimental. Given the spectator's prudent view, the principal must exercise self-command, that is, lower the pitch of the emotion, so that the spectator can enter the principal's station. This ability to live in the principal's shoes, for Smith, involves the same act of approving that the principal acted with propriety, that is, acted efficiently.

B. Another Source of Satisfaction

But why would the myopic self of the principal yield to the dictates of the spectator? Aside from the overall welfare of the principal, there is another reason why the spectator succeeds in nullifying the suboptimal strategy of the principal's myopic self: he or she can provide another "source of satisfaction" to the principal. This involves an emotion that accompanies sympathy à la Smith (involving judgment), which is absent in the case of sympathy à la Hume (involving reflexivity).

When the spectator understands, he or she approves the principal's emotion. Such sympathy affords "another source of satisfaction" to the principal, according to Smith (1976, 14). This additional source of emotion always has a positive effect—irrespective of whether the original emotion was grief over bad news or joy over good news—as it is the self-respect derived for acting rationally or, as Smith calls it, acting in propriety with regard to the original emotion—whether it is grief or joy. The principal—upon success in calibrating the action in proportion to the stimulus (incentive) irrespective of whether it is grief or joy—thus, in acting with propriety (i.e., acting optimally), is also infused with a sense of self-congratulation, accomplishment, integrity, or what can be called in general "symbolic utility" (Khalil 2000). This is considered a second source of emotion because it cannot exist independently of taking the proper or optimal action and hence is thus called "symbolic." Symbolic satisfaction arises also when one succeeds in exercising self-command over the appetite when one encounters a dessert tray while having a commitment not to indulge. Likewise, when one controls one's joy over good news, one derives utility from not celebrating in a careless fashion and also derives a sense of integrity for being so prudent. The same occurs when the original emotion is grief. When one controls one's grief over bad news, one derives utility from not giving in to anger and also derives a sense of integrity for being so prudent. While the success of resisting temptation affords a greater utility because (discounted) long-term health is preferred to momentary pleasure, it also affords the second source of satisfaction, namely, the symbolic effect.

Smith, in fact, directly criticizes Hume, the "ingeneious and agreeable author," for postulating that there is only one source of satisfaction, viz., utility. For Smith, Hume fails to recognize integrity, the self-satisfaction arising from acting with propriety that accompanies "approbation":

> The same ingenious and agreeable author who first explained why utility pleases, has been so struck with this view of things, as to resolve our whole approbation of virtue into a perception of this species of beauty which results from the appearance of utility. No qualities of the mind, he observes, are approved of as virtuous, but such as are useful or agreeable either to the person himself or to others; and no qualities are disapproved of as vicious but such as have a contrary tendency. And Nature, indeed, seems to have so happily adjusted our sentiments of approbation and disapprobation, to the conveniency both of the individual and of the society, that after the strictest examination it will be found, I believe, that this is universally the case. But still I affirm, that it is not the view of this utility or hurtfulness which is either the first or principal source of our approbation and disapprobation. These sentiments are no doubt enhanced and enlivened by the perception of the beauty or deformity which results from this utility or hurtfulness. But still, I say, they are originally and essentially different from this perception. [Smith 1976, 209 (TMS IV.2.3)]

That is, for Smith, the source of satisfaction related to approbation, what is called here the second source of emotion or the sense of integrity, is "originally and essentially different from this [utility] perception."

This second source of emotion, the sense of integrity that arises with approbation, goes to expose the "judgment function" of sympathy.

Sympathy as judgment can explain, first, the attenuation of the original emotion/action and second, via approbation, can account for the sense of integrity that is always positive—irrespective of whether the original emotion is joy or grief. So, sympathy as judgment with a grieving person leads to the attenuation of grief on two counts: first, through the lowering of the original pitch through self-command and, second, upon the success of lowering the original pitch, the feeling of self-satisfaction for being rational. (But, note, if the sympathy as judgment is with a joyful principal, it would attenuate the pitch of joy, on one hand, and lead to self-satisfaction for being optimal on the other.)

This interpretation should be distinguished from others. For example, Meardon and Ortmann (1996) reinterpret self-command as mainly arising from the agent's motive to uphold reputation. The concern over reputation can be seen as an extension of Smith's emphasis on the self-congratulatory effect. But, for Smith, such an effect is a secondary source that reinforces self-command: as already discussed, the main source or motive for action is the concern of the interest or welfare of principal—the welfare of the current and future selves.

Sugden (2002) also offers an interpretation that departs from Smith's project. Sugden confronts the issue of how a grieving parent could experience joy when he or she receives sympathy. Sugden does not see such joy as a second source of satisfaction that is based squarely on rationality. He instead reasons that such joy arises from an extrarational ground, namely, concern over moral rectitude that cannot be reduced to welfare calculation. The reason that Sugden, but not Smith, appeals to the extrarationality ground is that he does not see sympathy as the mechanism of self-command, that is, rational choice in the face of temptation. Once we see sympathy as the solution to the commitment problem, however, we do not need extrarationality considerations to account for the attenuation pathway of sympathy.

C. Sympathy with the Dead and Insane

In opposition to Hume's view, Smith's notion of sympathy-as-judgment can also explain why we feel sympathy with the insane and the dead (Smith 1976, 54). Sympathy with the insane and the dead would be perplexing—in fact, Sugden (2002) finds it incoherent—if sympathy is only sympathy as reflexivity, that is, what takes place in Hume's "ordinary Cases." If we follow Hume's notion, sympathy with the insane and the dead could not take place.

Smith discusses the paradox of sympathy with the insane and the dead at the start of TMS, as he tries to distance his notion of sympathy from Hume's. In the paragraph prior to his discussion of sympathy with the insane and the dead, Smith defines sympathy as the outcome of transporting oneself to the station of the other and feeling what the other would feel:

> Sympathy, therefore, does not arise so much from the view of the passion, as from that of the situation which excites it. We sometimes feel for another, a passion of which he himself seems to be altogether incapable; because, when we put ourselves in his case, that passion arises in our breast from the imagination, though it does not in his from the reality. We blush for the impudence and rudeness of another, though he himself appears to have no sense of the impropriety of his own behaviour; because we cannot help feeling with what confusion we ourselves should be covered, had we behaved in so absurd a manner. (Smith 1976, 54)

When the spectator feels sorry for the dead or the insane, the spectator is expressing justifiable sense of loss for what the principal would feel if he or she were alive or normal. The spectator cannot reach such judgment if he or she is merely imitating the feeling of the principal because such a feeling does not exist. Also, the spectator cannot reach such a judgment via projection of his or her own feeling—because it is obvious that the dead or the insane is no longer a viable recipient of such a projection.

Therefore, if we adopt Smith's notion of sympathy, we can explain why the spectator would feel sorrow for the dead and the insane. In the case of the dead, the sorrow is a negative function of the principal's age, and probably, a positive function of his or her former standing or potential accomplishments that were not realized. The same probably applies to the insane.

D. Is the Fellow-Feeling Paradox Fully Resolved?

Did Smith succeed in answering Hume? Not really. Smith simply elaborated on his

view of sympathy as judgment of propriety, which is responsible for the attenuation of original emotions. This does not mean that Hume's observation about fellow feeling, as the accelerator of original emotions, is invalid. Smith simply offered a view of one kind of fellow feeling, which does not rule out Hume's kind. Hume's fellow feeling might be a form of self-centered fellow feeling that may not involve judgment of propriety. So, Smith's answer does not amount to a full resolution of the fellow-feeling paradox and such a resolution cannot be investigated here (Khalil 2007).

Nonetheless, by expressing his view, Smith illustrated how moral action is rational action and how both are based on the emotions. Given the focus here, Smith showed that at least one kind of fellow feeling need not accelerate the original emotion. He succeeded in showing that sympathy, understood as involving judgment of propriety, assists rational decision making when dynamic inconsistency or temptation is present. In doing so, Smith ultimately undermined the Cartesian emotion/rationality or emotion/morality dichotomy.

IV. SYMPATHY, RATIONALITY, AND DYNAMIC INCONSISTENCY

The spectator judges the emotion/action of the principal according to whether or not it is optimal, that is, the spectator takes into account the interest of present as well as future selves. The idea of the spectator as a judge is called by Smith the "impartial spectator." As detailed elsewhere (Khalil 1990), this impartial spectator resides within the breast—what Smith explicitly called the "man within" (Smith 1976, 130–132) or "conscience" (Ibid, 134–156). The term "impartial spectator" denotes the principal's nonmyopic self as it exercises command over the myopic self. The principal exercises, via the spectator within, self-command in order to act rationally in the face of dynamic inconsistency (temptations).

Smith did not explicitly connect self-command, which he calls throughout the "virtue of self-command," with the spectator within. But his text easily supports such a connection. The text includes discussions that suppose an intimate link between the two. Self-command amounts to following the dictates of the spectator that lives within the breast.

Smith's spectator is ultimately *asocial*. Smith recognizes fully that the principal in many occasions exercises self-command in order to gain the approval of public opinion or actual spectators, that is, what Smith called the "man without" (Ibid, 130–131). But Smith condemns such a phenomenon as vanity and the pursuit of praise; the agent should instead be pursuing praiseworthiness in itself (Smith 1976, 113–134). Smith therefore emphasizes, with his stress on praiseworthiness, that the spectator *within* is ultimately autonomous of the spectator *without*—which is contrary to the interpretation of Robert Heilbroner (1982) and the editors of the Glasgow Edition of TMS (Khalil 1990; Smith 1976).

This asocial view of the spectator is supported by the argument that, in many of the cases where the principal is acting to gain the approval of spectator *without*, the spectator *without* could actually be a projection of the preferences of the spectator *within*. That is, the principal may set up goals for him or herself and then go out and pick up a peer group or network that would act as a gauge. It is commonly the case that agents do indeed choose their own social pond that acts exactly as a gauge of one's accomplishment (Frank 1985).

The principal's judgment of his or her own behavior is *not*, ultimately, therefore the internalization of social norms. Rather, it is the product of distancing that he or she can do even in the absence of society. The fact that the principal lives in a society, however, means that he or she is able to pass judgments on others, which permits the principal to learn faster about propriety. That is, society is not essential for the formation of judgment but rather acts as a mirror that enhances the ability of the agent to undertake the proper distance from his or her immediate emotions. Society allows the agent to self-reflect easier. In this manner, Smith does not start with social norms—which actually would beg the question about the source of such norms. Rather, he starts with the individual—an individual who is involved in self-judgment because the individual is interested in promoting his or her well-being. As such, the individual learns from experience to look at his or her own passions from a distance; society is not essential for such learning.

To argue this, Smith (1976, 134–135) uses the analogy of vision. Agents normally adjust for distance when they compare sizes of objects. Agents learn that object X would look bigger than object Y, which is of equal size, if agents are closer to X than to Y. It is worth quoting Smith's discussion of distance adjustment of vision in TMS:

> As to the eye of the body, objects appear great or small, not so much according to their real dimensions, as according to the nearness or distance of their situation; so do they likewise to what may be called the natural eye of the mind: and we remedy the defects of both these organs pretty much in the same manner. In my present situation an immense landscape of lawns, and woods, and distant mountains, seems to do no more than cover the little window which I write by, and to be out of all proportion less than the chamber in which I am sitting. I can form a just comparison between those great objects and the little objects around me, in no other way, than by transporting myself, at least in fancy, to a different station, from whence I can survey both at nearly equal distances, and thereby form some judgment of their real proportions. Habit and experience have taught me to do this so easily and so readily, that I am scarce sensible that I do it; and a man must be, in some measure, acquainted with the philosophy of vision, before he can be thoroughly convinced, how little those distant objects would appear to the eye, if the imagination, from a knowledge of their real magnitudes, did not swell and dilate them. (Smith 1976, 134–135)[6]

Likewise, for Smith, current passions or personal pain might loom larger than other experiences if the agent fails to adjust for "distance."[7]

To gain a rational view of one's personal pain, one should try to get sympathy, that is, the approval of the spectator within. Smith's notion of sympathy amounts to rational decision making in the face of temptations. These temptations lead to the possibility of dynamic inconsistency: the agent commits dynamic inconsistency when he or she undermines the interest of future self beyond what is optimum. The agent can do this in two ways. First, the agent can indulge the current self in anger or in overeating beyond what is optimum. Second, the agent can indulge the current self in shoplifting, loafing, or violating promises to abide by rules of fairness and equity: these acts might hurt others who, in turn, would retaliate and hurt the future self beyond what is optimum. (Although beyond the scope of this article, this last point may explain much cooperation seen in society.)

In Smith's analysis, the myopic self must obey the dictates of the nonmyopic, rational self, which he called the impartial spectator. As a result of such self-command, the original passions of the myopic self must be calmed and controlled. So, for Smith, it is not a paradox that sympathy affords attenuation, rather than escalation, of the emotions.

V. RAMIFICATIONS

A. Self-Control versus Self-Command

Smith's notion of sympathy is directly related to the issue of how to maintain commitment to an optimal action, given temptations and dynamic inconsistency. Obviously, *one* way to maintain commitment, which economists have studied extensively, is precommitment. Examples of precommitment include the taking of hypnosis to get rid of the smoking habit, retiring to so-called "fat farms" to lose weight, or submitting to other drastic measures to control appetites. These precommitment strategies, generally called "self-control," are inflexible (e.g., Fudenberg and Levine 2006; Gul and Pesendorfer 2001, 2004a, 2004b; Sally 2000). Such inflexibility can be very costly in environments that are uncertain (e.g., Amador, Werning, and Angeletos 2006), but they can be justified if the cost of yielding to temptation is sufficiently higher than the cost of inflexibility.

Such self-control, in the sense of precommitment, is a major theme in the work of Schelling (1960, 1978, 1984a, 1984b, 1992) on the commitment problem. Schelling uses the terms "self-management" and "self-command" to describe what is called here "self-control." Whatever terminology one adopts, though, one should not confuse Schelling's concept of *self-control* with Smith's concept of *self-command*. For Schelling, the agent solves the commitment problem by ex ante exchanges of

6. As Levy (1995) notes, Smith's discussion follows Berkeley's *New Theory of Vision*, as evident from Smith's explicit reference to Berkeley in his earlier essay on "External Senses" (Smith 1980, 148).

7. Smith uses the analogy of distant adjustment in vision to specifically discuss altruism, that is, how to allocate resources between the self and an important nonself other. By definition, altruism entails society, while self-command does not. But one can use the method of distancing in relation to altruism in order to discuss the broader case of self-command, that is, the allocation of resources between present and future selves.

hostages to ensure the cooperation of the other party. This amounts to precommitment since the restraints are irrevocable or inflexible, as in the case of Ulysses's request to be tied to the mast of the ship in order to avoid the luring voices of the Sirens (Elster 1984, 2000). In contrast, Smith's self-command does not entail the ex ante wearing of straightjackets. The principal maintains full flexibility, while still abiding by strategies to avoid dynamic inconsistency.

From casual observation, Smith's self-command seems to have some empirical corroboration. The exchange of hostages or ex ante destruction of options is not as widespread as economic theory predicts. Agents seem to command their appetites without ex ante measures that entail costly inflexibility. For example, Lindenberg (1993) argues that the rearing of children involves, at least partially, an investment in what he calls "self-command capital." Irrespective of the role of rearing of children, however, self-command, in Smith's account, arises from nonmyopic concern about self-interest. As such, agents, or parents as they rear children, would opt for self-command to ensure the maximization of well-being.

Such self-command ensures public order, justice, and respect of the property rights of others without resorting to a "police society." A police society, as defined here, denotes the extreme reign of self-control measures at the level of society, irrespective of what political structure the society has. In a police society, one would therefore expect a sheriff to be stationed, after being justified by the cost, at every critical corner of the city and to be thoroughly searched, if justified by the cost, upon leaving every shop. Such policing or a system of controls might even be demanded by the population to prevent self-cheating, which has come to be called "libertarian paternalism" (Sunstein and Thaler 2003). But most stable societies are not police societies. Agents command their impulses and restrain themselves from stealing without extensive *external* constraints ranging from monitors to prisons.

Given the excessive focus on external constraints, including self-controls, it is useful to read Smith's notion of sympathy as being about *internal* constraints. Smith's self-command mechanism highlights an alternative mechanism to punishment or other imposed external constraints.

B. Justice and Prudence

Adam Smith highlighted the virtue of propriety throughout TMS: it is the choice of action that is efficient. But does efficiency include the observation of property rights (i.e., acting in accordance with justice)? If property rights—or justice—is grounded on efficiency grounds, propriety would therefore also include the observation of property rights; after all, as Montes (2004, Chapter 4] shows skilfully, the terms property and propriety actually share the same etymology.

The quest after justice has always caused a problem for political theorists. Why would agents ever act to uphold justice (i.e., to obey property rights or to fulfill promises), when it pays to cheat? As already suggested, cheating others is not analytically different from cheating one's self or succumbing to the myopic passion of the moment. Both are different facets of the commitment problem. In the case of the cheating of others, there is the possibility of retribution that would hurt the future self. The possibility might be small and hence justify cheating. But it might invite recklessness that would undermine the future self. So, agents follow the rules of fairness concerning the rights of others (justice) for the same reason that they follow the rules of fairness concerning the rights the future self (prudence). In both cases, they abide by the commands of the spectator *within* in order to protect the future self. That is, following justice is not contrary to interest as much as following prudence is not contrary to interest: Social contract theorists from John Locke to Posner (1994) have also argued that justice is rooted in welfare maximization.

On the other hand, justice is problematic because it is not a simple calculation of self-interest. Agents in finite voluntary contribution mechanism (VCM) games tend to regard justice as a sacred principle, which makes it more imperative than prudence. This makes it seem that agents should pursue justice even when it contradicts interest. In light of the above discussion, however, the sacred or moral aspect of justice can be interpreted in two ways. First, the quest for such a principle commits the agent to the optimum in the face of opportunistic, inefficient strategies. As such, the gap between "justice" and "expediency" of the myopic self is simply the gap between, respectively, the view of the spectator

within and the view of the impetuous (myopic) principal. Second, the quest, once deemed optimal, gives rise to a second source of satisfaction, viz., the sense of self-approval or self-integrity for doing what is optimal. This means that one may act according to justice, when it pays to cheat, by the secondary motive of experiencing self-integrity. But, again, for this secondary motive to operate, the act of justice must have been already commanded by the impartial spectator, the spectator within, because it is deemed to be the optimal act.

C. Voluntary Cooperation

One aspect of justice, which is buttressed by self-command, is voluntary cooperation. Voluntary cooperation is defined as the choice to cooperate when a subgame perfect Nash equilibrium predicts defection. For example, experimental findings in VCM games show that agents often choose to cooperate, even when defection is more profitable (Deci, Koestner, and Ryan 1999; Fehr and Gächter 2000a, 2000b, 2002). Furthermore, Akerlof (1982) noted that wages encourage workers to cooperate beyond what is justified by financial incentives: in this sense, he called such wages a "partial gift" (Organ 1988; Simon 1997). Wages are partial gifts when they are set above market-clearing wages, which may explain why monetary wages do not decrease in recessions (Bewley 1999; Fehr, Kirchsteiger, and Riedl 1993). Cooperation, as in these examples, is a voluntary choice for people who live in a normal society, where the police are not present at every corner.

Smith's concept of self-command explains voluntary cooperation, when other explanations fail. Currently, there are three major explanations (Khalil 2003): strategic, social preferences, and evolutionary. According to the strategic explanation, agents cooperate because they want to encourage reciprocity, that is, they care about future returns (Trivers 1971; Tullock 1985). This might be true in infinitely repeated games. However, in finite games, such as in VCM games, the subgame perfect Nash equilibrium is for each agent to defect at the first round, when in fact they do not. Agents cooperate in first rounds even when they are matched randomly with other agents. Binmore (2006) argues that such an anomaly can be explained as the result of confusion at the beginning of the game (thus, supposedly, as soon as agents, toward the end, learn the game, they start to free ride). It is true that, as the game is close to its conclusion cooperation decays. But is confusion the reason why agents usually cooperate at the beginning? Andreoni (1995; Fischbacher, Gächter, and Fehr 2001; Houser and Kurzban 2002) has shown that agents are, in fact, far from confused and that a lack of knowledge of the game cannot explain cooperation at its beginning.

According to the social preferences explanation, the utility function of agents contains "social preferences," that is, preferences that make them amiable to favor cooperation because they possess a prosocial taste called "equity" or "fairness"—which is usually costly. As defined here, this notion of "social preferences" is about fairness and hence should not be confused with altruism, which is simply about the allocation of a pie according to how one cares about self-interest as opposed to the interest of a loved one (Khalil 2004). When one acts according to fairness, one feels obliged to honor the property rights of others, even when one does not know or does not care about the others. Thus, the agent rather cares about his or her own integrity or honesty. Understood in this manner, the concept of social preferences is best illustrated in the Dictator game (Camerer 2003, 57–58; Forsythe et al. 1994), where proposers share, on average, 20% of an endowment with responders when the responders have no power to punish. Such a tendency has been called "inequity aversion" (Fehr and Schmidt 1999), "guilt aversion" or "let-down aversion" (Charness and Dufwenberg 2006), taste for "equity" (Bolton 1991; Bolton and Ockenfels 2000; Dufwenberg and Kirchsteiger 2004; Rabin 1993), distaste for "shame" (Bowles and Gintis 2001; Gintis et al. 2005; Henrich et al. 2004), or "warm glow" (Andreoni 1990).

The problem with the social preferences explanation is that it is ad hoc: it creates new tastes, without explanation, and makes them heterogeneous across agents as a first level of approximation. Thus, it violates the principle of stable preferences (Khalil 2008; Stigler and Becker 1977). As List (2007) and Bardsley (2005) show that a change in the set of options can alter the behavior of proposers. Furthermore, if voluntary giving to

strangers is a preference, it would be an unfit trait in an evolutionary environment.

Those who support the evolutionary explanation, on the other hand, maintain that an evolutionary game selection mechanism might guarantee the appearance of cooperators—through one of two mechanisms. First, agents would prefer to deal with honest cooperators (Frank 1988). Second, agents might be afraid to cheat partners because partners might be irrational, that is, would retaliate even when it is too costly to do so (Kreps et al. 1982). But the evolutionary game explanation faces two problems. First, it assumes that a society has already reached the required critical ratio of cooperators because the trait would not spread unless it exists above the critical ratio. This begs the question of how this critical ratio comes about in the first place. Second, if agents are built to cooperate, it cannot explain why cooperation quickly decays as the game approaches the end.

Smith's notion of self-command offers a plausible way out, without an appeal to ad hoc reasoning. With the same concept of self-interest, one can simultaneously explain voluntary cooperation as well as the decay of such cooperation in VCM. Agents cooperate because it is the optimum option to protect the future self, given that everyone is cooperating. But how then can one explain the decay of cooperation? It might be the result of the lack of coordination (Chwe 2001). The lack of coordination may make agents suspicious of how others would behave, which would leave them at a disadvantage. Thus, if one cares about the interest of the future self, one should defect. A cascade occurs—where a suspicion becomes a self-fulfilling prophecy. If so, there is no need to invent new tastes, such as social preferences or inequity aversion.

But why does the lack of coordination not appear at the beginning of the game? This issue is beyond the scope of this article. It is arguable, though, that agents learn about the potential of miscoordination as they proceed in the game. The miscoordination appears to them as free riding, which prompts them to defect. In any case, it is possible with the same concept of self-command to not only explain initial cooperation but also the eventual decay of such cooperation when there is a lack of coordination.

D. Smith and Neuroscience

The recent discovery of MNS, spearheaded by Rizzolatti, has awakened an interest in Smith's theory of sympathy (Rizzolatti and Craighero 2005; Rustichini 2005). The MNS is mostly located in the F5 region, and partly in the PF area, of the brains of monkeys and humans. Research is still ongoing concerning the MNS, its extent, and its relation to imitation and learning. It is beyond the confines of this article to discuss MNS. Stated briefly, however, the MNS becomes engaged, that is, firing signals, irrespective of whether the spectator is watching the principal's action, or whether the spectator is doing the action him or herself (Rizzolatti and Craighero 2005; Rizzolatti et al. 1999). For instance, the MNS fires when one tries to grasp a cup or when one watches another subject trying to grasp a cup. What does matter for the engagement of the spectator's MNS is that the witnessed action is itself motivated by an intelligible reason, in this case the attempt of grasping a cup. Thus, the MNS in the F5 region of the spectator would not fire if the spectator simply observes the principal taking a pantomime kind of grasp, such as a movement of the hand in the air for no purpose.

The matter is, however, more complicated, given that there is a sister system called "canonical neurons," that fires for similar reasons as the MNS. This system, however, also fires when the spectator notices small objects that *can be* handled with precision—when in fact neither the principal nor the spectator him or herself is undertaking the action (Iacoboni 2005). These details can be ignored for our purpose as what really matters is that the MNS discovery gives a boost to Smith's notion of sympathy. Namely, sympathetic feelings involve the understanding and approval of the actions taken in light of the cause. This does not deny that reflexivity, noted by Hume, also takes place. The point is rather that the MNS discovery shows that at least some fellow feeling involves understanding. The mirror neurons would not fire by simply watching a pantomime kind of a grasp. They would fire only when the grasp is motivated by an object, which involves understanding (Khalil 2007).

Further, Smith's abridgment of rationality, morality, and the emotions are now a major thrust of modern neuroscience (e.g.,

Gazzaniga 2005). In Smith's theory, there is no sharp divide between the emotions and the intellect or between sentiments and morality. Smith's theory should thus appeal to modern neuroscientists interested in undermining the Cartesian dualism between mind and brain (e.g., Damasio 1994). If so, Smith may be the *first* social scientist, but maybe not the first philosopher, if we count Spinoza, to reject Descartes' dualism.

VI. CONCLUSIONS

Many thinkers, going back to the 1826 essay of Hazlitt (2004), have questioned the Cartesian dualism, known as the emotion/morality or the emotion/rationality dichotomy. What distinguishes Smith's theory from this duality is his concept of sympathy: sympathy shows the continuity, or even identity, of emotions, morality, and rationality. For Smith, rational action is moral action, which, in turn, is based on the emotions. So, for Smith, there is no difference between facts (emotions) and values (morality)—a dichotomy that has been challenged too infrequently in economics, with the exception of the works of Walsh (1961, 2000) and Sen (1987, 2002; see Putnam 2002, 2003). Other thinkers who have also emphasized the role of sympathy, however, have generally failed to escape the emotion/morality or emotion/rationality dichotomy (e.g., Scheler 1954; Sugden 2002).

Smith affirmed the unity of the self by postulating that sympathy, while a state of emotion, involves judgment of propriety: the spectator expresses the emotions of sympathy only when he or she finds the principal's emotions to be proper vis-à-vis the incentive. Such a judgment of propriety amounts to the exercise of rational judgment. So for an emotion to be sympathetic, the act must be proper or rational. A proper emotion, that is, sympathy, is the product of a process of internal reflection by the principal: it is not the emotion that *accompanies* the immediate incentive (i.e., temptations), but rather a weighted or nonmyopic emotion that arises *in response to* immediate incentives.

Smith's concept of sympathy expresses a tension between the principal, who is myopic, and the "impartial spectator," who is actually the nonmyopic principal. So, if the principal is interested in gaining sympathy of the impartial spectator (the nonmyopic principal), the principal must exercise self-command. And such self-command ensures that the action complies with the optimum decision. Such a concept of self-command, based on his concept of sympathy, solves the commitment problem: through sympathy, agents adopt the rational choice in the face of temptations. Sympathy affords an internal constraint, "self-command," which is successful in preventing agents from succumbing to temptations when external constraints cannot.

REFERENCES

Akerlof, G. A. "Labor Contracts as Partial Gift Exchange." *Quarterly Journal of Economics*, 97, 1982, 543–69.

Amador, M., I. Werning, and G.-M. Angeletos. "Commitment vs. Flexibility." *Econometrica*, 74, 2006, 365–96.

Andreoni, J. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving." *Economic Journal*, 100, 1990, 464–77.

———. "Cooperation in Public-Goods Experiments: Kindness or Confusion?" *American Economic Review*, 85, 1995, 891–904.

Bardsley, N. "Altruism or Artefact? A Note on Dictator Game Giving." Discussion Paper Series, 2005–10, The Centre for Decision Research and Experimental Economics, University of Nottingham, 2005.

Bewley, T. *Why Wages Don't Fall in a Recession*. Cambridge: Harvard University Press, 1999.

Binmore, K. "Economic Man—or Straw Man?" *Behavioral and Brain Sciences*, 28, 2006, 817–18.

Bolton, G. E. "A Comparative Model of Bargaining: Theory and Evidence." *American Economic Review*, 81, 1991, 1096–136.

Bolton, G. E., and A. Ockenfels. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review*, 90, 2000, 166–93.

Bowles, S., and H. Gintis. "The Economics of Shame and Punishment." Working Paper, University of Massachusetts, 2001.

Camerer, C. F. *Behavioral Game Theory*. Princeton, NJ: Princeton University Press, 2003.

Charness, G., and M. Dufwenberg. "Promises and Partnership." *Econometrica*, 74, 2006, 1579–601.

Chwe, M. S.-Y. *Rational Ritual: Culture, Coordination, and Common Knowledge*. Princeton, NJ: Princeton University Press, 2001.

Damasio, A. R. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: G.P. Putnam, 1994.

Deci, E. L., R. Koestner, and R. M. Ryan. "A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation." *Psychological Bulletin*, 125, 1999, 627–68.

Dufwenberg, M., and G. Kirchsteiger. "A Theory of Sequential Reciprocity." *Games and Economic Behavior*, 47, 2004, 268–98.

Elster, J. *Ulysses and the Sirens: Studies in Rationality and Irrationality*. Rev. Edn. Cambridge: Cambridge University Press, 1984.

———. *Ulysses Unbound: Studies in Rationality, Precommitment, and Constraints*. Cambridge: Cambridge University Press, 2000.

Fehr, E., and S. Gächter. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review*, 90, 2000a, 980–94.

———. "Fairness and Retaliation: The Economics of Reciprocity." *Journal of Economics Perspectives*, 14, 2000b, 159–81.

———. "Altruistic Punishment in Humans." *Nature*, 415, 2002, 137–40.

Fehr, E., G. Kirchsteiger, and A. Riedl. "Does Fairness Prevent Market Clearing—An Experimental Investigation." *Quarterly Journal of Economics*, 108, 1993, 437–59.

Fehr, E., and K. M. Schmidt. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 114, 1999, 817–68.

Fischbacher, U., S. Gächter, and E. Fehr. "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment." *Economics Letters*, 71, 2001, 397–404.

Fontaine, P. "Recognition and Economic Behavior: Sympathy and Empathy in Historical Perspective." *Economics and Philosophy*, 13, 1997, 261–80.

Forsythe, R., J. Horowitz, N. Savin, and M. Sefton. "Fairness in Simple Bargaining Experiments." *Games and Economic Behavior*, 6, 1994, 347–69.

Frank, R. H. *Choosing the Right Pond*. New York: Oxford University Press, 1985.

———. *Passions Within Reason: The Strategic Role of the Emotions*. New York: W.W. Norton, 1988.

Fudenberg, D., and D. K. Levine. "A Dual Self Model of Impulse Control." *American Economic Review*, 96(5), December 2006, 1449–76.

Gazzaniga, M. S. *The Ethical Brain*. New York: Dana Press, 2005.

Gintis, H., S. Bowles, R. Boyd, and E. Fehr, eds. *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life*. Cambridge, MA: MIT Press, 2005.

Gladstein, G. A. "The Historical Roots of Contemporary Empathy Research." *Journal of the History of the Behavioral Sciences*, 20, 1984, 38–59.

Gordon, R. "Sympathy, Simulation, and the Impartial Spectator." *Ethics*, 105, 1995, 727–42. (Reprinted in L. May, M. Friedman, and A. Clark, eds. Mind and Morals: Essays on Ethics and Cognitive Science. Cambridge, MA: MIT Press, 1996.)

Gul, F., and W. Pesendorfer. "Temptation and Self Control." *Econometrica*, 69, 2001, 1403–35.

———. "Self Control and the Theory of Consumption." *Econometrica*, 72, 2004a, 110–58.

———. "Self Control, Revealed Preference and Consumption Choice." *Review of Economic Dynamics*, 7, 2004b, 243–64.

Haakonssen, K., ed. "Introduction," in *Adam Smith, The Theory of Moral Sentiments*. Cambridge: Cambridge University Press, 2002, vii–xxiv.

Hazlitt, W. "On Reason and Imagination," in *W. Hazlitt's On the Pleasure of Hating*. London, UK: Penguin, (1826), 2004, 84–103.

Heilbroner, R. L. "The Socialization of the Individual in Adam Smith." *History of Political Economy*, 14, 1982, 427–39.

Henrich, J., R. Boyd, S. Bowles, C. F. Camerer, E. Fehr, and H. Gintis, eds. *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*. Oxford: Oxford University Press, 2004. Accessed 23 August 2008. http://www.oxfordscholarship.com/

Houser, D., and R. Kurzban. "Revisiting Kindness and Confusion in Public Goods Experiments." *American Economic Review*, 92, 2002, 1062–69.

Hume, D. *A Treatise of Human Nature*, 3 vols [1740], edited by L. A. Selby-Bigge. Oxford: Clarendon Press, 1896.

Iacoboni, M. "Understanding Others: Imitation, Language, and Empathy," in *Perspectives on Imitation: From Neuroscience to Social Science*, 2 vols, Vol. 1, Chapter 2, edited by S. Hurley and N. Chater. Cambridge, MA: MIT Press, 2005, 77–99.

Isaac, M. R., K. McCue, and C. R. Plott. "Public Goods Provision in an Experimental Environment." *Journal of Public Economics*, 26, 1985, 51–74.

Khalil, E. L. "Beyond Self-Interest and Altruism: A Reconstruction of Adam Smith's Theory of Human Conduct." *Economics and Philosophy*, 6, 1990, 255–73.

———. "Why does Trustworthiness Pay? Three Answers: An Introduction," in *Trust*, edited by E. L. Khalil. Cheltenham, UK: Edward Elgar, 2003, xiii–xxxii.

———. "What is Altruism?" *Journal of Economic Psychology*, 25, 2004, 97–123.

———. "An Anatomy of Authority: Adam Smith as Political Theorist." *Cambridge Journal of Economics*, 29, 2005, 57–71.

———. Smith's Theory of Sympathy—Where does it *Succeed* and Where does it *Fail*?" Working Paper, 2007.

———. "Tastes," in *International Encyclopaedia of the Social Sciences*, Vol. 8, 2nd ed, edited by W. A. Darity Jr. Detroit: Macmillan Reference USA, 2008, 266–70.

———. Forthcoming. "Self-Deceit and Self-Serving Bias: Adam Smith on 'General Rules'." *Journal of Institutional Economics*, 2009, 5.

Kreps, D. M., P. Milgrom, J. Roberts, and R. Wilson. "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma." *Journal of Economic Theory*, 27, 1982, 245–52.

Levy, D. M. "The Partial Spectator in the *Wealth of Nations: A Robust Utilitarianism*." *European Journal of the History of Economic Thought*, 2, 1995, 299–326.

Levy, D. M, and S. J. Peart. "Sympathy and Approbation in Hume and Smith: A Solution to the other Rational Species Problem." *Economics and Philosophy*, 20, 2004, 331–49.

Lindenberg, S. "Club Hierarchy, Social Metering and Context Instruction: Governance Structures in Response to Varying Self-Command Capital," in *Interdisciplinary Perspectives on Organization Studies*, edited by S. Lindenberg and H. Schreuder. Oxford: Pergamon Press, 1993, 195–220.

List, J. A. "On the Interpretation of Giving in Dictator Games." *Journal of Political Economy*, 115, 2007, 482–93.

Meardon, S. J., and A. Ortmann. "Self-Command in Adam Smith's Theory of Moral Sentiments: A Game-Theoretic Reinterpretation." *Rationality and Society*, 8, 1996, 57–80.

Montes, L. *Adam Smith in Context: A Critical Reassessment of Some Central Components of his Thought*. New York: Palgrave Macmillan, 2004.

Organ, D. W. *Organizational Citizenship Behavior: The Good Soldier Syndrome*. Lexington, MA: Lexington Books, 1988.

Posner, R. A. "Law and Economics is Moral," in *Adam Smith and the philosophy of law and economics*, edited by R. P. Malloy and J. Evensky. Dordrecht: Kluwer, 1994, 167–78.

Putnam, H. *The Collapse of the Fact/Value Dichotomy and Other Essays*. Cambridge, MA: Harvard University Press, 2002.

———. "For Ethics and Economics Without the Dichotomies." *Review of Political Economy*, 15, 2003, 395–412.

Rabin, M. "Incorporating Fairness into Game Theory and Economics." *American Economic Review*, 83, 1993, 1281–302.

Rizzolatti, G., and L. Craighero. "The Mirror-Neuron System." *Annual Review of Neuroscience*, 27, 2004, 169–92.

———. "Mirror Neuron: A Neurological Approach to Empathy," in *Neurobiology of Human Values*, edited by J. P. Changeux et al. Berlin: Springer-Verlag, 2005, 107–23.

Rizzolatti, G., L. Fadiga, L. Fogassi, and V. Ballege. "Resonance Behaviours and Mirror Neurons." *Archives Italiennes De Biologie*, 137, 1999, 88–99.

Rustichini, A. "Neuroeconomics: Present and Future." *Games and Economic Behavior*, 52, 2005, 201–12.

Sally, D. "I, Too, Sail Past: Odysseus and the Logic of Self-Control." *Kyklos*, 53, 2000, 173–200.

Scheler, M. *The Nature of Sympathy*, trans. Peter Heath, intro. by W. Stark. New Haven, CT: Yale University Press, 1954.

Schelling, T. C. *The Strategy of Conflict*. London: Oxford University Press, 1960.

———. "Egonomics, or the Art of Self-Management." *American Economic Review, Papers and Proceedings*, 68, 1978, 290–94.

———. "Self-Command in Practice, in Policy, and in a Theory of Rational Choice." *American Economic Review, Papers and Proceedings*, 74, 1984a, 1–11.

———. *Choice and Consequence*. Cambridge: Harvard University Press, 1984b.

———. "Self-Command: A New Discipline," in *Choice Over Time*, edited by G. Loewenstein and J. Elster. New York: Russell Sage Foundation, 1992, 167–76.

Sen, A. *On Ethics and Economics*. Oxford: Blackwell, 1987.

———. *Rationality and Freedom*. Cambridge, MA: Harvard University Press, 2002.

Simon, H. A. *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organizations*. New York: Free Press, 1997.

Smith, A. *The Theory of Moral Sentiments*, edited by D. D. Raphael and A. L. Macfie. Oxford: Clarendon Press, 1976.

———. *The Correspondence of Adam Smith*, edited by E. C. Mossner and I. S. Moss. Oxford: Oxford University Press, 1977.

———. *Essays on Philosophical Subjects*, general editors D. D. Raphael and A. S. Skinner, edited by W. P. D. Wightman, J. C. Bryce, and I. S. Ross. Oxford: Clarendon Press, 1980.

Sugden, R. "Beyond Sympathy and Empathy: Adam Smith's Concept of Fellow-Feeling." *Economics and Philosophy*, 18, 2002, 63–87.

Sunstein, C. R., and R. Thaler. "Libertarian Paternalism is Not an Oxymoron." *University of Chicago Law Review*, 70, 2003, 1159–1202.

Sutrop, M. "Adam Smith on Sympathy and Moral Motivation." *Acta Philosophica Fennica*, 83, 2007, 143–68.

Stein, E. (Saint Teresa Benedicta of the Cross Discalced Carmelite). *The Science of the Cross*, trans. by Josephine Koeppel. Washington, DC: ICS Publications, 2002.

Stigler, G. J., and G. S. Becker. "De Gustibus Non Est Disputandum." *American Economic Review*, 67, 1977, 76–90.

Trivers, R. L. "The Evolution of Reciprocal Altruism." *Quarterly Review of Biology*, 46, 1971, 35–57.

Tullock, G. "Adam Smith and the Prisoners' Dilemma." *Quarterly Journal of Economics*, 100, 1985, 1073–81.

Walsh, V. C. *Scarcity and Evil*. Englewood Cliffs, NJ: Prentice-Hall, 1961.

Walsh, V. C. "Smith After Sen." *Review of Political Economy*, 12, 2000, 5–25.