Fair and Impartial Spectators in Experimental Economic Behavior

Vernon L. Smith[*] and Bart J. Wilson[†]
Economic Science Institute
Chapman University
February 2012

> If he would act so as that the impartial spectator may enter into the principles of his conduct…he must…humble the arrogance of his self–love, and bring it down to something which other men can go along with.
>
> Adam Smith, *Theory of Moral Sentiments*, II.ii.2.1, p. 83.

Introduction

Contrary to popular belief, Adam Smith did not argue, famously or infamously, that humans were primarily motivated by self-interest, as is quite explicit in the above quotation. Even in *The Wealth of Nations* (hereafter, WN), he spoke not of the self-interest but of one's "own interest" which includes prudence, but was always mediated by what "other men can go along with."[1] Smith renowned-ly says that "[i]t is not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own interest. We address ourselves, not to their humanity but to their self-love, and never talk to them of our own necessities but of their advantages" (WN, I.ii.2, pp. 26-7). But acting in one's "own interest" need not entail putting one's own interest *above* another's interest in commerce, which is what acting with self-interest quite fundamentally means then and now.

A deeper reading of WN reveals Smith's implied qualification of "own interest", for appealing to the self-love of the butcher, the brewer, and the baker means "allowing every man to pursue his own interest his own way, upon the liberal plan of equality, liberty and justice" (WN, IV.ix.3, p. 664). If that qualification is unpersuasive, he elaborates later when discussing competition: "Every man, as long as he does not violate the laws of justice, is left perfectly free to pursue his own interest his own way, and to bring both his industry and capital into competition with those of any other man, or order of men" (WN, IV.ix.51, p. 687).[2] Thus, if the modern economist espouses naked self-interest as the foundation for economic decision making, he or she does so incompatibly with the founding father of the discipline and more generally with the genius of the Scottish Enlightenment.

[*] Email: *vsmith@chapman.edu*
[†] Email: *bartwilson@gmail.com*
[1] Tellingly, Book 5 in Volume 2 is the first and last time Smith uses the word "self-interest" and then it is to describe "the industry and zeal of the inferior clergy [in Rome]" (p. 789).
[2] Mandeville, who irreverently founded economic decision making on the vice of self-love in his poem, *The Fable of the Bees*, and whose satirical, tongue-in-cheek humor scandalized Smith, nevertheless still concluded: "So Vice is beneficial found, / When it's by Justice lopt and bound" (1705).

Smith's friend, David Hume, likewise circumscribes market behavior within rules when he distinguishes interested commerce, what North (1990, 2005) calls impersonal (market) exchange, from disinterested commerce, or what North calls personal (social) exchange. In the 18th century, while the first meaning of *interest* is "concern, advantage, good", the fourth meaning, which applies here, is "regard to private profit" (Johnson, 1755). Hume recognizes that promises were invented for interested commerce to "bind ourselves to the performance of any action" (1740, 3.2.5, p. 335). While with disinterested commerce we "may still do services to such person as I love, and am more particularly acquainted with, without any prospect of advantage; and they may make me a return in the same manner, without any view but that of recompensing my past services," the same is not true of our impersonal intercourses. We precisely engage in mutually benefiting and impersonal exchange for the distinct prospect of a private profit, and we voluntarily do so only with promises, "the sanction of interested commerce of mankind" (Hume, 1740, 3.2.5, p. 335).

Our primary purpose in this essay is to draw upon the literature of classical liberal economy to show how it informs and is informed by results from experimental economics. In particular, we focus on disinterested commerce, which, like interested commerce, is circumscribed by rules. Johnson (1755) defines *disinterested* as "superior to regard of private advantage; not influenced by private profit". Importantly, *superior* can connote a sense of being "greater in dignity or excellence" (Johnson, 1755). Adam Smith's first great book, *The Theory of Moral Sentiment* (hereafter, TMS), serves as our chief source of insights for understanding and interpreting modern laboratory research in terms of the conventions that govern human conduct in personal exchange. At the same time, we wish to demonstrate how today's economic experiments elucidate a reading of Adam Smith.

Influenced by Newton and astronomy, Smith was concerned with the power of rule-governed systems to organize observations beneath human sensible awareness, and sought to develop such a system for the social foundations of morality.[3] His project in TMS is to acutely

[3] See his History of Astronomy (1795). This work was published posthumously; that it was written prior to 1758, the year before TMS was to be published is indicated by Smith himself in the text of Section (IV. 74) noting that Newton's followers have predicted the return of a comet in 1758, adding in a footnote that this statement had been written earlier, and that subsequently "the return of the comet had occurred agreeably to the prediction." Smith is referring to Halley's Comet that appears on schedule about every 76 years—a prediction whose confirmation was truly mind-bending for any remaining 18th century skeptics of Newton. Prior to the publication of his two books in 1759 and 1776, Smith was clearly enamored by the ability of Newtonian theory to provide an orderly account of observations from the physical world.

discern how our moral sentiments emerge out of human interactive experience to form a system of general rules that wisely orders society:[4]

> The general maxims of morality are formed, like all other general maxims, from experience and induction. We observe in a great variety of particular cases what pleases or displeases our moral faculties, what these approve or disapprove of, and, by induction from this experience, we establish those general rules (TMS, VII.iii.2.6, p. 319).

Smith, however, warned that in these maxims, arising "by experience and induction," we should never confuse their functional efficiency with their cause, i.e., the general rules from induction are not the consequence of applying reason or deliberate human design:

> In every part of the universe we observe means adjusted with the nicest artifice to the ends which they are intended to produce**…**But though, in accounting for the operations of bodies, we never fail to distinguish in this manner the efficient from the final cause, in accounting for those of the mind we are very apt to confound these two different things with one another. When by natural principles we are led to advance those ends, which a refined and enlightened reason would recommend to us, we are very apt to impute to that reason, as to their efficient cause, the sentiments and actions by which we advance those ends… (TIM, II.ii.3.5, p.87).

In this, Hume was in full agreement, as the rule of justice and of property "…arises gradually, and acquires force by a slow progression, and by our repeated experience of the inconveniences of transgressing it" (Hume 1740, III.2.2.10, p. 315).

Thus, as Smith and his intellectual contemporaries appreciated, "Man [both the individual and the species] is made for society" and the peace of that society depends upon morality (Ferguson, 1792, p. 199). Moreover, the rules of morality, as Hume explains, "are not *arbitrary*" (Hume 1740, III.2.1.19, p. 311), and from Smith: "Vice is always capricious: virtue only is regular and orderly" (TMS, VI.ii.1.18, p. 225). In the language of Hayek, the leading 20th century scholar the core of whose work continued in the Scottish tradition, we are speaking of spontaneous order mediated by the community-grown rules of interaction in small familial-like groups (Hayek, 1988, p. 18).

---

[4] The ramifications of this point are lost to those who thumb through TMS for quotations that justify *post hoc* their modern research. In a well-known thought experiment, Smith considers how a European "who had no sort of connexion" with China would respond to hearing that a dreadful earthquake had struck this remote land. Ashraf et al. (2005) use this section of TMS to discuss how "Smith argued that natural sympathy often falls short of what is morally justified by mass misery" (p. 134). But Smith isn't discussing moral justification in Part III. He is discussing "the Foundation of our Judgments concerning our own Sentiments and Conduct, and of the Sense of Duty" (p. 109). This is confirmed four pages later when Smith carefully explains that "[a]ll men, even those at the greatest distance, are no doubt entitled to our good wishes, and our good wishes we naturally give them. But if, notwithstanding, they should be unfortunate, to give ourselves any anxiety upon that account, seems to be no part of our duty. That we should be but little interested, therefore, in the fortune of those whom we can neither serve nor hurt, and who are in every respect so very remote from us, *seems wisely ordered by Nature*" (TMS, III.iii.9, p. 140, italics added). Smith is modeling the conduct expressed in our actions. In that model we are disciplined by judgments that focus our attention on issues where our actions can make a difference—serve or hurt—through our choices.

We report results from a variety of two-person laboratory experiments motivated originally by game-theoretic predictions. In these economic environments we see property rights, in the sense of rights and wrongs of taking certain actions. In personal exchange environments, these property rights are involved as mediators of choice; i.e., they emerge as conventions, or a form of mutual consent, that are recognized implicitly, or not, within the group by the interacting individuals, and determine whether cooperative outcomes are realized or not. In impersonal market exchange, these socially grown rights have become codified in externally imposed and enforced rules, defining an institution that governs exchange and outcomes. This insight into the social origins of property rights is captured in Hayek's quotation from Julius Paulus, a third century A.D. Roman jurist: "What is right is not derived from the rule, but the rule arises from our knowledge of what is right" (Hayek, 1978, p. 162).

Principles of Action in TMS

The arguments that follow make use of our interpretation of Adam Smith's theory of the mental and emotional states that serve to mediate the individual actions that produce those states; accordingly, we provide a very brief overview of these principles of action.

Humans desire and seek praise and praise-worthiness; also to avoid blame and blame-worthiness (TMS, III.2.1, p. 114). Praise and praise-worthiness are connected, but the latter is not derived from the former and the two are somewhat independent (TMS, III.2.2-3, p. 114). Thus praise yields little pleasure if, in ignorance or error, we judge it—via the impartial spectator—to be undeserved (TMS, III.2.4, pp. 114-115). Similarly, we find satisfaction in our praise-worthy conduct, even if no such praise is likely to be bestowed on us (TMS, III.2.5, pp. 115-116). In these passages it is important for economic readers to avoid thinking of words like "satisfaction" and "pleasure" as being equivalent to or yielding "utility," which for Smith meant merely and only "useful" (TMS, IV.1.6, p. 180). For Smith what was satisfying or pleasing was the conformance of our conduct with social propriety in choosing an action.

Concerning action in the self-interest, Smith followed the Stoics in arguing that "self-love" is recommended to all by the requirements of self-preservation, but its arrogant forms must at all times be humbled in order to pursue actions that conform to the judgments of one's impartial spectator (TMS, II.ii.2.1, pp. 82-83; VII.ii.1.15, p. 272).[5]

[5] Formally, we might think of an action taken by i as depending on its propriety, given the circumstances:

a_i (Propriety$|C$) = $\alpha_i(C)(PR) + \beta_i(C)(PR)\cdot(PW) + \gamma_i(C)(PW) + \delta_i(C)$,

The Impartial Spectator

Our actions are subject to a discipline of self-command by principles that operate through the metaphor of the "fair and impartial spectator," or simply the Impartial Spectator:

> We endeavour to examine our own conduct as we imagine any other fair and impartial spectator would examine it. If, upon placing ourselves in his situation, we thoroughly enter into all the passions and motives which influenced it, we approve of it, by sympathy with the approbation of this supposed equitable judge. If otherwise, we enter into his disapprobation, and condemn it (TMS, III.1.2, p. 110).

The words "fair," "impartial" and "equitable" were chosen, we believe, quite deliberately by Smith to represent judgment by a neutral referee as to whether an action was fair or foul under the applicable rules of interaction given the circumstances. Within Smith's metaphor of the Impartial Spectator is the sports metaphor of judgment under the rules of the game.[6] Smith repeatedly makes reference to actions that "other people" or "mankind," or the "impartial spectator," "can go along with" (or not). The Impartial Spectator constitutes an internalization of what is approved or not approved by others. We are encouraged to take actions that others can go along with, and deterred from actions that they cannot and find objectionable: others "always mark when they enter into, and when they disapprove of (our) sentiments." (TMS, III.1.3, p 110) This characterization of human sociality serves to mediate human action, however imperfectly.[7] As a social-psychological restraint it emerges first in our families, extended families, and friendship enclaves, but ultimately appears in the laws codified by civil society (TMS, VI.ii.Introduction,1, pp. 218-227; II.ii.2.2-3, pp. 83-85).

---

where *PR* and *PW* are (0, 1) indicator variables that an action deserves social praise (1), or not (0), and is praise-worthy (1), or not (0); and α_i, β_i, γ_i and δ_i are nonnegative functions. In the second term, *PW* adds leverage to *PR*, while the third term expresses the *TMS* sentiment that *PW* may yield stand-alone value even where it can never receive praise. *C* defines the circumstances—the game structure, including i's choice alternatives and their payoffs. Each action is based on conduct that is more or less satisfying or pleasing conditional on circumstances, and the action chosen is the one most satisfactory according to these socially mediated criteria. The term $\delta_i(C)$, independent of the social indicators, allows "self-love" to be part of the evaluation of action. This function is defined only on own payoffs. One implication is that where *i*'s information is limited regarding the choice and/or payoffs of other, then *i* cannot infer the intent of other and thereby reward beneficence, although she may still value her decision as praise-worthy; hence PR = 0, and $\delta_i(C)$ looms larger than otherwise in determining the choice. A formal treatment similar to the above would apply where blame and blame-worthiness were elements to be applied to the evaluation of some actions. Even where payoffs are large, self-love may be constrained by considerations of blame and blame-worthiness.

[6] For a discussion on "fair" as playing within the rules of social practice, see Wilson (2012), particularly footnote 7 which discusses the 18th century meaning of the word. Adam Smith's usage of "fairness" stands in sharp contrast to the interpretation and discussion in Ashraf et al. (2005, pp. 136-137).

[7] The Impartial Spectator is not, however, equivalent to our conscience because: "The word conscience does not immediately denote any moral faculty by which we approve or disapprove. Conscience supposes, indeed, the existence of some such faculty, and properly signifies our consciousness of having acted agreeably or contrary to its directions" (TMS, VII.iii.3.15, p. 326). Ashraf et al. (2005) miss this distinction in their reading of TMS when they explain that "[i]n social situations, the impartial spectator plays the role of a conscience" (p. 132).

The Impartial Spectator enters in two ways: Our judgments of the actions of others and judgments of, and actions by, ourselves. Propositions concerning our judgments of the actions of others include the following:

- Properly motivated beneficent actions alone require reward. Why? Because it is these actions alone that inspire our gratitude (TMS, II.ii.1.1, p. 78).
- Improperly motivated hurtful actions alone deserve punishment. Why? Because these actions alone provoke resentment (TMS, II.ii.1.2, p. 78).
- The want of beneficence cannot provoke resentment.[8] Why? Because beneficence is always free (voluntarily given) and "cannot be exhorted by force" (TMS, II.ii.1.3, p. 78-79).

In TMS the emotion of resentment has a central role in expressing disapproval and emerges in human social interactions, providing common experience and a consensual foundation for rights to take action in social groupings. Thus, resentment safeguards justice by provoking the punishment of an injustice already done to another, while protecting against injustice by deterring others who fear punishment if they commit a like offence (TMS, II.ii.1.4-5, pp. 79-80). Retaliation is a law of Nature that requires the violator of the laws of justice to feel that evil done to another; he who simply observes and does not violate the laws of justice merits no reward, but only respect for his innocence (TMS, II.ii.1.9-10, p. 82).

Judgments of, and actions by, ourselves are governed by the principles of approval (disapproval) of our own conduct:

- These reflect the judgments we apply to others as we endeavor to exchange, mirror-like, our perspective with that of others, and "To see oursels as ithers see us"[9] in which we imagine our conduct examined by any other fair and impartial spectator.
- We possess no other looking-glass with which to examine our own conduct.
- In this, each becomes as two persons—the first is the Impartial Spectator, the judge; the second is the agent, himself the person judged (TMS, III.1.2-6, pp. 109-113).[10]

Traditional Game Theory and Experimental Economics

[8] Thus, as we interpret it, if I pass an opportunity to trustingly benefit you this would or need not be cause for your resentment. But if I should accept the opportunity, and you take advantage of my trust, then I have just cause for resentment of your action.

[9] From Robert Burns, "Ode to a Louse." Burns, we should note, was born in the year TMS was published.

[10] "We suppose ourselves the spectators of our own behaviour, and endeavour to imagine what effect it would, in this light, produce upon us. This is the only looking-glass by which we can, in some measure, with the eyes of other people, scrutinize the propriety of our own conduct. If in this view it pleases us, we are tolerably satisfied" (TMS, III.1.5, p. 112).

Initially, many of the experimental game results were motivated by game theory; subsequently, experiments were designed to better understand why the initial results so often deviated from game-theoretic predictions. Hence, we begin with a simple reduced form representation of a game as in Sobel (2005). We then modify that framework with a formalization that we believe corresponds more accurately to the way Adam Smith constructed a process view of human sociality in TMS.

Suppose that individual $i = 1,\ldots,n$ selects an action, x_i, in a stage game to maximize $Z_i(\mathrm{x})$, where $x = (x_1,\ldots x_i,\ldots x_n)$ are strategy choices by n players:

$$Z_i(\mathrm{x}) = (1-d)u_i(x) + dV_i[H(x)], \tag{1}$$

where $1 > d > 0$ is the discount rate, $H(x)$ is the history of play, u_i is i's self-loving "utility" outcome from the choice x_i in the stage game, and V_i is the value to i of continuation of play. (In the discussion below our examples are for $n = 2$ persons.)

$Z_i(\mathrm{x})$ is interpreted as the criterion of judgment for decision making by i in a single sequential repetition of the same stage game with the same well-identified other. $Z_i(\mathrm{x})$ is described as i's discounted current plus future utility in a pairing created by the experimenter. Hence, $H(x)$ includes all past history, as well as the shadow of i's anticipated future history of play with other. As described by Sobel (2005):

> Repeated-game theory incorporates strategic context, not by changing preferences but by changing the way people play. In order to obtain equilibria distinct from repetitions of equilibria of the underlying static game, the history of play must influence future play. History does not influence preferences, but it does influence expectations about behavior (p. 412).

To achieve this, actions may take the form of punishments and rewards, contingent on actions by other that shape the self-loving behavior of other so as to enable i to maximize her long term self-loving interest over the horizon of the repeated game.

In this development, V_i is an endogenous function of the history of play. If V_i is positive and d is sufficiently large (near enough to 1), then in maximizing $Z_i(\mathrm{x})$, i must take care not to spoil her self-loving future interaction with this particular other person by her choice in the present—a care that in traditional repeated game theory exhausts the content of actions that are social; i.e. her sociality is defined and confined relative to her historical and anticipated future interactions with the particular person with whom she has been paired.

In game theory, repetition is essential for long term strategic success in achieving cooperative results, but laboratory experiments have long recorded significant levels of cooperation in single plays of a stage game in which the anonymous players forego larger payoff for themselves in favor (or expectations) of a cooperative outcome. Therefore, as noted by Sobel (2005, p. 411), "[b]ecause laboratory experiments carefully control for repeated-game effects, these results need a different explanation." That is, in a single play of the stage game a rational *i* is *assumed* to set $V_i = 0$ when matched with an unknown other person and therefore is presumed to be a "stranger" who person *i* cannot identify and thereby build on any relevant past personal history. Hence, both *i* and the other are predicted to choose self-loving dominant outcomes, whatever the circumstances defined by the game. The "different explanation" commonly offered for experimentally observed cooperative outcomes is the postulate of other-regarding or "social preferences" that rationalize the observed behavior by each player attributing own utility (or envious disutility) to money assigned to other, as well as money assigned to one's self in a single play of the stage game.

In this explanation any generosity, positive or negative, has been accounted for by simply augmenting *post hoc* the decision maker's utility function in an appropriate way. If this methodology is accepted—it has been widely adopted by theorists and experimenters since the predictive failures of game theory started to accumulate—the scientific conversation stops, along with inquiries directed to an understanding of why and how these external preferences serve the human career.

Adam Smith carefully and thoughtfully modeled human interactions of this kind, not as governed by own versus other utilitarian considerations, but by *conduct*—rules conditioned by propriety.[11] In following these principles the individual is pleased by the actions driven by her self-judgment, but "pleased" does not map into a utilitarian reward. Even if one can identify a formal case-by-case technical equivalence between outcome utilities and actions motivated by conduct rules, following such mechanical curve fitting involves an omitted essential step, and risks failing to articulate a process that disciplines our understanding of how and why context matters in games and life.[12] Adam Smith, who believed TMS was his most important work,

[11] Wilson (2008, 2010) uses the insights of Wittgenstein to make the related point that rules of conduct cannot be represented by utilitarian preferences, but are rather embedded in language games, the lifelong intercourse that each of us has with the rest of humankind. The Impartial Spectator is Adam Smith's version of that intercourse with oneself.

[12] Evidence of the failure of utilitarianism is prominent in the ubiquitous observation that varying payoffs for a given context matters much less than varying the contextual circumstances given payoffs. See Camerer (2003, pp. 60-61) for a report of the minor effects on ultimatum game outcomes of varying the stakes by factors of 10 and much higher; and Falk et al. (2007) for an examination of the importance of intentions. In Hoffman et al. (1994), ultimatum game choices vary significantly with circumstances, whereas in Hoffman et al. (1996) a tenfold increase

provides a meaningful systematic approach to experimental testing as an alternative to extending utilitarianism.

In (1), if *H* is "history," decision must be informed by one's entire cultural and past social experience, and the exploration of this social experience may expose thinking to non preference-based forms of other-regarding behavior. In this development, actions are only intelligible in reference to moral judgments of own and other actions in past and anticipated future interactions. What is important about the actions is the conduct (including intentions) they signal and not merely the outcomes the actions yield.

Such a pathway is provided by Smith's program in TMS. That pathway includes not only a continuation value, that we will now call $W_i[H(x)]$ where the stage game is to be repeated, but also sympathetically modifies the self-loving first term, $u_i(x)$, in equation (1). Moreover, W_i is now based on expected future conduct, both own and other, and not only on outcomes.

In TMS, individuals are motivated to seek praise and praise-worthiness, and to avoid blame and blame-worthiness, in all social interactions. And in judging her own conduct, a person i will always imagine that conduct as being examined by a fair and impartial spectator. Her actions will vary with circumstances, based on past experience, but require that her conduct serve personal long term (reputation) ends across a wide variety of human social encounters. When she knows little of a particular other she may be cautious, and more preserving of immediate Stoic care for herself, but, even so, she knows it is another human, recruited from a group whose characteristics may not be that dissimilar from her own, and she relies on self-command principles that have served her well on average in the past. Her action x_i will generate a current value that we will designate $U_i\ [x \mid H_i(0)]$, where $H_i\ (0)$ is her current entry-level personal historical state (after reading the instructions of the experiment). U_i values *i*'s conduct in taking immediate action x_i; part of that valuation is the resulting payoffs. But the value attained is derived from the judgment of the Impartial Spectator as to the propriety of her action, albeit including that the payoffs *are deserved and justified by the circumstances*.

That $U_i\ [x \mid H_i(0)]$ alone captures baseline elements in Smith's criterion for weighing the present against the future by a prudent person, under the self-commanding judgment of the Impartial Spectator, seems plainly evident in the following quotation:

> …in his steadily sacrificing the ease and enjoyment of the present moment for the probable expectation of the still greater ease and enjoyment of a more distant but more lasting period of

---

in payoff levels yields an insignificant effect on choices. Yet, these games have been ritualistically modeled by attempts to refit explanatory utility functions to the shifting circumstances recorded by experiments.

> time, the prudent man is always both supported and rewarded by the entire approbation of the impartial spectator, and of the representative of the impartial spectator, the man within the breast. The impartial spectator does not feel himself worn out by the present labour of those whose conduct he surveys; nor does he feel himself solicited by the importunate calls of their present appetites. To him their present, and what is likely to be their future situation, are very nearly the same: he sees them nearly at the same distance, and is affected by them very nearly in the same manner. He knows, however, that to the persons principally concerned, they are very far from being the same, and that they naturally affect *them* in a very different manner. He cannot therefore but approve, and even applaud, that proper exertion of self-command, which enables them to act as if their present and their future situation affected them nearly in the same manner in which they affect him (TMS, VI.i.11, p. 215).

Instead of equation (1) we now have a sympathy-derived criterion of action

$$S_i(x) = (1-d)\, U_i\, [x \mid H_i(0)] + d\, W_i[H(x)] \qquad (2)$$

If $W_i = 0$, as in an advertised one-shot game, max $S_i(x)$ does not reduce to max $Z_i(x)$; that would occur only for an *i* raised in isolation from all contact with other humans, or who is otherwise barren of all socialization: "To a man who from his birth was a stranger to society, the objects of his passions, the external bodies which either pleased or hurt him, would occupy his whole attention" (TMS, III.1.3, p. 110).

With $W_i > 0$, equation (2) allows action to accommodate the knowledge that the interaction will be repeated, and thereby enables the relationship with other to be influenced by possible futures that the two are able to create beyond the self-command principles that would apply to a single encounter which already contains baseline considerations of futurity as in the above quote from TMS. Under repetition, judgments by the Impartial Spectator of each person in their shared interaction will be updated based on how each reads the intentions conveyed sequentially by the other.

Ultimatum Games

In this game people are recruited to the lab in groups, say of 12, and are randomized into pairs, and at random one person is selected to be the Proposer, the other the Responder. The task of each pair is to determine the allocation of a fixed sum of money, *M*, say $10 or $100 (consisting of ten $1 bills or ten $10 bills) between them, under the following rules: The Proposer chooses an amount, y for herself, with the understanding that $M - y$ is allocated to the Responder. Play then passes to the Responder, who either accepts the allocation, in which case the indicated payments will be made to each, or he rejects the allocation, in which case

each receives zero from the interaction.[13] The subgame perfect equilibrium (SPE) of the game is for the Proposer to offer $1 (the minimum unit of account), and for the Responder to accept. The latter should accept any amount that is better than zero, and, in awareness of this, the Proposer offers that amount. The data tend to show very high rejection rates of $1, and rejections of amounts up to $3 are not uncommon. But Proposers tend to anticipate this behavior and very few offer low amounts. In experiments described as a "Divide $M" game the average offer is commonly about 45% of *M*, but offers change substantially with variations in the context and instructions (Smith, 2008; Camerer, 2003).

The first thought, for those schooled in TMS, might be that this behavior suggests that the Impartial Spectator of each player is at work evaluating the propriety of their actions. But this is a strange game to consider as a test of the propositions from TMS summarized above. Smith informs us emphatically that, "Beneficence is always free, it cannot be extorted by force, the mere want of it exposes to no punishment; because the mere want of beneficence tends to do no real positive evil" (II.ii.1.3, p. 78). Rethinking the Ultimatum Game in this light, we can say:

- As in most lab experiments, people are recruited to the lab not knowing the experiment that is to occupy them.
- They arrive and are not offered a choice between alternative experimental games.
- These procedures are carefully designed to control for self-selection bias, but as we see, other conditions may be inadvertently controlled for.
- These procedures, however, are hardly sacred: the first rule of any experimentalist should be that the experiment and its design be relevant to its purpose. One should backward induct from the purpose, and the question, to the design of the experiment.
- Playing the ultimatum game does not constitute a voluntary action. Have we gathered much data on pairs of "reluctant duelists," without this being part of our intention?
- Borrowing Adam Smith's words, should we not think of the Ultimatum Game as an "extortion game." The Proposer under the terms of his participation must decide on *y*, with *M* – *y* awarded to the Responder. Is her choice motivated by beneficence? Is the Responder rewarding beneficence by his acceptance of *M* – *y*? Is he punishing "want of beneficence" by rejecting it?

[13] The Ultimatum Game originated with Guth et al. (1982) and has spawned a vast literature. See for example, Forsythe et al. (1994), Hoffman et al. (1994), Hoffman et al. (1996) and for a partial survey, Camerer (2003, pp. 48-59).

- The circumstances of the game—to which the Impartial Spectator must always be sensitive in the light of past experience—are such that our answers to these questions are surely, "No," or at least "Mmm." From the perspective of TMS this is a mixed motive game.

These considerations cannot be dismissed with the convenient *ex post* argument that "in many situations one must play a game, even against one's wishes."[14] Rather the question is whether or not it is useful to think about the ultimatum game from a broader perspective, such as that in TMS, for there are experiments showing clearly that it matters how one arrives at the circumstance of deciding on a take-it-or-leave-it offer. Salmon and Wilson (2008) is a case on point. In their experiment motivated by observations on eBay, they embed an ultimatum game in a context of multiple buyers competing for purchases from a single seller. The seller has two units of the same good for sale, the first of which is auctioned off to the highest bidder in a typical English (ascending price) auction. For the second unit, the seller then makes a take-it-or-leave-it offer to the bidder with the highest losing bid (the second highest bidder). If the buyer accepts, she receives a profit equal to the difference between her randomly drawn value and the seller's offer. If she rejects, neither the seller nor the buyer earn anything on that unit.

Salmon and Wilson find that in a treatment with only two competing bidders, only 12 out of 273 offers (4.4%) are rejected. Moreover, 93 of those offers are greater than the buyer's final (but losing) bid, and only 6 of those are rejected. In other words, the seller is attempting to extract even more surplus out of the buyer and the buyers still do not reject the offers. With four bidders, 111 profitable offers are made to the bidders and only 4 (3.6%) are rejected. But here's the kicker. The median accepted surplus is a mere 61¢ and 39¢ in the two- and four-bidder treatments, respectively. In contrast, Hoffman et al. (1994) find that 10.4% of all offers are rejected, usually for amounts of $2 and $3, even when the ultimatum game is framed as a one-shot buyer-seller negotiation over a price.[15]

Why are the Salmon and Wilson results so strikingly different relative to the standard ultimatum game? Because the ultimatum game over the second unit is not a game of extortion mixed with beneficence from receiving a windfall. The second unit is a game of prudence with an immediate prior history, and the context that invokes the virtue of prudence is distinct from those that call for the virtues of beneficence or justice (TMS, Part VI). There is no open-ended question as to whether the seller is being beneficent enough with his offer to the buyer

[14] The quotation is from Ellsberg (1956) who notes that minimax strategies were not satisfactory solutions to zero-sum games, because if that were the solution to playing the game, and a person had the option to refuse play, then "[h]e would never play" (p. 922).

[15] Hoffman et al. (1996) report rejections of $30 offered from stakes of $100; List and Cherry (2000) report rejections of offers of $100 where the stakes are $400.

because she's not beneficently splitting a windfall with the buyer. She's prudently attempting to sell the second unit of a commodity to a buyer who couldn't pay as much as some other buyer for the first unit. Unlike the traditional ultimatum game, we observe that there's simply no beneficence to assess in a seller's take-it-or-leave-it offer.

Likewise, there is also no room for resentment of the seller's offer, for "[r]esentment seems to have been given us by nature for defence, and for defence only" (II.ii.i.4, p. 79). In a reluctant game of extortion, a Proposer may go too far in extracting money from the windfall and thus an offer of $2 may "prompt us to beat off the mischief which is attempted to be done to us, and to retaliate that which is already done" (II.ii.i.4, p. 79). But in the Salmon and Wilson markets, where is the mischief on the part of the seller? The buyer has just demonstrated he is unwilling to name and pay a price as high as someone else and in the process he has revealed approximately how much he is willing to spend. So, when faced with take-it-or-leave-it, the buyer takes it nearly every time. Notice, in comparing observations from the two different experimental designs, that the process is governed by "fairness" in the sense of the rules of conduct given the circumstances, not whether the outcomes are fair.

Pecorino and Van Boening (2010) embed the ultimatum game in the context of a litigation dispute. A plaintiff and a defendant are bargaining over how to split the cost savings of not going to trial, $0.75 to the plaintiff and $0.75 to the defendant. To avoid this cost, the defendant makes a pre-trial settlement offer to the plaintiff. If the plaintiff accepts the settlement offer, neither incurs the trial costs. The plaintiff receives the offer as payment and the defendant incurs the cost of his wrongdoing (which is subtracted as a lump sum given to him by the experimenter). If the plaintiff rejects the offer, then the plaintiff receives a judgment from which the trial costs are subtracted, and the defendant incurs the trial cost and the cost of judgment. In the baseline comparison treatment, a Proposer and a Responder play a traditional ultimatum game with M = $1.50. Both versions are repeated for 10 rounds.

In the embedded game, the median offer by the defendant is 8% of $1.50, or 12¢. In Pecorino and Van Boening's replication of the traditional ultimatum game, the median Proposer offer is 50% of $1.50, or 75¢. For similar offers of 0-25¢, 23% of the offers are rejected in the litigation game and 100% in the traditional game. Thus, defendants offer less and plaintiffs accept more often than their counterparts in the traditional ultimatum game. How does the TMS framework help us understand this? In the litigation game, the motives are no longer mixed. The proposing defendant is attempting to avoid a loss with an offer to the plaintiff which corresponds to the plaintiff avoiding the cost of a trial. While the experimenter has thrown them into a dispute, albeit an unavoidable one (which might explain the high rejection rates of 21-25%), mutually avoiding a cost is not a matter of beneficence on the part

of the defendant. In the litigation game, prudence in the form of accepting an offer equal to her opportunity cost is a virtue for the plaintiff, and not a matter of how beneficent the defendant is in his offer. Regardless of what happens, the defendant is minimizing the depletions from his upfront windfall.

Trust Games: Single Play

Consider the following two-person game commonly studied by experimental economists in a variety of forms and summarized in Figure 1. Person 1 chooses to either (a) end the interaction sending each person on their way with an additional $10 or (b) forego his sure $10 and turn the decision making over to Person 2. If Person 1 chooses (b), then Person 2 decides between (a') the experimenter paying her $25 and Person 1 $15 or (b') the experimenter paying her $40 and sending Person 1 on his way with nothing by way of the outcome from the interaction in this game.[16]

If Person 1 is fully aware of the choice that Person 2 faces, and vice versa, how do we understand what two anonymous people do when faced with this situation? Adam Smith notes that unless the situation calls for a rule of justice "our conduct should rather be directed by a certain idea of propriety, by a certain taste for a particular tenor of conduct, than by any regard to a precise maxim or rule" (III.6.10, p. 175). If that sounds fairly "loose, vague, and indeterminate" (III.6.11, p. 175), then that is because "there are no rules by knowledge of which we can infallibly be taught to act upon all occasions with prudence, with just magnanimity, or proper beneficence" (III.vi.11, p. 176). Consequently, Smith implicitly recognizes here that the rule a particular individual might follow can be expected to vary with the circumstances that constitute particular "occasions".

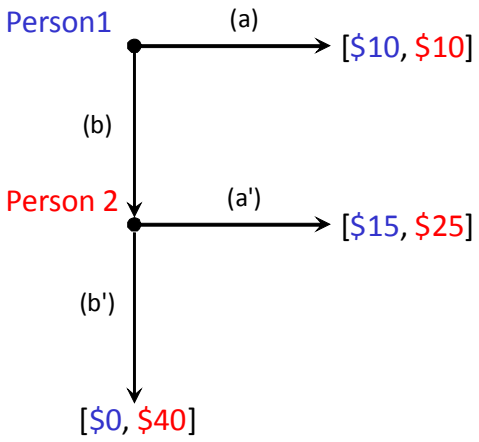


Figure 1. A Two-Person Trust Game in Extensive Form

[16] Experimentalists commonly pay subjects a fixed show-up payment when they arrive, that is for each person to keep whatever the outcomes of the subsequent experiment.

This rules out as being pertinent for all occasions the modern economist's rather precise and accurate concept of subgame perfect equilibrium, which predicts that Person 1 would immediately end the game and receive $10 because, if given the opportunity to make the decision, Person 2 would choose $40 over $25 for herself, thereby leaving Person 1 with nothing. Fortunately, "[n]ature, … [has not] 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" (III.4.7, p. 159).

What general rules of fit and proper behavior are applicable to this game and to the experiences of this community of participants? And what would the rules predict? Let's first consider, as subgame perfection does, Person 2. If given the opportunity to make a decision, Person 2 would "endeavor to examine [her] own conduct as [she] imagines any other fair and impartial spectator would examine it. If, upon placing [herself] in his situation, [she] thoroughly enter[s] into all the passions and motives which influenced it, [she] approve[s] of it, by sympathy with the approbation of this supposed equitable judge. If, otherwise, [she] enter[s] into his disapprobation, and condemn[s] it" (III.1.2, p. 110).

In this game the question is whether, by sympathy with the impartial spectator, would Person 2 approve or disapprove of choosing (a') and approve or disapprove choosing of (b'). Choosing (a') yields a higher payment from the experimenter to both individuals as Person 1 forewent a sure $10 for both. A fair and impartial spectator could thus approve of (a'); both are better off because of the actions of Person 1 and Person 2. Choosing (b'), however, sends Person 1 home with nothing after foregoing a sure $10. In light of (a'), Person 2 is better off, regardless of what she does, because Person 1 passed the play to her. Thus however anonymous the participants may be in this interaction, an impartial spectator could reasonably disapprove of (b'). Now consider Person 1. From past experience with friends and classmates, he expects that "[n]ature, which formed men for that mutual kindness, so necessary for their happiness, renders every man the peculiar object of kindness, to the person to whom he himself has been kind" (VI.ii.1.19, p. 225; hereafter, Principle of Reciprocal Beneficence). In other words, experience has taught him that if he kindly passes the play for a mutual gain for the both of them, a Person 2 may kindly reciprocate him, the person to whom he himself has just been kind.

But must the impartial spectator disapprove of (b')? Not necessarily, if our conduct is indeed directed by a certain idea of propriety and not a precise rule. Recall that Person 1 has the choice of (a) or (b), and if Person 1 chooses (b), Person 2 has the choice of (a') or (b'). An

impartial spectator could reason that the experimenter's rules are the rules, and everyone, including Person 1, knows the rules and has agreed to participate in this experiment. Thus, if Person 1 willingly chooses (b) an impartial spectator could also approve of (b'), for if the experimenter did not wish to observe whether or not Person 2 might actually choose (b') the experimenter would not have given her the option.

TMS thus informs the experimental economist that the rules of interaction in the trust game merely "present us with a general idea of the perfection we ought to aim at, [rather] than afford us any certain and infallible directions for acquiring it" (III.6.11, p.175-6), and this general idea of the perfection is founded upon our autobiographical experiences "of what, in particular instances, our moral faculties, our natural sense of merit and propriety, approve, or disapprove of" (III.4.8, p. 159). Different people, either with different experiences, or different interpretations as to how their experience applies to the game in question, may converge on different responses, especially in a one-shot game.

In the laboratory the replicable facts from three different studies are that of 98 first movers, 52 choose (a) and 46 choose (b), and that of the 46 second movers who have the opportunity to make a decision, 31 (67%) choose (a') and 15 (33%) choose (b') [McCabe and Smith, 2000; Cox and Deck, 2005; and Gillies and Rigdon, 2008]. So while TMS modestly makes no specific prediction about what people will do in the trust game,[17] experimental economics can inform Smith's theory of the general principles with which impartial spectators approve and disapprove of (a), (b), (a'), and (b'). By randomly assigning participants to conditions with systematic variations in the procedures, we can trace out the contextual principles that excite and mediate whether more impartial spectators approve or disapprove of (a), (b), (a'), and (b').

Typically in a laboratory experiment, subjects make decisions anonymously with respect to each other, but the experimenter knows by name what each subject did so as to pay them (privately) what they earn. This is the protocol for the data reported above. In a second condition, Cox and Deck (2005) implement an elaborate procedure to ensure that the subjects also make their decisions anonymously with respect to the experimenter. The experimenter cannot match decisions to specific individuals. Interestingly, this change in procedures asymmetrically effects the decisions of Persons 1 and 2. First movers pass the play by choosing (b) at the same rate in both conditions. However, 10 out of 14 (71%) second movers choose (b') with double anonymity but only 8 out of 25 (32%) choose (b') with single anonymity. It

[17] The critic who asserts that a Smithian analysis of this game is unhelpful because it does not make a specific prediction has the burden of providing and demonstrating a set of rules for this interaction that are, in the words of Adam Smith, "precise, accurate, and indispensable" (TMS, III.6.11, p. 175). When the experimental games on which we are reporting first began to be studied in the 1980's, the predictions of game theory performed very poorly.

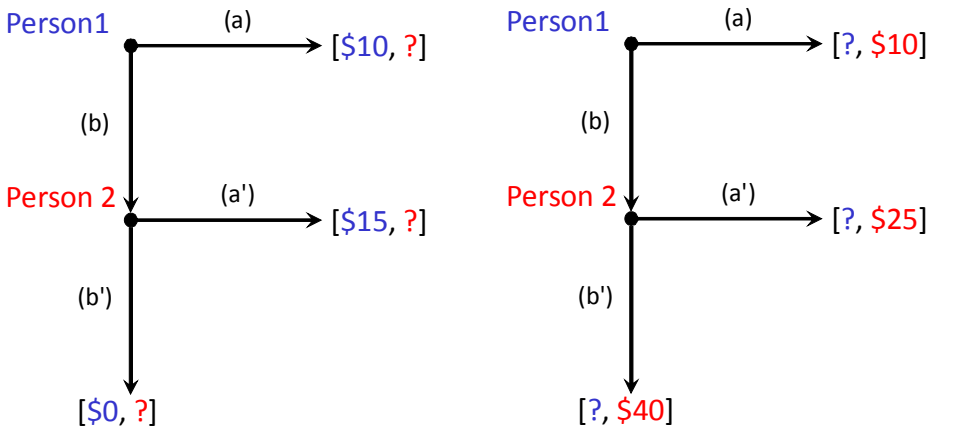
seems that increasing the private character of the interaction is one aspect of the context that excites more impartial spectators to approve of (b'). An unresolved question is why Person 1's do not anticipate that Person 2's are more disposed to choosing (b') over (a') with double anonymity.[18] Hence, empirical support for Smith's Principle of Beneficent Reciprocity is strong under single, but not double anonymity; it seems important whether or not people other than your matched counterpart can know your behavior.

Gillies and Rigdon (2008) consider how knowledge of the payoffs affects the play of Persons 1 and 2. In what they call a "Private Game," each person only knows their own payoffs associated with (a), (b), (a'), and (b'). As shown in Figure 2, Person 1 only knows that he receives $10 from choosing (a) and that if he passes the play, Person 2 is choosing between $15 and $0 for him. The catch is that Person 1 does not know what Person 2's payoffs are from choosing (a') and (b') and Person 1 knows that Person 2 does not know what his payoffs are from choosing (a') and (b'). Likewise, Person 2 does not know what Person 1's payoff is from choosing (a), only that her payoff is $10 from Person 1 choosing (a).

Without knowledge on how his decision affects Person 2, Person 1 is unable to conclude from past experience that Person 2 will reciprocate a trusting action of (b) with a trustworthy one of (a'), and that is what Gillies and Rigdon observe. Fifteen of 45 (33%) first movers play down in the "Private Game" as opposed to 21 of 50 (42%) first movers do in the full common knowledge game.

More dramatic is the response of Person 2's impartial spectators. Only 3 of 15 (20%) second movers play (a') in the "Private Game" in contrast to 14 of 21 (67%) who do so in the full common knowledge game. More impartial spectators approve of (b'), taking the higher payoff of $40, when they are unaware of what Person 1 forewent in choosing (b) and unaware of what Person 1 will receive ($0). Since neither player knows the payoff of other, the sentiments of praise and praiseworthiness, and the Principle of Beneficent Reciprocity, cannot enter into judging the propriety of each other's actions; hence their self-love cannot be "humbled" by the Impartial Spectator and is necessarily more important under such game circumstances.

[18] Person 2's conduct in choosing (a'), under double anonymity, may be merely praise-worthy and thus weakened in conduct value compared with single anonymity; similarly, Person 2's choice of (b') may be less discouraged by being merely blame-worthy compared with single anonymity. Any such second order effects may be more difficult for Person 1's to anticipate.



Panel (a). Person 1's View of the Game Panel (b). Person 2's View of the Game

Figure 2. Private Knowledge of Payoffs in the Trust Game

In the complete knowledge version of the game in Figure 1, Gillies and Rigdon also consider in a separate treatment condition how Person 2's behave when they are asked to make their decision assuming that Person 1 has chosen (b). Person 2's, however, are only paid based upon those decisions if Person 1 actually chooses (b). If Person 1 chooses (a), then Person 2's choice is not implemented. In this treatment the impartial spectators are hypothetically invoked as opposed to being explicitly excited with Person 1's actual choice of (b). Whereas 14 of 21 (67%) second movers choose (a') when Person 1 has actually chosen (b), only 20 of 43 (47%) Person 2's choose (a') when asked to assume Person 1 has chosen (b).[19] The distinction made in these experiments correspond to games played in extensive versus normal (or strategic; i.e., contingent play) form. Traditional game theory treated the two as equivalent, but many experimental studies have reported data rejecting this postulated equivalence.[20] The two game forms are cognitively much different in that in the extensive form Person 1 conveys to Person 2 her intentions before the latter is required to choose. TMS is particularly relevant in this interpretation because intentions are central to the capacity of the Impartial Spectator to form an appropriate judgment of the other person's action, and therefore in judging an appropriate response.

Trust Games: Repeat Play

Figure 3 presents another simple trust game that has been used to study single as well as repeat play versions of the same basic stage game. In single play, if Person 1 chooses to end the game, each receives $20; if Person 1 passes to Person 2, the latter chooses between (a') $25 for each, or (b') $15 for Person 1 and $30 for Person 2. As in the first trust game above (Figure

[19] Casari and Cason (2009) observe similar behavior in a trust game with different parameters.

[20] For a discussion and several references, see Smith (2008, pp. 264-267) and for earlier experiment results see McCabe et al. (1996).

1), the SPE is for Person I to end the game and each leaves with $20 apiece, but in the laboratory we observe 63% passing to Person 2. And twice as many people in the Person 2 position (65%) choose (a') over (b'). As before, both persons are choosing cooperatively in a manner consistent with the Principle of Beneficent Reciprocity in *TMS*. McCabe et al. (2003) use this game to answer the following question: How will these results be affected if in a second treatment condition, Person 1 cannot voluntarily choose between ending the game and passing to Person 2, with passing being required of Person 1? Person 2 faces the same alternatives as before, but sees that Person 1 gives up nothing. Consequently, under these conditions, the Impartial Spectator in Person 2 is prevented from forming the same intentional "kindness" judgment of the conduct of person 1 as in the first treatment. Consistent with this reasoning, under the second treatment conditions the results from the first experiment are reversed: now only 33% of the Person 2's choose (a') over (b').[21]

Rigdon et al. (2007) have also studied behavior in repeat play of the stage game in Figure 3. Their experiments examine decision behavior under two different conditions that vary only the protocols for matching subject pairs after each round of play. In both protocols, the subjects are not informed as to the number of repetitions; without warning, play is stopped after 20 rounds. In the first protocol the subjects are simply re-paired at random. In the second a scoring algorithm uses their previous decisions to enable all Person 1's and Person 2's to be separately rank ordered from most cooperative to least. The highest in each rank are then matched with each other for the next round; the second highest are matched with each other for the next round, and so on down the list. A cooperative choice by Person 1 means that she passed to Person 2; a cooperative choice by Person 2 occurs whenever option (a') is selected. It is very important to keep in mind that the subjects in these experiments *were not informed of the matching procedure.* In both treatments all the participants were told simply that they would be re-paired with a person in the room each period. In all sessions there were 16 people in the room with 8 Person 1's (and 8 Person 2's) to be re-paired either at random, or by application of the scoring algorithm.

If indeed "kindness begets kindness" as in Adam Smith's Principle of Beneficent Reciprocity, then the scoring rule allows those interacting over the 20 repetitions to "discover" by experience that they are in an environment characterized by "kindness." Over time each person's Impartial Spectator would be updated and reflect any experiential tendencies toward kind behavior. Rigdon et al. (2007) had no assured expectation as to how effective the scoring rule would be. This is why they used a comparison control that implemented random re-pairing.

[21] But remarkably, many Person 2's still choose to be generous to Person 1's perhaps leaving ample room for the TMS sentiment of acting in a praise-worthy manner even without the implied praise when kindness is returned by kindness.

An open question was how effective the two protocols would be in separating the two different pools of subjects with respect to their frequency of cooperative choice.[22]

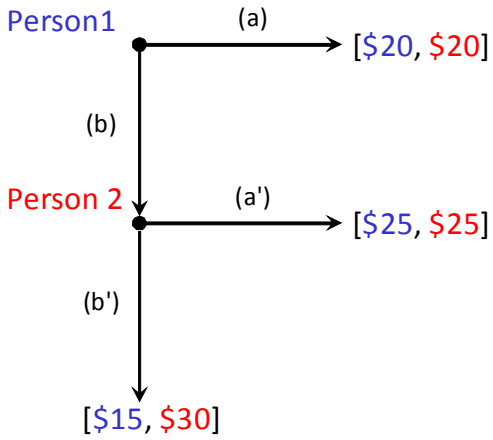


Figure 3. Another Two-Person Trust Game in Extensive Form

The data show that the primary research hypothesis was strongly supported as the two treatment groups bi-furcated significantly across repeat trials in exhibiting cooperative responses: On trials 1-5, the ratio of percent cooperative choice by Person 1's in the treatment to the percent cooperation in the random control was 1.05; for Person 2's, the ratio was 1.10; i.e., essentially very little treatment difference in the first five trials. But cooperation steadily improved, so that in the last five trials, 16-20, these ratios respectively were 1.94 and 1.63, corresponding to increases respectively of 1.94/1.05 = 185% for Person 1's and 1.63/1.10 = 150% for Person 2's. The latter are less than the former because regardless of treatment, Person 2's, experiencing the largess of Person 1's, tend to honor the principle that "Actions of a beneficent tendency, which proceed from proper motives, seem alone to require reward;

[22] The research reported in Rigdon et al. (2007) was done at the University of Arizona at the turn of the millennium, appearing as a working paper in 2002, and was delayed in publication. Why? Principally, the procedure—subjects not being informed of the rank order rule for re-matching pairs—was the source of many explanations and discussions with seminar participants and in the editor-refereeing process. Many had difficulty grasping why we did not make the comparison with subjects given full knowledge of the cooperative matching procedure. There is a body of constructivist economic theory—irrelevant and distractive from the perspective of this study—that argues that a small in-group of cooperators can invade a population of defectors, and being able to identify each other, outperform their out-group peers. Yes, and if our subjects knew the circumstances of their matching and we observed more cooperation than in the randomly re-paired group what would we learn? Only, we fear, that, when it is made plain to people that in repeat interaction cooperation is individually optimal, then people are likely to choose optimally. In that case we would learn yet again that in games that essentially reduce rationally to games against nature, people tend to go to the top of the profit hill. If this exercise is to be meaningful, the question must be what will people do if they find themselves—without knowledge of why—in a climate of relative cooperation, compared to a climate of relative defection? Will cooperation and profitability build experientially and "insensibly" in the former à la TMS, or will it deteriorate in attempted mutual exploitation à la game theoretic self-loving behavior?

because such alone are the approved objects of gratitude, or excite the sympathetic gratitude of the spectator" (TMS, II.ii.1.1, p. 78).

Rigdon et al. (2007) also report a very pronounced regularity in the behavior of people in both treatments: the individual decisions of Person 1's to trust or not, and for Person 2's to respond trustworthily or not, *on the first trial* was strongly and significantly related to their subsequent tendency to show trust or trustworthy behavior in repeat interaction. Thus, in equation (2) we can say that in these experiments, each person after reading the instructions, and entering into the first round of play, makes a decision conditional upon her history, $H_i(0)$, and her anticipated future interactive behavior, $H(x)$. What we learn across all the subjects is that her sympathetic state is marked indelibly by her first decision, and is predictive of her subsequent behavior in the remaining 19 trials. In the language of game theory, she is "typed" by her decision on the first trial, and her type significantly accounts for her subsequent decisions although these vary significantly with her subsequent experience and the experimental treatment condition.

Concluding Remarks

Adam Smith's *Theory of Moral Sentiments* is much more than a source of ornamental quotations for modern research in economics. TMS is a primary source of insights for understanding what modern, logico-deductive economics cannot account for—our human passions and motives, the edifice upon which our morality is built. Adam Smith is a theorist in the original sense of the Greek word *theoria*, meaning "to view or behold". He importantly begins, not ends, with acute observations on everyday human intercourse qua axiom, which he then organizes as elements in a rule-governed system of morality. Rules of conduct, not outcomes, are the focus of his analysis. Adam Smith uses the word *society* 157 times in the TMS, roughly once every other page. Why? Because his overarching concern with understanding human rules of conduct is how, in an ever-fluxional world, society orders itself via morality, which is "indeed the result of human action but not the execution of human design" (Ferguson 1767, p. 102).

References

Ashraf, Nava, Colin Camerer, and George Lowenstein. 2005. "Adam Smith, Behavioral Economist," *Journal of Economic Perspectives*, 19(3): 131-145.

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

Casari, Marco, and Timothy N. Cason. 2009. "The Strategy Method Lowers Measured Trustworthy Behavior," *Economics Letters*, 103(3): 157-59.

Cherry, Todd L., and John List. 2000. "Learning to Accept in Ultimatum Games: Evidence from an Experimental Design that Generates Low Offers," *Experimental Economics*, 3(1): 11-29.

Cox, James C. and Cary A. Deck. 2005. "On the Nature of Reciprocal Motives," *Economic Inquiry*, 43, 623–635.

Elleserg, Daniel. 1956. "Theory of the Reluctant Duelist," *American Economic Review*, 46(5): 909-923.

Falk, Armin, Ernst Fehr and Urs Fischbacher. 2007. "On the Nature of Fair Behavior," *Economic Inquiry*, 41(1): 20-26.

Ferguson, Adam. 1767. *An Essay on the History of Civil Society*. Echo Library: Middlesex, UK. (2007).

Ferguson, Adam. 1792. 'Of Man's Progressive Nature'. In *Selections from the Scottish Philosophy of Common Sense*, George A. Johnston (Ed.). The Open Court Publishing Company: Chicago. (2005).

Forsythe, Robert, Joel L. Horowitz, N. E. Savin, and Martin Sefton. 1994. "Fairness in Simple Bargaining Experiments," *Games and Economic Behavior*, 6(3): 347-69.

Gillies, Anthony S. and Mary L. Rigdon. 2008. "Epistemic Conditions and Social Preferences in Trust Games," Working paper, University of Michigan.

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze. 1982. "An Experimental Analysis of Ultimatum Bargaining," *Journal of Economic Behavior and Organization*, 3(4): 367–388.

Hayek, Friedrich von. 1973. *Law, Legislation and Liberty, Volume 1: Rules and Order.* University of Chicago Press: Chicago, IL.

Hayek, Friedrich von. 1988. *The Fatal Conceit: The Errors of Socialism*. University of Chicago Press: Chicago, IL.

Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon L. Smith. 1994. "Preferences, Property Rights, and Anonymity in Bargaining Experiments," *Games and Economic Behavior*, 7(3): 346-80.

Hoffman, Elizabeth, Kevin McCabe and Vernon L. Smith. 1996. "On Expectations and the Monetary Stakes in Ultimatum Games," *International Journal of Game Theory*, 25(3): 289-301.

Hume, David. 1740. *A Treatise of Human Nature*. Oxford University Press: New York, NY. (2000).

Johnson, Samuel. 1755. *A Dictionary of the English Language*. CD-rom. Oakland, CA: Octavo. (2005).

McCabe, Kevin, Stephen Rassenti and Vernon L. Smith. 1996. "Game Theory and Reciprocity in Some Extensive Form Experimental Games," *Proceedings of the National Academy of Arts and Sciences*, 93: 13421–13428.

McCabe, Kevin, Vernon L. Smith, and Michael LePore. 2000. "Intentionality Detection and 'Mindreading': Why Does Game Form Matter?" *Proceedings of the National Academy of Sciences*, 97(8): 4404-9.

McCabe, Kevin, Mary L. Rigdon, and Vernon L. Smith. 2003. "Positive Reciprocity and Intentions in Trust Games," *Journal of Economic Behavior and Organization*, 52(2): 267-275.

North, Douglass C. 1990. *Institutions, Institutional Change, and Economic Performance*. Cambridge: Cambridge University Press.

North, Douglass C. 2005. *Understanding the Process of Economic Change*. Princeton, NJ: Princeton University Press.

Pecorino, Paul and Mark Van Boening. 2010. "Fairness in an Embedded Ultimatum Game," J*ournal of Law and Economics*, 53: 263-287.

Rigdon, Mary L., Kevin A. McCabe, and Vernon L. Smith. 2007. "Sustaining Cooperation in Trust Games." *Economic Journal*, 117(522): 991-1007.

Salmon, Timothy C. and Bart J. Wilson. 2008. "Second Chance Offers Versus Sequential Auctions: Theory and Behavior," *Economic Theory*, 34: 47-67.

Smith, Adam. 1759. *Theory of Moral Sentiments*. Liberty Fund: Indianapolis, IN. (1982).

Smith, Adam. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations*. Vol. I & II. Liberty Fund: Indianapolis, IN. (1981).

Smith, Adam. 1795. "The History of Astronomy," In *Essays on Philosophical Subjects*. Liberty Fund: Indianapolis, IN. (1982).

Smith, Vernon L. 2008. *Rationality in Economics: Constructivist and Ecological Forms*. New York, NY: Cambridge University Press.

Sobel, Joel. 2005. "Interdependent Preferences and Reciprocity," *Journal of Economic Literature,* 93: 392-436.

Wilson, Bart J. 2012. “Contra Private Fairness,” *American Journal of Economics and Sociology*, 71: forthcoming.

Wilson, Bart J. 2010. “Social Preferences aren’t Preferences,” *Journal of Economic Behavior and Organization*, 73: 77-82.

Wilson, Bart J. 2008. “Language Games of Reciprocity,” *Journal of Economic Behavior and Organization*, 68: 365-77.