Journal of Institutional Economics (2009), 5: 2, 251–258 Printed in the United Kingdom
© *The JOIE Foundation 2009* doi:10.1017/S1744137409001325

Self-deceit and self-serving bias: Adam Smith on 'General Rules'

ELIAS L. KHALIL

Adam Smith (1982) undertook a momentous project in *The Theory of Moral Sentiments* (TMS): He wanted to debunk the Cartesian dichotomy between the emotions (body) and rationality (mind). The debunking of the body/mind dichotomy is a major aim of many modern neuroscientists and neuroeconomists (e.g., Damasio, 1994; Glimcher, 2004; Rustichini, 2005). Smith also wanted to demonstrate that morality is grounded on the emotions – as the title of his book indicates. So for Smith, the twin pillars of what thinkers consider to be distinguishably human, namely, rationality and morality, are actually rooted in the emotions (body). More specifically, while rationality and morality are based on the emotions, Smith showed that the emotions are rather refined and sophisticated sentiments adopted by agents in light of how they judge the actions of others and, in turn, how they should judge their own actions. Such judgments are nothing but the determination of what is the rational (optimal) solution. So, sentiments are not hurdles in the pathway of rational decisions. They are rather the building blocks of such decisions.

Smith dedicated Part I of TMS to discussing propriety or the virtue of self-command. Part I shows how raw sentiments are regulated to become sophisticated sentiments. This introduction, and the excerpted chapter from Part III of TMS, recapitulate Smith's analysis of propriety and self-command. But this has been done with the specific focus on an anomaly that Smith's momentous project faces: If human conduct is based on sentiments, and sentiments are regulated through self-command, why do societies adhere to institutions concerning moral conduct that he calls 'general rules'? That is, if morality arises spontaneously and naturally from everyday interactions, why do societies uphold 'intrusive' social rules? Such rules seem to go against the sentiments; they impose prescriptions that circumvent the operation of the emotions. Such social rules can become so elaborate that some theorists (namely, theologians) have traced them to celestial axioms – as illustrated by the Ten Commandments. Other theorists (namely, secular political philosophers) have traced them to axioms of reason – along either utilitarian grounds or the Kantian categorical imperative. Adam Smith challenges the axiomatic approach – whether it is theologian or secular – and calls its advocates men of 'systems', i.e. men who think that general social rules exist prior to experience and everyday

https://doi.org/10.1017/S1744137409001325 Published online by Cambridge University Press

interactions. Smith wanted to argue that human morality is actually inseparable from everyday experience in that morality is interlaced with the emotions of everyday experience. This directly challenged the scholastic divide, going back to St Augustine, between the spiritual realm and the bodily needs, as well as it challenged its secular version, namely the Cartesian divide between reason and sentiments.

But still, to restate the anomaly, if human morality is based on sentiments and everyday experience, how can we explain the existence of general rules? To do so, we need first to understand Smith's project, which cannot be discussed here in detail (see Khalil, 1990). Briefly stated, however, propriety or self-command consists of internal constraints that check hyperbolic discounting; that is, self-command deals with problems arising from weakness of will or from what Smith in the excerpted chapter below calls the 'violence and injustice of our own selfish passions' (Khalil, 2009). Smith recognizes that immediate sentiments are violent and harmful. So, the agent must subject them to some reflection and evaluation (i.e., self-command) from the standpoint of the 'impartial spectator', which Smith also calls the 'indifferent spectator', or the 'indifferent person'. This impartial spectator is not an actual outside spectator (see Khalil, 1990), but rather resides in the breast of the agent. The internal, impartial spectator simply applies the same sentiments that the agent uses to judge the conduct of others to the conduct of his or her own. The agent, finding him or herself motivated by momentary emotions, would want to regulate such emotions by sentiments that adhere to the judgment that he or she passes on the behaviour of others. The adherence to uniform judgement is nothing but complying with the impartial spectator within the breast. The impartial spectator simply weighs the immediate passions, stirred by the incentive (stimulus), against the concerns for the interest of the future self. Such weighting usually forces the agent to place the immediate passions within perspective, i.e. to examine the present temptation *vis-à-vis* its long-term consequence. This weighting helps the agent avoid the whims of the moment. This amounts to an internal constraint that nullifies the weakness of will (hyperbolic discounting).

An agent who follows the weighted judgment of the impartial spectator does not need general rules, because he or she can follow the advice of the impartial spectator each case at a time; that is, if the impartial spectator is in command, what Smith called the 'virtue of self-command', there is no need to appeal to general rules. So, why do individuals need general rules? Smith's answer is simple: Agents may fall victim to self-deceit – what Smith characterizes as 'the fatal weakness of mankind' – when they think they are following the command of the internal, impartial spectator. The impartial spectator might not really be impartial – without the full awareness of the agent. This veil of ignorance is usually the case when the agent has to make a decision in the heat of the moment. This means that the agent may interpret property rights disputes, or a given piece of information, in a partial manner that favours the current

https://doi.org/10.1017/S1744137409001325 Published online by Cambridge University Press

self without the agent being aware of his or her partiality. Such self-serving bias has been documented in the psychological literature, and more recently in behavioural economics literature as 'positive outcome bias', 'valence effect', 'wishful thinking', 'attributional bias', and 'over-confidence bias' (Rosenhan and Messick, 1966; Miller and Ross, 1975; Hogarth, 1987; Rapoport *et al.*, 1992; Babcock and Loewenstein, 1997).

The self-serving bias might not be universal. Smith recognizes the possibility of a bold person who 'does not hesitate to pull off the mysterious veil of self-delusion, which covers from his view the deformities of his own conduct'. However, it is unquestionable that agents generally fall victim to the self-serving bias on at least some occasions. So, Smith argues, the agent may want to create a general rule that would protect the self from self-deceit. In this manner, for Smith, we 'naturally lay down to ourselves a general rule, that all such actions are to be avoided, as tending to render us odious, contemptible, or punishable, the objects of all those sentiments for which we have the greatest dread and aversion'.

The operation of general rules, therefore, does not contradict Smith's anti-Cartesian project. Smith particularly states that the everyday appeal to general rules 'seems to have misled several very eminent authors, to draw up their systems in such a manner, as if they had supposed that the original judgments of mankind with regard to right and wrong, were formed like the decisions of a court of judicatory, by considering first the general rule, and then, secondly, whether the particular action under consideration fell properly within its comprehension'. General rules are rather the consequence of the operation of the sentiments. Smith suggests that general rules express how 'Nature' corrects the deformities arising from self-deceit. This might be another example of the operation of the invisible hand (Khalil, 2000). But that is another story.

References

Babcock, Linda and George Loewenstein (1977), 'Explaining Bargaining Impasse: The Role of Self-Serving Biases', *Journal of Economic Perspectives*, **11**(1): 109–126.

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

Glimcher, Paul W. (2004), *Decisions, Uncertainty, and the Brain: The Science of Neuroeconomics*, Cambridge, MA: MIT Press.

Hogarth, Robin M. (1987), *Judgement and Choice: The Psychology of Decision*, 2nd edition, Chichester: Wiley.

Khalil, Elias L. (1990), 'Beyond Self-Interest and Altruism: A Reconstruction of Adam Smith's Theory of Human Conduct', *Economics and Philosophy*, **6**(2): 255–273.

Khalil, Elias L. (2000), 'Making Sense of Adam Smith's Invisible Hand: Beyond Pareto Optimality and Unintended Consequences', *Journal of the History of Economic Thought*, **22**(1): 49–63.

https://doi.org/10.1017/S1744137409001325 Published online by Cambridge University Press

Khalil, Elias L. (2009), 'How Adam Smith's Concept of Self-Command Suggests a Solution to Dynamic Inconsistency and the Commitment Problem', *Economic Inquiry*, forthcoming.

Miller, Dale T. and Michael Ross (1975), 'Self-Serving Biases in the Attribution of Causality: Fact or Fiction?', *Psychological Bulletin*, **82**: 213–225.

Smith, Adam (1982), *The Theory of Moral Sentiments*, ed. D. D. Raphael and A. L. Macfie, vol. I of *The Glasgow Edition of the Works and Correspondence of Adam Smith*, Indianapolis: Liberty Fund, 1982. (Reprinted from the Oxford University Press edition, 1976).

Rapoport, Amnon, D. V. Budescu, Ramzi Suleiman, and E. Weg (1992), 'Social Dilemmas with Uniformly Distributed Resources', in Wim B. G. Liebrand, David M. Messick, and Henk A. M. Wilke (eds.), *Social Dilemmas: Theoretical Issues and Research Findings*, Oxford: Pergamon Press.

Rustichini, Aldo (2005), 'Neuroeconomics: Present and future', *Games and Economic Behavior*, **52**(2): 201–212.

Rosenhan, D. L. and S. Messick (1966), 'Affect and Expectation', *Journal of Personality and Social Psychology*, **3**: 38–44.

Excerpt from Adam Smith, *The Theory of Moral Sentiments**

Of the Nature of Self–deceit, and of the Origin and Use of general Rules

1 In order to pervert the rectitude of our own judgments concerning the propriety of our own conduct, it is not always necessary that the real and impartial spectator should be at a great distance. When he is at hand, when (156/157) he is present, the violence and injustice of our own selfish passions are sometimes sufficient to induce the man within the breast to make a report very different from what the real circumstances of the case are capable of authorising.

2 There are two different occasions upon which we examine our own conduct, and endeavour to view it in the light in which the impartial spectator would view it: first, when we are about to act; and secondly, after we have acted. Our views are apt to be very partial in both cases; but they are apt to be most partial when it is of most importance that they should be otherwise.

3 When we are about to act, the eagerness of passion will seldom allow us to consider what we are doing, with the candour of an indifferent person. The violent emotions which at that time agitate us, discolour our views of things;

*This excerpt is from vol. I of *The Glasgow Edition of the Works and Correspondence of Adam Smith*, ed. D.D. Raphael and A.L. Macfie, Indianapolis: Liberty Fund, 1982, Part III, chap. Iv, pp. 156–161. (Reprinted here by permission of Oxford University Press. © Oxford University Press 1976.)

https://doi.org/10.1017/S1744137409001325 Published online by Cambridge University Press

even when we are endeavouring to place ourselves in the situation of another, and to regard the objects that interest us in the light in which they will naturally appear to him, the fury of our own passions constantly calls us back to our own place, where every thing appears magnified and misrepresented by self–love. Of the manner in which those objects would appear to another, of the view which he would take of them, we can obtain, if I may say so, but instantaneous glimpses, which vanish in a moment, and which, even while they last, are not altogether just. We cannot even for that moment divest ourselves entirely of the heat and keenness with which our peculiar situation inspires us, nor consider what we are about to do with the complete impartiality of an equitable judge. The passions, upon this account, as father Malebranche says, all justify themselves, and seem reasonable and proportioned to their objects, as long as we continue to feel them.

4 When the action is over, indeed, and the passions which prompted it have subsided, we can enter more coolly into the sentiments of the indifferent spectator. What before interested us is now become almost as indifferent to us as it always was to him, and we can now examine our own conduct with his candour and impartiality. The man of to–day is no longer agitated by (157/158) the same passions which distracted the man of yesterday: and when the paroxysm of emotion, in the same manner as when the paroxysm of distress, is fairly over, we can identify ourselves, as it were, with the ideal man within the breast, and, in our own character, view, as in the one case, our own situation, so in the other, our own conduct, with the severe eyes of the most impartial spectator. But our judgments now are often of little importance in comparison of what they were before; and can frequently produce nothing but vain regret and unavailing repentance; without always securing us from the like errors in time to come. It is seldom, however, that they are quite candid even in this case. The opinion which we entertain of our own character depends entirely on our judgments concerning our past conduct. It is so disagreeable to think ill of ourselves, that we often purposely turn away our view from those circumstances which might render that judgment unfavourable. He is a bold surgeon, they say, whose hand does not tremble when he performs an operation upon his own person; and he is often equally bold who does not hesitate to pull off the mysterious veil of self–delusion, which covers from his view the deformities of his own conduct. Rather than see our own behaviour under so disagreeable an aspect, we too often, foolishly and weakly, endeavour to exasperate anew those unjust passions which had formerly misled us; we endeavour by artifice to awaken our old hatreds, and irritate afresh our almost forgotten resentments: we even exert ourselves for this miserable purpose, and thus persevere in injustice, merely because we once were unjust, and because we are ashamed and afraid to see that we were so.

5 So partial are the views of mankind with regard to the propriety of their own conduct, both at the time of action and after it; and so difficult is it for them to view it in the light in which any indifferent spectator would consider it. But if it was by a peculiar faculty, such as the moral sense is supposed to be,

https://doi.org/10.1017/S1744137409001325 Published online by Cambridge University Press

that they judged of their own conduct, if they were endued with a particular power of perception, which distinguished the beauty or deformity of passions and affections; as their own passions would be more immediately exposed to the view of this faculty, it would judge with more accuracy concerning them, than concerning those of other men, of which it had only a more distant prospect.

6 This self–deceit, this fatal weakness of mankind, is the source of half the disorders of human life. If we saw ourselves in the light in which others (158/159) see us, or in which they would see us if they knew all, a reformation would generally be unavoidable. We could not otherwise endure the sight.

7 Nature, however, has not left this weakness, which is of so much importance, altogether without a remedy; nor has she abandoned us entirely to the delusions of self–love. Our continual observations upon the conduct of others, insensibly lead us to form to ourselves certain general rules concerning what is fit and proper either to be done or to be avoided. Some of their actions shock all our natural sentiments. We hear every body about us express the like detestation against them. This still further confirms, and even exasperates our natural sense of their deformity. It satisfies us that we view them in the proper light, when we see other people view them in the same light. We resolve never to be guilty of the like, nor ever, upon any account, to render ourselves in this manner the objects of universal disapprobation. We thus naturally lay down to ourselves a general rule, that all such actions are to be avoided, as tending to render us odious, contemptible, or punishable, the objects of all those sentiments for which we have the greatest dread and aversion. Other actions, on the contrary, call forth our approbation, and we hear every body around us express the same favourable opinion concerning them. Every body is eager to honour and reward them. They excite all those sentiments for which we have by nature the strongest desire; the love, the gratitude, the admiration of mankind. We become ambitious of performing the like; and thus naturally lay down to ourselves a rule of another kind, that every opportunity of acting in this manner is carefully to be sought after.

8 It is thus that the general rules of morality are formed. They are ultimately founded upon experience of what, in particular instances, our moral faculties, our natural sense of merit and propriety, approve, or disapprove of. We do not originally approve or condemn particular actions; because, upon examination, they appear to be agreeable or inconsistent with a certain general rule. The general rule, on the contrary, is formed, by finding from experience, that all actions of a certain kind, or circumstanced in a certain manner, are approved or disapproved of. To the man who first saw an inhuman murder, committed from avarice, envy, or unjust resentment, and upon one too that loved and trusted the murderer, who beheld the last agonies of the dying person, who heard him, with his expiring breath, complain more of the perfidy and ingratitude of his false friend, than of the violence which had been done to him, there could be no occasion, in order to conceive how horrible such an action was, that he should

https://doi.org/10.1017/S1744137409001325 Published online by Cambridge University Press

reflect, that one of the most sacred rules of conduct was what prohibited the taking away the life of an innocent person, that this was a plain violation of that rule, and consequently a very blamable action. His detestation of this crime, it is evident, would arise instantaneously and antecedent to his having formed to himself any such general rule. The general rule, on the contrary, (159/160) which he might afterwards form, would be founded upon the detestation which he felt necessarily arise in his own breast, at the thought of this, and every other particular action of the same kind.

9 When we read in history or romance, the account of actions either of generosity or of baseness, the admiration which we conceive for the one, and the contempt which we feel for the other, neither of them arise from reflecting that there are certain general rules which declare all actions of the one kind admirable, and all actions of the other contemptible. Those general rules, on the contrary, are all formed from the experience we have had of the effects which actions of all different kinds naturally produce upon us.

10 An amiable action, a respectable action, an horrid action, are all of them actions which naturally excite for the person who performs them, the love, the respect, or the horror of the spectator. The general rules which determine what actions are, and what are not, the objects of each of those sentiments, can be formed no other way than by observing what actions actually and in fact excite them.

11 When these general rules, indeed, have been formed, when they are universally acknowledged and established, by the concurring sentiments of mankind, we frequently appeal to them as to the standards of judgment, in debating concerning the degree of praise or blame that is due to certain actions of a complicated and dubious nature. They are upon these occasions commonly cited as the ultimate foundations of what is just and unjust in human conduct; and this circumstance seems to have misled several very eminent authors, to draw up their systems in such a manner, as if they had supposed that the original judgments of mankind with regard to right and wrong, were formed like the decisions of a court of judicatory, by considering first the general rule, and then, secondly, whether the particular action under consideration fell properly within its comprehension.

12 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 concerning what is fit and proper to be done in our particular situation. The man of furious resentment, if he was to listen to the dictates of that passion, would perhaps regard the death of his enemy, as but a small compensation for the wrong, he imagines, he has received; which, however, may be no more than a very slight provocation. But his observations upon the conduct of others, have taught him how horrible all such sanguinary revenges appear. Unless his education has been very singular, he has laid it down to himself as an inviolable rule, to abstain from them upon all occasions. This rule preserves its authority with him, and

https://doi.org/10.1017/S1744137409001325 Published online by Cambridge University Press

renders him incapable of being guilty of such a violence. Yet the fury of his own temper may be such, that had this been the first time in which he considered such an action, he would undoubtedly have determined it to be quite just and (160/161) proper, and what every impartial spectator would approve of. But that reverence for the rule which past experience has impressed upon him, checks the impetuosity of his passion, and helps him to correct the too partial views which self–love might otherwise suggest, of what was proper to be done in his situation. If he should allow himself to be so far transported by passion as to violate this rule, yet, even in this case, he cannot throw off altogether the awe and respect with which he has been accustomed to regard it. At the very time of acting, at the moment in which passion mounts the highest, he hesitates and trembles at the thought of what he is about to do: he is secretly conscious to himself that he is breaking through those measures of conduct which, in all his cool hours, he had resolved never to infringe, which he had never seen infringed by others without the highest disapprobation, and of which the infringement, his own mind forebodes, must soon render him the object of the same disagreeable sentiments. Before he can take the last fatal resolution, he is tormented with all the agonies of doubt and uncertainty; he is terrified at the thought of violating so sacred a rule, and at the same time is urged and goaded on by the fury of his desires to violate it. He changes his purpose every moment; sometimes he resolves to adhere to his principle, and not indulge a passion which may corrupt the remaining part of his life with the horrors of shame and repentance; and a momentary calm takes possession of his breast, from the prospect of that security and tranquillity which he will enjoy when he thus determines not to expose himself to the hazard of a contrary conduct. But immediately the passion rouses anew, and with fresh fury drives him on to commit what he had the instant before resolved to abstain from. Wearied and distracted with those continual irresolutions, he at length, from a sort of despair, makes the last fatal and irrecoverable step; but with that terror and amazement with which one flying from an enemy, throws himself over a precipice, where he is sure of meeting with more certain destruction than from any thing that pursues him from behind. Such are his sentiments even at the time of acting; though he is then, no doubt, less sensible of the impropriety of his own conduct than afterwards, when his passion being gratified and palled, he begins to view what he has done in the light in which others are apt to view it; and actually feels, what he had only foreseen very imperfectly before, the stings of remorse and repentance begin to agitate and torment him.

https://doi.org/10.1017/S1744137409001325 Published online by Cambridge University Press