

*Sentiments, Conduct, and Trust in the Laboratory**

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

22 JANUARY 2013

BEHAVIOUR.

1. Manner of behaving one's self, whether good or bad; manners.

...

5. Conduct; general practice; course of life.

CONDUCT.

5. Behaviour; regular life.

—Samuel Johnson, *Dictionary of the English Language*, 1755

Background and Motivation

Current interest in trust games by experimental economists originated in the 1990's (Berg et al., 1995) following upon earlier studies of simple two-person ultimatum and dictator games (Guth et al., 1982; Kahneman et al., 1986). The finding that decisions in these games collided with the predictions of game theory subsequently ignited a large literature on trust games.^{1,2} This literature has extensively replicated and explored the robustness of the original

* We thank Jeffrey Kirchner for his software programming *par excellence*, Jennifer Cunningham for diligently recruiting our subjects, Chapman University for financial support, Sean Crockett for constructive comments, and finally the many students with whom we have read and discussed the ideas in Adam Smith's two great books.

[†] Email: vsmith@chapman.edu

[‡] Email: bartwilson@gmail.com

¹ Yet equilibrium theory had performed well in a great variety of experimental markets in which subjects traded using the institutions (language) of message exchange and contracting that had been observed in business and financial practice such as bid-ask double auctions, posted pricing, and call markets (Smith, 1982). It is important to observe, however, that when trust games are conducted under the same conditions as the experimental markets—repeat play and private information on payoffs—convergence is to equilibrium predicted outcomes. (McCabe et al., 1998, pp. 16-19). Since private information forecloses any prospect of the players reading or signaling conduct in their actions, this condition removes all social content from trust games much as in our representations of the extended order of markets. But in market experiments that provide complete information, convergence to equilibrium is still observed, even if that convergence is less rapid than with private information (Smith, 1982).

² Experimentalists long have asked whether their replicable findings in student subject population are special to that group or can be extend to other populations. One answer is to go to the field for subjects; examples include the bilateral bargaining experiments reported in Siegel and Harnett (1964) comparing undergraduates with General Electric executives, and the asset trading experiments reported in Porter and Smith (1994) comparing undergraduates with corporation executives and over-the-counter traders; neither of these studies report any differences in the qualitative patterns of behavior, but results vary and investigations should proceed case by case. In the tradition of Siegel and Harnett (1964), forty years later, Fehr and List (2004) report a comparison of CEOs and university students, both in Costa Rica, using the Berg et al. (1995) trust game. In this particular comparison the "CEOs exhibit considerably more trustful and trustworthy behavior than students; as a consequence, CEOs reach substantially higher efficiency levels" (Fehr and List, 2004, p. 764).

findings and launched a search for explanations and models, and the testing of their predictive implications in an effort to better account for the original discordant empirical findings.

Much of the subsequent research was motivated by the original reciprocity or exchange interpretation of these results and the costly punishment and reward strategies that characterized subject behavior (Berg et al., 1995, pp. 138-9):

In conclusion, experiments on ultimatum game, repeated prisoners' dilemma games, and other extensive form games provide strong evidence that people do punish inappropriate behavior even though this is personally costly. Furthermore, subjects take this into account when they make their decisions. The investment game provides evidence that people are also willing to reward appropriate behavior and this too is taken into account. Taken together these results suggest that both positive and negative forms of reciprocity exist and must be taken into account in order to explain the development of institutional forms which reinforce the propensity to reciprocate.³

The reciprocity narrative as an explanation of trust/trustworthiness derived much of its weight from concepts in evolutionary theory and in particular the developing field of evolutionary psychology theory that involved social exchange algorithms for 'mind reading,' 'intentionality,' and 'cheater detection' (Hoffman et al., 1998).

The growing empirical evidence in support of the original findings led to a second more formal response in which the traditional game-theoretic assumption of strictly self-interested agents was replaced by a utility function defined over both 'own' and 'other' reward payoffs, while retaining all the other assumptions.⁴ Reciprocity was thereby interpreted as a form of revealed other-regarding behavior, and this could be rationalized within the game-theoretic framework by simply postulating that agents were driven by an 'other-regarding' utility criterion. We wish to emphasize, however, that *when a key prediction of a theory fails, all of its assumptions must be on the table for reconsideration, and the search for a resolution must not exclude consideration of entirely different ways of thinking, representing, and modeling the phenomena.*

Cox et al. (2007) supply a concise summary of models that enrich utility by the inclusion of 'other' rewards. Their model is particularly noteworthy in prefacing the experiments we

³ Before his death, John Dickhaut helped to instigate an extension of these original experiments to the study of three-person trust games in which person A could transfer money which was tripled, to person B, who could transfer money that was tripled again to person C. Person C could then return money to B, and B could return money to A. The original qualitative patterns of trust and trustworthiness continued to be represented in the three-person case (see Reitz et al., 2013).

⁴ These assumptions were: backward induction (players look ahead and apply reason to the analysis of other and own decisions); decisions are independent of the players' history or future (the game is played exactly once by anonymously paired players) and complete information on payoffs (fully displayed to both players). For further discussion, see Smith (2010, pp. 5-9).

report below because they parameterize utility to include postulated emotional responses—such as status, gratitude, and resentment—to the intentions conveyed by the first mover in two person games. They model only second-mover responses, but that is the obvious first step in a program to reform and redirect the theory exercise and in itself is not the source of the problem with this approach as we shall address it in this paper.

In his first book, *The Theory of Moral Sentiments*, Adam Smith (1759; hereafter *Sentiments*, or *TMS* for specific reference citation) articulated a theory of human sociality devoted to understanding moral human action; i.e., the “practice of the duties of life” (C.J. Smith, 1894, p. 574). He and his intellectual cohorts in the Scottish enlightenment were astute observers of their respective worlds of primary interest as they searched for the hidden rules that ordered the complex phenomena they studied.⁵ In this paper we (1) provide a brief account and interpretation of *Sentiments* showing that it (2) departs fundamentally from contemporary patterns of thought in economics⁶ that are believed to govern individual behavior in small groups, and (3) contains strong testable propositions governing the expression of that behavior; we then (4) apply the propositions to the prediction of actions in trust games, and (5) report experiments testing these predictions. In short, we argue that the system of sociability developed in *Sentiments* provides a coherent non-utilitarian model that is consistent with the pattern of results in trust games, and leads to testable new predictions, some of which we test in what follows.

‘Pleasure’ and the Mainsprings of Human Action in *Sentiments*

In contemporary representations by economists and cognitive psychologists ‘pleasure’ gives rise to ‘utility’ whose measure is related functionally to a desirable (or undesirable, where utility is negative) outcome resulting from the action. Given a choice among alternatives, an individual is postulated to choose the action that maximizes measured utility (hereafter, Max-U).⁷ Utility preference functions perform heavy duty work in modeling a vast range of human decisions: isolated individuals in psychophysical measurements, individuals choosing among uncertain probabilistic prospects, interactive agents in supply-and-demand auction and asset markets, and individuals interacting through choices in two-person (e.g., trust) games or in small groups (public good and common property games).⁸

⁵ For a broader perspective on Adam Smith and his cohorts in Scotland’s intellectual community, see Buchan (2003) and Phillipson (2010).

⁶ We recognize that *Sentiments* may have important connections to psychology, social psychology, philosophy, sociology, and anthropology, but any such discussion is well beyond the scope of what we attempt here.

⁷ So abbreviated and further discussed by McCloskey (2006).

⁸ Kahneman et al. (1997) provide a particularly clarifying and thoughtful distinction between the concept of Bentham’s (also Jevons’ and Edgeworth’s) intensity of experienced utility, and the writings of later and

These utilitarian interpretations constitute fundamental departures from the intellectual modeling framework of *Sentiments*, but in our view Adam Smith's systematic account of human action better illuminates the processes that govern action in small groupings of which trust games are an excellent example. In *Sentiments* the individual is painted as inseparably connected with overlapping social groupings based in family, extended family, friends and neighbors; these groupings in turn prepare and enable the individual to reach much beyond these narrow circles into daily life experiences. As Smith saw it, this is the world that first and originally defines the content and meaning of sociability, defines the individual within that social context, and out of which the civil order of society emerged based on property, defined as rights to undertake (or not) certain actions, conditional on circumstances.

The world of *Sentiments* envisions a pre-civil law community proving ground for fashioning the rules of social order in an environment disciplined by propriety, and bereft of any external enforcement of property. But of course it was a world in which individuals continued to engage and thrive long after the emergence of civil government, national economies, and the extended order of specialization and markets;⁹ the latter is the world Smith sought to understand in his better-known and phenomenally successful second book, *An Inquiry Into the Nature and Causes of the Wealth of Nations* (Smith, 1776; hereafter *Wealth*). But *Wealth* was a systematic treatment of economic development in civil society based on third party enforcement of property right rules—justice. The two worlds were distinct but complementary, and *Sentiments* articulated the critical preconditions for the emergence of justice and the enabling of civil society.

In *Sentiments* Adam Smith frequently makes reference to the 'pleasure' associated with an action chosen by an individual. What did 'pleasure' mean in *Sentiments*? The title of Part I, Section I, Chapter II, provides the key definition: "Of the Pleasure of mutual Sympathy" (*TMS*, p. 13). It refers to the fellow-feeling which Smith saw as the critical common feature of human sociability that governs individual conduct:

But whatever may be the cause of sympathy, or however it may be excited, nothing pleases us more than to observe in other men a fellow-feeling with all the emotions of our own breast; nor are we ever so much shocked as by the appearance of the contrary. Those who are fond of deducing all our sentiments from certain refinements of self-love, think themselves at no loss to account, according to their own principles, both for this pleasure and this pain. (*TMS*, p. 13)

contemporary economists based on decision utility. These utilitarian concepts, however, deflect modeling attention away from the foundation of social action in conduct as we find it developed in *Sentiments*.

⁹ Indeed, today the multibillion dollar demand for "sociability" has become global in Facebook and other social media enterprises.

As the person who is principally interested in any event is pleased with our sympathy, and hurt by the want of it, so we, too, seem to be pleased when we are able to sympathize with him, and to be hurt when we are unable. (*TMS*, p. 15)

Since the modern reader may think that this framework surely must only be about intimate friends, we hasten to add Smith's dictum that such sentiments, "...when expressed in the countenance or behavior, even towards those not peculiarly connected with ourselves, please the indifferent spectator upon almost every occasion" (*TMS*, p. 38). Provisionally, therefore, *Sentiments* should be viewed as a theory for all occasions.

Quite evidently, Smith was not a utilitarian.¹⁰ His system was not about outcomes, nor about equilibrium in outcomes, nor especially about "behavior" in its ordinary usage in the standard social science model. The first dictionary of the English language in 1755 did not even include the modern social scientific meaning of BEHAVIOR, which is definition 5 in the *Oxford English Dictionary (OED)*: "the manner in which a thing acts under specified conditions or circumstances, or in relation to other things".¹¹ For Smith, behavior is about rules all the way down (see Samuel Johnson's definitions at the head of the paper). A person interfacing with others either acts within the rules, for which he is deemed to be "good" or at least "not bad", or he acts outside the rules, for which he is deemed to be "bad". Thus, in the course of life a person as a general practice either conducts himself well in a morally upstanding way, or regularly or on occasion ill falls short.

In contrast, modern positivistic economic interpretations of observations in the laboratory abstain from attributing moral judgments as a mainspring to human action, treating people as things acting in relation to other things under specified conditions.¹² There is no good or bad manner of behaving one's self, only strategy s_i 's for $i = 1, \dots, n$ and observed frequencies thereof. The sole basis for predicting which s_i that a person would take is the highest possible utilitarian payoff. So in the 1980's and 1990's when human beings, as opposed to mere things, regularly chose s_i 's for actual payoffs that yielded lower payoffs than what were possible in laboratory experiments, economists rebalanced the discrepant utility function by padding run-of-the-mill self-loving preferences with so-called social preferences, and hence self-regarding behavior with other-regarding behavior (Wilson, 2010).¹³

¹⁰ And hence, *pace* Ashraf, Camerer, and Lowenstein (2005), he is not a "behavioral economist".

¹¹ The *OED* only traces this definition back to 1674, whereas definition 1, the manner of conducting oneself in the external relations of life, goes back to 1490. Examples for definition 2, which is word for word the same as Samuel Johnson's in 1755, are dated 1521 (perhaps) and 1535.

¹² See Kurzban (2001) for an evolutionary psychologist's critique of experimental economics as an essentially behaviorist enterprise.

¹³ Methodologically, both experimental and behavioral economists were continuing in the tradition of Bentham and Jevons.

For Smith, however, “self-regarding behavior” is an oxymoron and “other-regarding behavior” a pleonasm. When he says that “[m]en of virtue only can feel that entire confidence in the conduct and behaviour of one another, which can, at all times, assure them that they can never either offend or be offended by one another,” the “behaviour”, with no modifier, that he references already regards others because those regards are embedded in the rules that govern human intercourse (*TMS*, p. 225). As Charles John Smith in his *Synonyms Discriminated* (1894) explains, “BEHAVIOUR...refers to all those actions which are open to the observation of others as well as those which are specifically directed to others. As BEHAVIOUR refers more especially to actions, so Demeanour...refers more directly to manners; or in other words, Demeanour regards one’s self, BEHAVIOUR regards others” (p. 159).

By using both CONDUCT and BEHAVIOUR, the meticulous Adam Smith intends to place the confidence of men of virtue in two distinct concepts. Chiefly though, his project throughout *Sentiments* is about conduct.¹⁴ Charles John Smith discriminates the synonyms for us (p. 159):

As BEHAVIOUR belongs to the minor morals of society, so CONDUCT to the graver questions of personal life...We speak of a man’s behaviour in the social circle, of his conduct in his family, as a citizen, or in life. Good conduct is meritorious and virtuous. Good behaviour may be natural or artificial. The conduct has relation to the station of men’s lives, or the circumstances in which they are placed. Good conduct will include right behaviour as part of it, and a proper demeanour will flow necessarily out of it.

An isolated individual *j* abstracted from society is but a counterfactual thought experiment to impress upon the reader the central role of sociability in (moral) human action (*TMS*, p.110):

Were it possible that a human creature could grow up to manhood in some solitary place, without any communication with his own species, he could no more think of his own character, of the propriety or demerit of his own sentiments and conduct, of the beauty or deformity of his own mind, than of the beauty or deformity of his own face...Bring him into society, and he is immediately provided with the mirror he wanted before. It is placed in the countenance and behavior of those he lives with, which always mark when they enter into, and when they disapprove of his sentiments;¹⁵ and it is here that he first views the propriety and impropriety of his passions, the beauty and deformity of his own mind.

In ordinary human intercourse, we *feel* when we *experience* the mirror of life, the *sentiments* of which then lead us to *conduct* ourselves accordingly. We do not merely behave in the modern social scientific sense. Moreover, in the practice of virtues we direct our conduct in the

¹⁴ Adam Smith uses CONDUCT 309 times in the 338-page *Sentiments*, twice in a chapter title and once in the title of the very important Part III (Of the Foundation of our Judgments concerning our own Sentiments and Conduct, and of the Sense of Duty). BEHAVIOUR, on the other hand, never appears in a title and is used only 80 times. Moreover, *X* AND BEHAVIOUR is used 17 times where *X* is CONDUCT, CHARACTER, SENTIMENTS or COUNTENANCE; again, he often refers to *whole behavior*. The substance of Smith’s thought process—one to which we are not accustomed—is revealed in his careful diction.

¹⁵ Notice that the mirror of society is *in* the behavior of those with whom one lives, i.e., behavior regards others. Observe also the inference from Adam Smith’s thought experiment that the concept of the individual, of one’s own character, of self-knowledge, is ultimately derived from the idea of social mind or social psychology.

circumstances in which we find ourselves “by a certain idea of propriety, by a certain taste for a particular tenor of conduct, [rather] than by any regard to a precise maxim or rule” (*TMS*, p. 175).¹⁶

What then is the disconnect between a utilitarian model of behavior (in the modern social science sense) and human conduct? The deficiency of an egoist utilitarian approach, as Pettit (1995) explains, stems from a confusion of equating the outcome of acting with the motivation for acting in a social situation (pp. 311-12):

When I act on a desire to help an elderly person across the road, I act so as to satisfy that desire but I do not act for the sake of such satisfaction; I act for the sake of helping the elderly person. To think otherwise would be to confuse the sense in which I seek desire-satisfaction in an ordinary case like this and the sense in which I seek it when I relieve the longing for a cigarette by smoking or the yearning for a drink by going to the pub.

Rules of interpersonal conduct come into play when we see an elderly person attempting to cross a street and those supra- and subconscious rules generate the desire to help the elderly person, not some imaginary rush to the head for seeing the elderly person on the other side of the street that somehow more than offsets the buzzkill of subsequently being five minutes late to a meeting.

So unlike modern economics’ fixation with precise Max-U over outcomes, Adam Smith is concerned with understanding conduct, the fair-play rules governing that conduct, and the trial-and-error processes through which those rules might have emerged. Thus,

- On motivation:
Man has a “love of praise and of praise-worthiness” and a “dread of blame and blame-worthiness”, and “[t]he love of praise-worthiness is by no means derived altogether from the love of praise....though they resemble one another...[and]...are connected..., [they] are yet, in many respects, distinct and independent of one another” (*TMS*, pp. 113-114).
- On conduct and self-command (*TMS*, p. 83; italics added):
If he would act so as that the impartial spectator may enter into the principles of his conduct...he must...upon all...occasions, humble the arrogance of his self-love, and bring it down to something which other men can *go along with*...In the race for wealth, and honours, and preferments, he may run as hard as he can, and strain every nerve and every muscle, in order to

¹⁶ The exception is the practice of the virtue of justice. This nontrivial distinction between the rules of justice and the rules of all other virtues separates Adam Smith from Bicchieri (2006) who treats the rules of all virtues as rules of grammar. For Adam Smith, “the rules of justice may be compared to the rules for grammar; the rules of the other virtues [however], to the rules which critics lay down for the attainment of what is sublime and elegant in composition. The one, are precise, accurate, and indispensable. The other, are loose, vague and indeterminate, and present us rather with a general idea of the perfection we ought to aim at, than afford us any certain and infallible directions for acquiring it” (*TMS*, pp. 175-176).

outstrip all his competitors. But if he should jostle, or throw down any of them, the indulgence of the spectators is entirely at an end. It is a violation of fair play, which they cannot admit of. This man is to them, in every respect, as good as he: they do not enter into that self-love by which he prefers himself so much to this other, and cannot go along with the motive from which he hurt him.

- On process:

“...to attain this satisfaction (love and admiration), we must become the impartial spectators of our own character and conduct” (*TMS*, p. 114).

“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*, p. 110).

- On rules being derived from experience, not reason (*TMS*, p. 159):

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....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.¹⁷

Consequently, actions signal conduct or responses to the conduct inferred from the actions of others; and the general rules governing conduct become fixed through the discipline of their propriety; what others will “go along with” shapes the rule and determines its fitness. Only within the self-governing discipline of these general rules, is there scope for the individual to seek self-loving personal gain.

Propositions on Beneficence and Justice

Within this framework, Smith states three relational propositions on beneficence and justice. While the indispensable virtue of justice may be a familiar concept to the reader,¹⁸ BENEFICENCE has an archaic ring to it, sounding more like an 18th century word than a 21st century one (when did you last use the word in conversation?). Beneficence is, literally from Latin, *well doing*, and according to the *OED* only entered the lexicon in the 16th century. Its older Latin

¹⁷ Hence general rules are not a product of reason, or rational construction; they are formed ‘insensibly’ out of experience and if efficient are ecologically rational, as in Smith (2008).

¹⁸ “Society...cannot subsist among those who are at all times ready to hurt and injure one another. The moment that injury begins, the moment that mutual resentment and animosity take place, all the bands of it are broke asunder...Beneficence, therefore, is less essential to the existence of society than justice. Society may subsist, though not in the most comfortable state, without beneficence; but the prevalence of injustice must utterly destroy it” (*TMS*, p. 86).

relation, BENEVOLENCE, is used by Chaucer (that would be in 1384). Benevolence is, literally, *well willing*. Thus, BENEVOLENCE consists of the intention to do good for another, BENEFICENCE the action that does good for another. A niggardly, selfish, or mischievous man, as determined by the circumstances, cannot be beneficent even if what he does is good. Thus, beneficence always presupposes benevolence. Charles John Smith (1894) distinguishes BENEVOLENCE from BENIGNITY (another 18th century-sounding word), HUMANITY, and KINDNESS (pp. 165-166):

Benignity is, as it were, dormant, or passive benevolence. It is a matter more of temperament than will...As benevolence is inherent, so benignity may be shown on special occasions only...Humanity expresses an impulse rather than a quality...[and] is not so much a virtue when exhibited as something the absence of which is positively disgraceful and evil...Kindness is very like benevolence, but is rather a social than a moral virtue. It applies to minor acts of courtesy and good will, for which benevolence would be too serious a term.

The conceptual distinction between kind intentions as applicable to minor acts and benevolence as applicable to more serious, and hence beneficent, acts reinforces Adam Smith's claim that the general rules of conduct are "loose, vague and indeterminate, and present us rather with a general idea of the perfection we ought to aim at, than afford us any certain and infallible directions for acquiring it" (pp. 175-176).¹⁹ In other words, there is room for disagreement on interpreting an act in context as connoting minor or major intentions of doing good (Wilson, 2010 & 2012).²⁰

With this background, here then are three 'relational propositions' from *Sentiments* (Part II, Section ii, Chapter I):²¹

- *Beneficence Proposition 1: Properly motivated beneficent actions alone require reward.* Why? "...because such alone are the approved objects of gratitude, or excite the sympathetic gratitude of the spectator" (TMS, p. 78).
- *Beneficence Proposition 2: The want of beneficence cannot provoke resentment and punishment.*

Why not? "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" (TMS, p. 78).²²

¹⁹ Cf. Fn 16.

²⁰ Where in modern game theory is the assumption of agreement on the interpretation of the act? Hidden obscurely in the assumption that every individual j always chooses the largest possible pot of utilitarian pleasure $U_j(\cdot)$.

²¹ We name and number them as propositions; *Sentiments* does not.

²² In Smith and Wilson (2011), we use this proposition to interpret the standard ultimatum game context as projecting a form of involuntary extortion: the first player's choice is subject to veto by the second, the players' roles having been determined at random. Under this interpretation the proposition denies that ultimatum offers can be described as involving 'beneficence,' or that the responses involve 'gratitude' or 'punishment' independent

- *Injustice Proposition: Improperly motivated hurtful actions alone deserve punishment.* Why? “...because such alone are the approved objects of resentment, or excite the sympathetic resentment of the spectator” (TMS, p. 78).

These are strong falsifiable propositions applying to choice behavior [*sic*, conduct; hereafter we will use the language of *Sentiments*, not the language of contemporary economics] in a wide range of games in which actions are voluntary, extortion free, and convey intentions likely to be unambiguously interpreted as ‘beneficial’ or ‘hurtful’. Let’s see how well *Sentiments* predicts conduct over 250 years after it was first offered.

Trust Game: Designs for the Reward and Punishment of Actions

Our starting point, displayed in Figure 1, is a slightly modified two-person trust game of McCabe and Smith (2000), which has been replicated by Cox and Deck (2005) and Gillies and Rigdon (2008). The first modification is that the payoffs are increased by 20%. If, as the first-mover, **Person 1** ends the game by playing right, each person receives \$12 instead of the original \$10. Similarly, if **Person 1** plays down and **Person 2** plays right, **Person 1** receives \$18 and **Person 2** \$30 instead of the original \$15 and 25, respectively. The second modification is that if **Person 2** plays down, **Person 1** receives, instead of nothing, the modest non-zero amount of \$6, and **Person 2** receives \$42, as opposed to \$40 in the original game. The non-zero amount for **Person 1** is necessary to implement the test of two of our propositions in our two other initial treatments. Notice that this game, like that in McCabe and Smith (2000), seems quite hazardous for a **Person 1** who plays down.

Previous results combined from the three studies listed above indicate that 46 out of 98 first movers pass the play to the second mover, and consistent with *Beneficence Proposition 1*, 31 second movers reward the beneficent actions of the trusting first-mover by playing right. In what we will call the *No Punish Game* (hereafter *NP Game*) we will test whether our slight modification of **Person 2**’s choices affects the conduct of our **Person 2**’s. Notice that 15 or 33% of the second-movers do not agree with their corresponding 15 fallible first-movers that this situation calls for rewarding the first mover. There is something in the way this problem is represented to the minds of these second-movers that they do not see beneficence in the first-mover playing down.²³ Perhaps they perceive mere kindness or not even that. Only further experimental designs could flesh out the motivations of these second-movers. One take-away

of their perception as extortionist. It also suggests that the ultimatum game outcomes will be sensitive to changes in the context wherein procedures or narratives rationalize a process whereby subjects have reached the ultimatum stage game.

²³ See also Smith and Wilson (2011).

point is that before we extend this basic extensive form game,²⁴ this problem is not nearly as simple as traditional game theory presumes it to be for its human subjects. All they have to cleave to are their past experiences in life and the bare structure of an extensive form game tree.

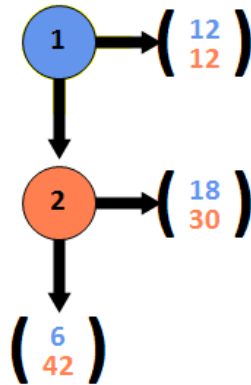


Figure 1. *No Punish (NP) Game*

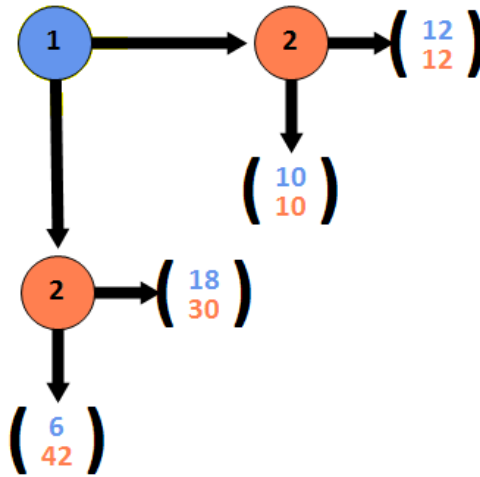


Figure 2. *Punish Want of Beneficence (PWB) Game*

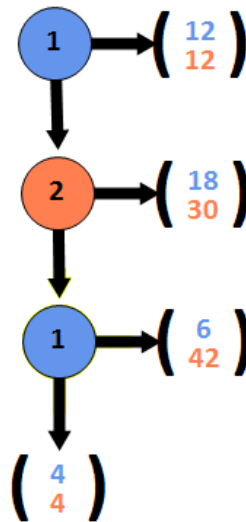


Figure 3. *Punish Hurt (PH) Game*

To test *Beneficence Proposition 2*, we construct the extensive form game in Figure 2, which we will call the *Punish Want of Beneficence Game* (hereafter, *PWB Game*). The

²⁴ Extensive form trust games yield quite different outcomes than their strategic or normal game-theoretic form representations; in particular people are less trusting and less trustworthy in the latter, and we study only the former (see, e.g., McCabe et al., 2000).

difference between this game and the *NP Game* is that if *Person 1* plays right, *Person 2* can either punish the want of beneficence of *Person 1* by playing down, yielding \$10 to each person, or *Person 2* can play right yielding \$12 to each person. Playing down by *Person 1* is unambiguously beneficent (at least we hypothesize so) towards *Person 2* as *Person 2*'s payoff increases by 250% or 350%.²⁵

The *Punish Hurt* (hereafter, *PH Game*) in Figure 3 is designed to test the *Injustice Proposition*. Compared to the *NP Game* in Figure 1, if *Person 2* plays down, *Person 1* can either accept the (\$6, \$42) outcome as in Figure 1, or choose to punish hurt and receive less than \$18. Punishing the hurt of *Person 2* by playing down comes at the cost of \$2 for *Person 1*, but it also reduces the payoff of *Person 2* from \$42 to \$4.

Notice that the personal cost of punishment by *Person 2* in the *PWB Game* and by *Person 1* in the *PH Game* is \$2 in both. *Sentiments* predicts that (a) *Person 2* will not punish want of beneficence by playing right if *Person 1* plays right in Figure 2, but that (b) *Person 1* will punish hurt by playing down a second time if *Person 2* plays down in Figure 3.

Procedures

We originally recruited 150 students with a variety of majors from the undergraduate population at a private university with approximately 5,000 undergraduates. Each of 5 sessions conducted over 3 weeks consisted of 30 students equally and randomly assigned to each of the three games described above. No student had any prior experience in an extensive or normal form game though many had experience in a prior experiment interacting with more than one other person. To explicate these results, we recruited another 146 students from the same population and assigned them (nearly) equally and randomly to each of three games, one of which was the *NP Game* in Figure 1. The two new additional games in the second series in this study will be presented and discussed below. But first we will report the results for 49 total pairs in the *NP Game*, 25 total pairs in the *PWB Game*, and 25 total pairs in the *PH Game*.

Each subject was paid \$7 for showing up on time and was seated in one of two computer laboratories. One laboratory contained 14 people seated in 14 carrels and the other 16 people were seated in the front portion of a 24-carrel room. The roles of *Person 1* and *Person 2* in all three games were distributed nearly proportionately between the two rooms. The subjects read at their own pace the interactive computerized instructions contained in the appendix. After playing one and only one of the extensive form games, the subjects were

²⁵ We note that because of these inequitable outcomes, models of inequity aversion (e.g., Fehr and Schmidt, 1999) also predict that *Person 1* would not play down.

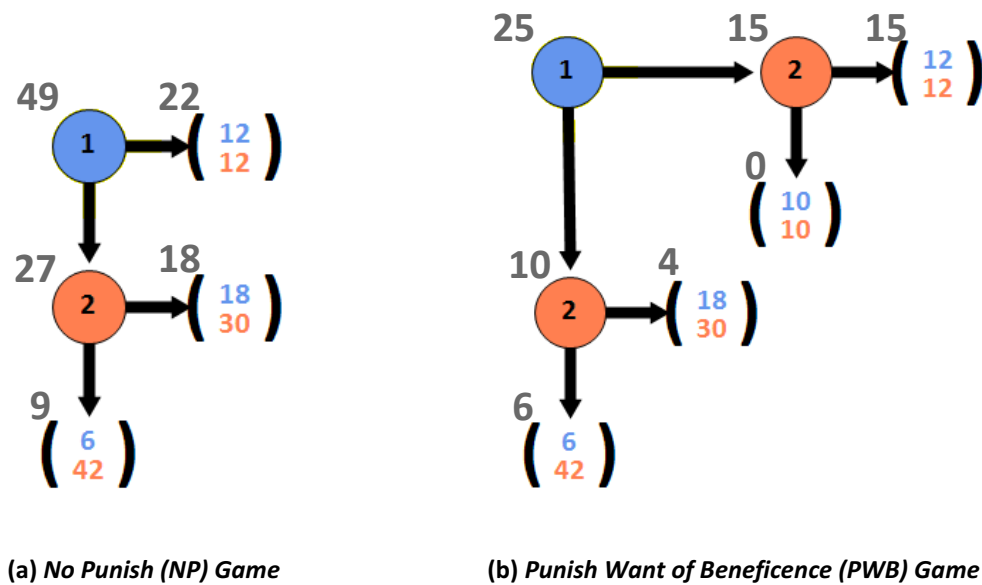
privately paid their earnings, which averaged \$18.11, excluding the show up payment.²⁶ The experiment lasted well under the 60 minutes for which they were recruited.

Results

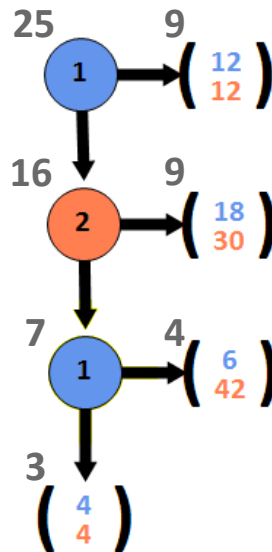
Figure 4 reports the number of decisions at each node and the number of outcomes reached in each of the three games. Our first finding establishes a baseline by comparing our *No Punish Game* to the previous trust game studied by McCabe and Smith (2000) and replicated by Cox and Deck (2005) and Gillies and Rigdon (2008).

Finding 1: *First movers in the NP Game beneficently play down, and second movers reward that action at frequencies consistent with those observed in the previous trust games.*

Of the 49 **Person 1**'s in our sample, 27 (55%) move down. Previous studies combined have found that 46 out of 98 first movers (47%) trust their anonymous second mover. Using a two-sided two-proportion z-test, we fail to reject the null hypothesis of equal proportions ($z = 0.93$, $p\text{-value} = 0.3507$). Of the 27 **Person 2**'s who have the opportunity to move, 18 (67%) play right. Previous studies have found that 31 out of 46 (67%) second movers honor the trust of the first mover. Again, we fail to reject the null hypothesis of equal proportions with a two-sided test ($z = 0.06$, $p\text{-value} = 0.9493$).



²⁶ Thus, the 296 subjects were paid a sum total of \$7,432 for participating in this experiment.



(c) Punish Hurt (PH) Game

Figure 4. Number of Decisions by Node and Outcome

Having replicated the rate at which **Person 2's** reward the beneficent actions of **Person 1**, our next finding reports whether Adam Smith's *Beneficence Proposition 2* holds in the *PWB Game*.

Finding 2: Fifteen out of 15 (100%) **Person 2's** do not punish **Person 1** for failing to beneficently play down in the *PWB Game*.

Empirically, however, the “simple” addition of a decision node for **Person 2** when **Person 1** plays right changes the frequency at which **Person 2's** play right to reward **Person 1** when the latter beneficently plays down. (Consistent with *Sentiments*, circumstances matter.) Instead of 33% of the **Person 2's** playing down after **Person 1** plays down, 6 out of 10 (60%) fail to reward the beneficent action of **Person 1**. Pooling the second mover decisions across the four studies with the basic trust game, this difference is marginally significant with a two-sided test ($z = 1.67$, $p\text{-value} = 0.0941$, $n_1 = 73$, $n_2 = 10$). Interestingly, **Person 1's** appear to anticipate that **Person 2's** may be untrustworthy by only playing down in 10 out of 25 pairs (40%) as opposed to the 73 out of 147 (50%) who beneficently play down in the simple two-node trust game. [This result is statistically insignificant ($z = 0.89$, $p\text{-value} = 0.3716$, $n_1 = 147$, $n_2 = 25$).] Even though Adam Smith successfully predicts that **Person 2** will not punish the want of beneficence 100% of the time when **Person 1** fails to act beneficently, it appears that giving **Person 2** the option to punish want of beneficence has the unintended consequence, in the counterfactual

treatment, of changing the response to **Person 1** beneficently playing down. We will flesh out this observation in the next section with a new game designed to explicate this result.

Our next finding reports a test of the *Injustice Proposition*:

Finding 3: Seven **Person 1**'s in the PH Game faced the decision of whether or not to punish the perfidy of **Person 2** and 3 (43%) do.

Comparing the first-move decisions of **Person 1**, in the *PWB Game* with those in the *PH Game* the option to punish conditional on the hurt of **Person 2** (Figure 4b and c) increases the frequency at which **Person 1** plays down (40% in the *PWB Game* versus 64% in the *PH Game*, $z = 1.70$, $p\text{-value} = 0.0894$).²⁷

Interestingly, the frequency at which **Person 2**'s follow *Beneficence Proposition 1* in the *PH Game* falls slightly relative to our baseline *NP Game* (9/16 vs. 18/27). Perhaps **Person 1**'s fail to anticipate that choosing "down" may be interpreted by many **Person 2**'s as a threat to "cooperate or else." And it backfires. Furthermore, for four of those **Person 1**'s on whom it backfired, they may then choose not to follow through on their dare by selecting (\$6, \$42) over (\$4, \$4). Of course, there are many other *ex post* interpretations (that, conditional on observing these specific results, are only possible *ex post*). For example, introducing upfront the possibility of punishment puts **Person 2** in a Max-U frame of mind, the blinders of which exclude from view any beneficence on the part of **Person 1**.²⁸ This is another reminder that the general rules of conduct are far from certain in a contextually sparse extensive form game, or, in other words, this is a hard problem for our dear subjects.

Two Additional Games

The results of the *PWB Game* indicate that the opportunity to punish want of beneficence simultaneously (a) reduces the frequency of **Person 1**'s beneficence to **Person 2** and (b) reduces the frequency at which **Person 2** rewards the properly motivated beneficent actions of **Person 1**. Is this because, vis-à-vis the *NP Game*, **Person 1** does not directly choose (\$12, \$12), or because **Person 2** has recourse to punishing the want of beneficence by choosing (\$10, \$10)? That is, the *PWB Game* introduces two alterations in the *NP Game*: (1) generically, regardless of the choice by **Person 1**, **Person 2** is the controlling player who determines the final outcome and (2) **Person 2** can now punish the want of beneficence by **Person 1**. To test the effects of the first change only, we conducted what we will call the *No Punish Pass Game* (hereafter the *NPP Game*) in Figure 5.

²⁷ Recall that the subjects are randomly assigned to one of the three treatments in each of 5 sessions.

²⁸ We thank Sean Crockett for point this out to us.

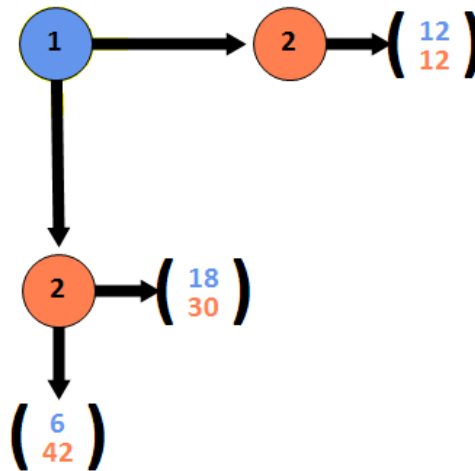


Figure 5. *No Punish Pass (NPP) Game*

The *NPP Game* like the *NP Game*, provides no opportunity for Person 2 to punish the want of beneficence by choosing (\$10, \$10). The difference is that, if Person 1 fails to act beneficently toward Person 2 by playing down, Person 2 has the involuntary option of ending the game with payoffs (\$12, \$12). Like the *PWB Game*, Person 2 must take an action, albeit one inconsequential to the payoffs, if Person 1 fails to act beneficently. The question is, is the frequency of Person 1's who act beneficently and the frequency of Person 2's who reward that beneficence closer to what we observe in the *NP Game* or in the *PWB Game*? Notice that there are four possible combinations of outcomes for the game in Figure 5:

- (i) Person 1's may act beneficently with the same frequency that they do in Figure 4(a) and Person 2's may reward that beneficence with the same frequency as they do in Figure 4(a);
- (ii) Person 1's may act (less) beneficently with the same frequency that they do in Figure 4(b) and Person 2's may (fail to) reward that beneficence with the same frequency as they do in Figure 4(b);
- (iii) Person 1's may act beneficently with the same frequency that they do in Figure 4(a) and Person 2's may (fail to) reward that beneficence with the same frequency as they do in Figure 4(b); and
- (iv) Person 1's may act (less) beneficently with the same frequency that they do in Figure 4(b) and Person 2's may reward that beneficence with the same frequency as they do in Figure 4(a).

Which set of results would you, the reader, predict? And how confident are you in that prediction? Having been surprised by the results in Figure 4(b), we had no clearly reliable insight, which is why we conducted the explicating treatment.

Having found in Figure 4 (b) and (c) that the sub-games available in the decision tree affect choice, although they do not change the analysis by self-loving players, we combine the opportunities to punish in each of these into a single game in Figure 6 in which either hurt or want of beneficence can be punished: the *Punish Either* (hereafter, *PE*) Game. In this form there is no imbalance in punishment opportunities as occurs in Figure 4 (b) and 4 (c). Again, we recruited subjects to be randomly assigned within each session to the games in Figure 1, Figure 5, and Figure 6 in proportions nearly equal.

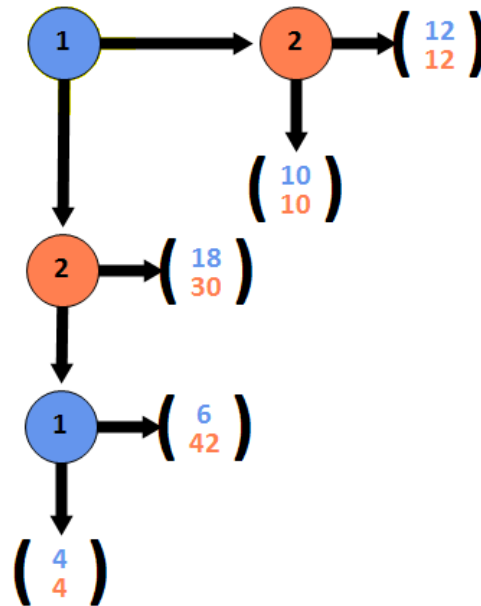


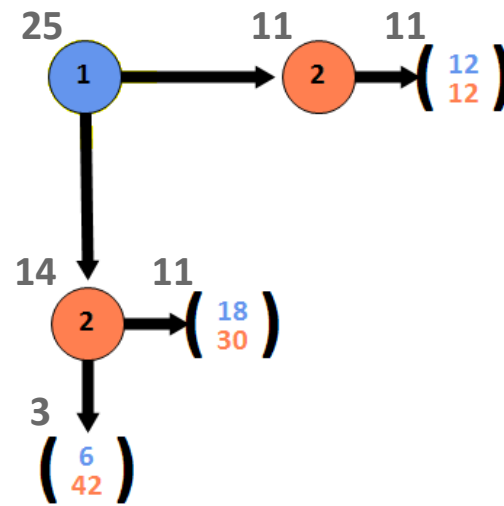
Figure 6. *Punish Either (PE) Game*

Figure 7 reports the number of decisions at each node and the number of outcomes reached in the *NPP Game* and the *PE Game*. We report first the results of the *NPP Game*.

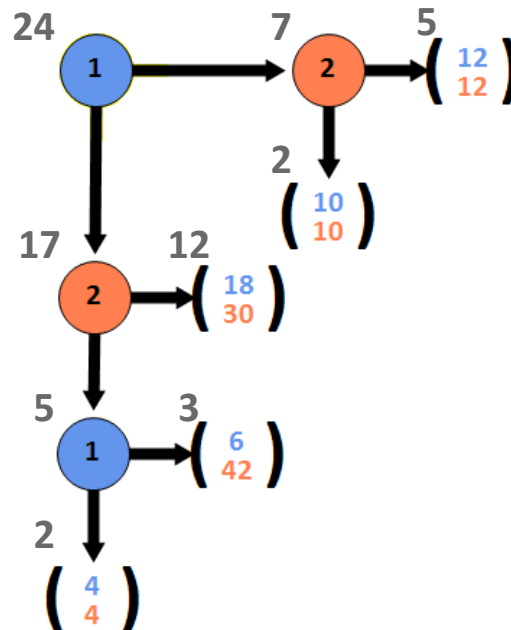
Finding 4: In the *NPP Game* *Person 2's* reward the beneficence of *Person 1* at (a) a higher frequency than in the *PWB Game* and (b) the same frequency as in the *NP Game*.

Fourteen out of 25 *Person 2's* in the *NPP Game* have the opportunity to reward the beneficence of *Person 1*, and 11 (79%) support Adam Smith's *Beneficence Proposition 1*. This observed frequency is nearly double that of what we observed in the *PWB Game*. Using a two-sided test, we reject the null hypothesis of equal proportions in the *NPP Game* and the *PWB Game* ($z = 1.92$, $p\text{-value} = 0.0272$). If anything, the involuntary option of *Person 2* when *Person 1* fails to act beneficently leads *Person 2's* to be more inclined to reward the beneficence of *Person 1*

(79% in the *NPP Game* versus 67% in the *NP Game*), though this difference is statistically insignificant ($z = 0.79$, $p\text{-value} = 0.2135$).



(a) *No Punish Pass (NPP) Game*



(b) *Punish Either (PE) Game*

Figure 7. Number of Decisions by Node and Outcome for the Two Additional Games

Thus, *Person 1*'s in the *NPP Game* appear to correctly anticipate that *Person 2*'s will properly reward their beneficence. Fourteen of 25 (56%) *Person 1*'s act beneficently toward *Person 2*, which is closer to the observed frequency in the *NP Game* (55%) than the observed frequency of the *PWB Game* (40%). Having observed the combination (i) above, the results

confirm that our *NPP Game* replicates the *NP Game*. Thus, we conclude that it is the opportunity to punish the want of beneficence by Person 2 that leads Person 2's to be less likely to reward the realized beneficence of Person 1.

As a counterfactual experimental treatment, the *PWB Game* allows us to measure the effect of a hypothetical rule on observed conduct. Punishing the want of beneficence is a rule of “what is *not*” (Hayek, 1973, p. 17). It does not emerge as a community convention because the want of beneficence does no real positive harm.²⁹ Thus, if *Beneficence Proposition 2* holds, introducing the opportunity to punish the want of beneficence interferes with reading the motives of one's counterpart, particularly in a situation stripped of the normal contextual cues that we rely upon to make such assessments. So perhaps we should not be surprised at the sensitivity of conduct by both Person 1's and Person 2's in the *PHW Game*. This observation reminds and humbles the hypothesizing social scientist that the human taste for a particular tenor of conduct is a rather sensitive and complicated palate.

Finally, the results of the *PE Game* in Figure 7(b) indicate that any differences induced by including (asymmetrically) the option to punish the want of beneficence are offset by also adding the opportunity to punish hurt. Seventeen of 24 (71%) Person 1's act beneficently, which is quite close to the 16 of 25 (64%) who do so in the *PH Game*. Twelve of 17 (71%) Person 2's reward the act of beneficence, which is line with the 67% who do so in the *NP Game*. For the 5 Person 1's who were hurt by Person 2, two of them (40%) punish that hurt, which is consistent with the 43% that do so in the *PH Game*. Lastly, we observe our first two (out of 7) Person 2's who punish a want of beneficence. As reported above in Finding 2, we previously found that zero out of 15 Person 2's were willing to punish the want of beneficence.

Conclusion

Modern (experimental) economics is endeavoring to understand moral human action in the laboratory and explain its associated sentiments by the calculated application of models of utility maximization, but to the extent that it has had any success the result no longer provides an understanding of moral human action. The error is compounded by the fact that because modern economics can never explain why this is so, it is wont to deny the fact. As Frank Knight recognized long ago, but whose methodological cautions have long since been forgotten, “the economist meets the problem of conduct and motive at every point and stage of his work” (1925, p. 374). If we want to understand how moral human beings act as they do in personal social situations we need to know more than what they have chosen amongst n possible

²⁹ We can imagine all sorts of beneficence in our favor and punishing the want of it invites resentment from those who might not agree that their beneficence to us is appropriate or even possible.

outcomes.³⁰ We also need to know the values and rules of everyday human intercourse that shaped and ordered those possible alternatives at that time and place, i.e., we need to know what they considered the appropriate conduct to be in the first place.³¹ *Sentiments* is about the ethical rules that constitute the character of an inherently sociable person who strives for a better life. This great book is the foundation for lost insights into a quintessentially humanistic science of economics. “Life is not fundamentally a striving for ends, for satisfactions, but rather for bases for further striving” (Knight, 1922, p. 459).

References

- Ashraf, Nava, Colin Camerer, and George Lowenstein. 2005. “Adam Smith, Behavioral Economist,” *Journal of Economic Perspectives*, 19(3): 131-145.
- Berg, Joyce, John Dickhaut, and Kevin McCabe. 1995. “Trust, Reciprocity, and Social History,” *Games and Economic Behavior*, 10: 122–142.
- Bicchieri, Cristina. 2006. *The Grammar of Society*. New York: Cambridge University Press.
- Buchan, James. 2003. *Crowded with Genius: The Scottish Enlightenment: Edinburgh’s Moment of the Mind*. New York: Harper.
- Cox, James C. and Cary A. Deck. 2005. “On the Nature of Reciprocal Motives,” *Economic Inquiry*, 43: 623–635.
- Cox, James C., Daniel Friedman, and Steven Gjerstad. 2007. “A Tractable Model of Reciprocity and Fairness,” *Games and Economic Behavior*, 59: 17–45.
- Fehr, Ernst and John List. 2004. “The Hidden Costs and Returns of Incentives—Trust and Trustworthiness among CEOs” *Journal of the European Economic Association*, 2: 743–771.
- Fehr, Ernst and Klaus M. Schmidt. 1999. “A Theory of Fairness, Competition, and Cooperation,” *Quarterly Journal of Economics*, 114(3): 817-868.
- Kahneman, Daniel, Peter Wakker, and Rakesh Sarin. 1997. “Back to Bentham? Explorations of Experienced Utility,” *Quarterly Journal of Economics*, 112: 375-405.
- Gillies, Anthony S. and Mary L. Rigdon. 2008. “Epistemic Conditions and Social Preferences in Trust Games,” Working paper, University of Michigan.

³⁰ That n is often simply 2 in the laboratory can deceive the observer into thinking that the conduct of a subject is easy to interpret. The sociality of mind and meaning is not that simple.

³¹ Furthermore, to suppose, that as an explanatory premise the appropriate conduct is part of a utility function that explains observed actions as following an hypothesized rule of conduct, is to reason in a circle (Wilson, 2008, pp. 372-373 spells out the circular argument for modeling reciprocal intentions in a utility function and likewise Wilson, 2012, pp. 409-410 for modeling utilitarian concerns of “fairness”).

- Hayek, F. A. 1973. *Law Legislation and Liberty, Vol. I Rules and Order*. Chicago: University of Chicago Press.
- Knight, Frank H. 1922. "Ethics and the Economic Interpretation," *Quarterly Journal of Economics*, 36(3): 454-481.
- Knight, Frank H. 1925. "Economic Psychology and the Value Problem," *Quarterly Journal of Economics*, 39(3): 372-409.
- Kurzban, Robert. 2001. "Are Experimental Economists Behaviorists and is Behaviorism for the Birds?" *Behavior and Brain Sciences*, 24(3): 420-421.
- McCabe, K., Stephen Rassenti, and Vernon L. Smith. 1996. Game Theory and Reciprocity in Some Extensive Form Experimental Games. *Proceedings of National Academy of Sciences*, 93: 13421-13428.
- McCabe, Kevin and Vernon L. Smith. 2000. "A Comparison of Naïve and Sophisticated Subject Behavior with Game Theoretic Predictions," *Proceedings of the National Academy of Arts and Sciences*, 97: 3777-81.
- 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): 4404-4409.
- McCloskey, Deirdre N. 2006. *The Bourgeois Virtues: Ethics for an Age of Commerce*. Chicago: University of Chicago Press.
- Pettit, Philip. 1995. "The Virtual Reality of 'Homo Economicus'," *The Monist*, 78(3): 308-329.
- Phillipson, Nicholas. 2010. *Adam Smith: An Enlightened Life*. New Haven: Yale University Press.
- Porter, David and Vernon L. Smith. 1994. "Stock Market Bubbles in the Laboratory," *Applied Mathematical Finance*, 1: 111-127.
- Rietz, Thomas A., Roman M. Sheremeta, Timothy W. Shields, and Vernon L. Smith. 2013. "Transparency, Efficiency and the Distribution of Economic Welfare in Pass-Through Investment Trust Games," *Journal of Economic Behavior and Organization*, forthcoming.
- Siegel, Sidney and Donald Harnett. 1964. "Bargaining Behavior: A Comparison between Mature Industrial Personnel and College Students," *Operations Research*, 12:300-304.
- Smith, Adam. 1759. *The Theory of Moral Sentiments*. Indianapolis: Liberty Fund (1982).
- Smith, Charles John. 1894. *Synonyms Discriminated: A Dictionary of Synonymous words in the English Language*. New York: Henry Holt and Company.
- Smith, Vernon L. 1982. "Microeconomic Systems as an Experimental Science," *American Economic Review*, 72(5): 923-955.

Smith, Vernon L. 2010. "Theory and Experiment: What are the Questions?" *Journal of Economic Behavior and Organization*, 73(1): 3-15.

Smith, Vernon L. and Bart J. Wilson. 2011. "Fair and Impartial Spectators in Experimental Economic Behavior." Economic Science Institute working paper, Chapman University, July. To appear in *Oxford Philosophical Concepts: Sympathy* (edited by Eric Schliesser).

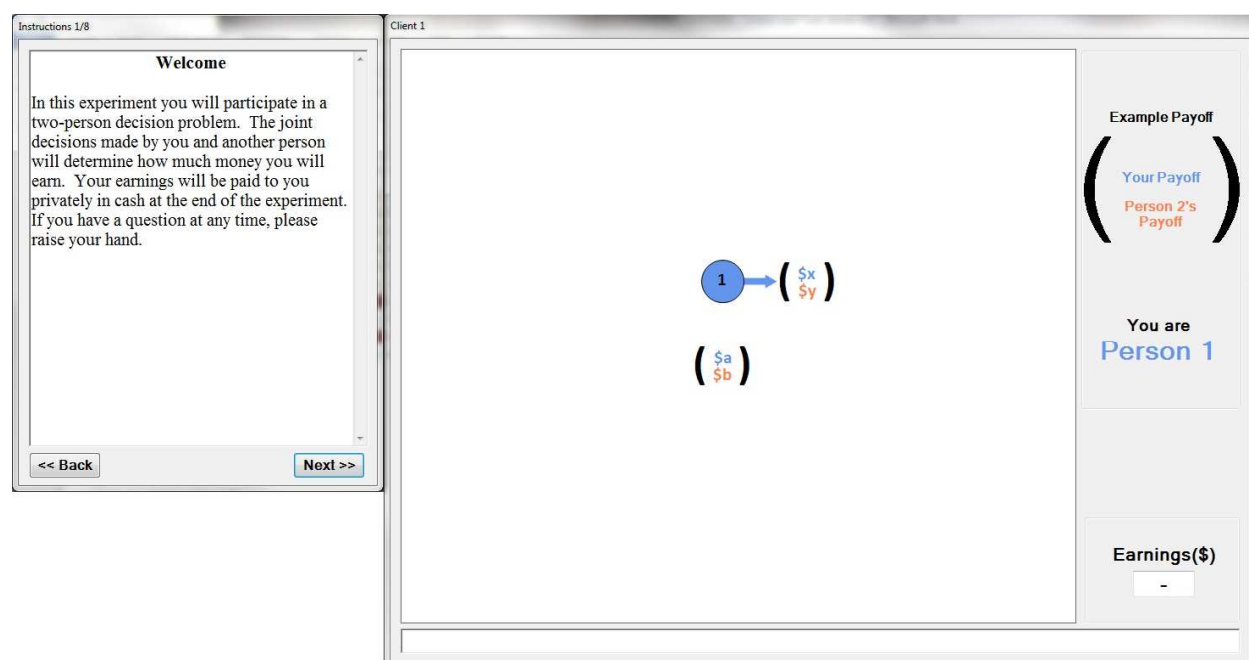
Wilson, Bart J. 2008. "Language Games of Reciprocity," *Journal of Economic Behavior and Organization*, 68(2): 365-377.

Wilson, Bart J. 2010. "Social Preferences aren't Preferences," *Journal of Economic Behavior and Organization*, 73(1): 77-82.

Wilson, Bart J. 2012. "Contra Private Fairness," *American Journal of Economics and Sociology*, 71(2): 407-435.

Appendix. Experiment Instructions

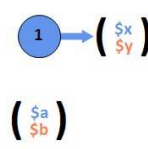
Please note that prior to making a decision, the **colored arrows** at the active decision node were flashing, alternating between the choices.



Instructions 2/8

The Diagram

The decision problem will be presented in a diagram. One of you will be **Person 1** (blue). The other person will be **Person 2** (orange). On the right, you will see whether you are **Person 1** or **Person 2**. You will be either a **Person 1** or a **Person 2** for the entire experiment.



Client 1

Example Payoff

(

Your Payoff

Person 2's Payoff

)

You are Person 1

Earnings(\$)

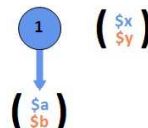
-

Instructions 3/8

The Diagram Continued

The payoffs to you and the other person will be displayed in parentheses in the diagram. There are two numbers: **Person 1** will earn what is in blue, and **Person 2** will earn what is in orange if that decision is made.

You and the other person will jointly determine a path through the diagram to a set of payoffs, which will be described next.



Client 1

Example Payoff

(

Your Payoff

Person 2's Payoff

)

You are Person 1

Earnings(\$)

-

Instructions 4/8

The Diagram Continued

A circle in the diagram is a point at which one person makes a decision. Each circle is color coded to indicate whether **Person 1** or **Person 2** will be making that decision. You will always have two options. Those options will include some combination of clicking on a set of payoffs and/or clicking on a circle.

If a person chooses a circle, a person will make the next decision at the next level in the diagram.

If a set of payoffs is chosen, the round ends with each of you receiving your respective earnings.

<< Back
Next >>

Client 1

Example Payoff

Your Payoff

Person 2's Payoff

You are
Person 1

Earnings(\$)
-

1

\$x

\$y

\$a

\$b

Please note that prior to making a decision, the **Submit** button flashed green until it was clicked.

Instructions 5/8

Example 1

Choose a payoff and click the **Submit** button.

<< Back
Retry
Next >>

Client 1

Example Payoff

Your Payoff

Person 2's Payoff

You are
Person 1

Submit

Earnings(\$)
-

1

\$x

\$y

\$a

\$b

Instructions 6/8

Example 2

Click on the **Orange** circle.

<< Back Retry Next >>

Client 1

```

graph TD
    1((1)) --> 2((2))
    2 --> P1["($a, $b)"]
    2 --> P2["($c, $d)"]
          
```

Example Payoff

Your Payoff
 Person 2's Payoff

You are Person 1

Submit

Earnings(\$)

-

Instructions 5/8

Example 1

Choose a payoff and click the **Submit** button. (done)

Person 1 earns \$a.
Person 2 earns \$b.

Continue to the next page of instructions.

<< Back Retry Next >>

Client 1

```

graph TD
    1((1)) --> P1["($x, $y)"]
    1 --> P2["($a, $b)"]
          
```

Example Payoff

Your Payoff
 Person 2's Payoff

You are Person 1

Earnings(\$)

-

Instructions 7/8

Example 3

Click on one of the circles.

<< Back Retry Next >>

Client 1

Example Payoff

(
Your Payoff
Person 2's Payoff
)

You are
Person 1

Submit

Earnings(\$)

-

Instructions 7/8

Example 3

Click on one of the circles. (done)

Person 2 now makes the next decision to choose the payoffs (\$a/\$b) or (\$c/\$d)

Continue to the next page of instructions.

<< Back Retry Next >>

Client 1

Example Payoff

(
Your Payoff
Person 2's Payoff
)

You are
Person 1

Earnings(\$)

-

Instructions 6/8

Example 2

Click on the **Orange** circle. (done)

Person 2 now makes the next decision. He or she will choose a set of payoffs.

Continue to the next page of instructions.

<< Back Retry Next >>

Client 1

Example Payoff

(Your Payoff
Person 2's Payoff)

You are
Person 1

Earnings(\$)

-

Instructions 8/8

Summary

This is the end of the instructions. The important points are:

- (1) You will be either **Person 1** or **Person 2** for the entire experiment.
- (2) If a person chooses a circle, a person will make a decision at the next level in the diagram.
- (3) If a person chooses a set of payoffs in parentheses, the round ends.
- (4) **Person 1's** payoff is displayed in blue, and **Person 2's** in orange.
- (5) You will participate in one and only one round of this experiment.

If you have any questions, please raise your hand and a monitor will come by to answer them. If you are finished with the instructions, please click the **Start** button. The instructions will remain on your screen until everyone has clicked the **Start** button.

We need everyone to click on the **Start** button before the experiment can begin.

<< Back Start Next >>

Client 1

Example Payoff

(Your Payoff
Person 2's Payoff)

You are
Person 1

Earnings(\$)

-