10-2018

Causal versus Consequential Motives in Mental Models of Agent Social and Economic Action: Experiments, and the Neoclassical Diversion in Economics

Vernon L. Smith
Chapman University, vsmith@chapman.edu

Causal versus Consequential Motives in Mental Models of Agent Social and Economic Action: Experiments, and the Neoclassical Diversion in Economics

Comments
Working Paper 18-11

Causal versus Consequential Motives in Mental Models of Agent Social and Economic Action: Experiments, and the Neoclassical Diversion in Economics

Vernon L. Smith

Chapman University

> [H]e who goes to the impulses from which action proceeds goes to the roots of action, and not merely to its issues.
>
> Samuel Alexander (1933, p 249)

MARKET EXCHANGE

Market Exchange for Non-Durable Goods and Services: Supply and Demand without Re-trade

In this paper I want to begin with the neoclassical supply and demand model of markets (SDM), whose static equilibrium consequences predicted outcomes far more accurately than were anticipated in laboratory experimental tests of the theory actuated by Jevons (1862, 1871; Smith, 1962). The observed predictive accuracy of SDM was not anticipated because complete information on supply and demand was widely believed, thought and taught to be a necessary condition for finding equilibrium.[1] Jevons' model required him to have complete information in any particular market, as he only articulated a model of market optimal outcomes, and no model of how individual actors might discover those outcomes. Only private information on their own unit values (costs) was provided to the subject-agents in the experiments; yet

[1] Walras (1969) had an exogenous mechanism for finding equilibrium, but it performed very poorly in experimental tests. Bronfman et al. (1996)

convergence was rapid. Market participants functioning under principles of motion, known to no one, were finding the equilibrium by means that were no part of the theory, nor any part of their own intentions and awareness.

This failure exposed flaws in the mental model we inherited from the neoclassical marginal tradition. That revolution appeared to introduce fundamental new insights linking price determination to individual utility—marginal value to buyers and marginal loss (cost) to sellers. For any price there was a corresponding maximum amount each self-interested, fully rational, buyer (seller) would be prepared to buy (sell); equilibrium in this conjunction was the "clearing" condition that was defined by the price that equalized the total amount taken with the total amounted offered by all individuals.

No part of Jevon's theory included a mental model of how the actor perceived the market, nor asked, based on some such perception, what information might be necessary, sufficient, or both necessary and sufficient, for agents to possess to achieve a competitive equilibrium. Prices determine optimal quantities bought or sold but how are prices determined? Consequently, the experiments exposed a flaw in our comprehension of markets that had ill prepared us for understanding the relationship between the theory and the choice observations by actors in markets. In new work by Inoua (2018), which I will discuss below, I hope to whet your appetite for delving further into its principle message: these errors and shortcomings were not shared by classical economics, which was more complete than appreciated.

The original market equilibrium framework of S&D models were conceptualized as a flow of a produced good or service from a source into the market, matched by a corresponding consumption-sink outflow.[2] By this reckoning, equilibrium occurs in a steady state flow. Hence, the good (or service) is perishable or, as in the macroeconomic accounts, is classified as non-durable, and is clearly distinguishable from a durable good. (Gjerstad and Smith, 2014, Chapter 2)

Exchange for Durable Assets: Effect of Re-Trade-Ability

Durable goods, or assets, constitute a store of value that is tapped by possessing and using it. Thus an automobile yields a future stream of transportation services if owned; the alternative is to rent (or lease), renewing each period. A house yields a stream of shelter services as long as it is owned; the alternative is to rent on current account. A bond (stock) yields a current interest (dividend) income payment flow. The SDM also readily applies to durables, if the end-of-market purpose is only to hold the asset for its service-yielding flow of consumption value. However, the important difference in price behavior between the two kinds of markets arises because non-durables are never re-traded—never bought for resale by end-use consumers. To use Adam Smith's (1776, p 30; hereafter WN) words, value in use and value in exchange are rigidly and inseparably linked. A durable good, however, can be re-traded and, accordingly, a discrepancy may emerge and persist between value in use and market value. Moreover, as we experienced in the long build-up and collapse of housing and mortgage

[2] Early experiments by Chamberlin (1948) did not consider flows over time, which constituted an important contribution of Marshall (1890).

markets in the Great Recession, that discrepancy may be the source of large-scale economic instability.

In sharp contrast with the first S&D experiments, the first experimental markets for an asset with fixed fundamental yield value did not immediately converge to fundamental value in an environment where fundamental value was common information. Unexpectedly, observed prices deviated "bubble-like" from fundamental holding value, though these markets converged if repeated a second and third time with the same participants. These highly replicable findings for durable asset trading were as puzzling as had been the rapid convergence behavior of the original markets for non-durables. (Smith, et al., 1988)

Finally, we summarize and review Adam Smith's (1759; hereafter TMS) non-utilitarian model of human sociality and apply it to simple two-person trust and ultimatum games. Smith's model was based on causal sources of individual action that had inferred consequences, and he was always quite clear that he was modelling the agent's experience of his/her relationship with other agents. Those relationships had important consequences for the welfare, stability and efficacy of society that he pursued, but this was not Smith's explanation of the cause or reason that the actions are taken.

The "mental models" perspective that I employ here descends from the contributions of Denzau and North (1994), Denzau and Roy (2005), Denzau et al. (2014), and motivates this

paper. Thus: “A fundamental theme of our paper was that one never sees things as they are, but rather only through the lens of the mental models in our heads.” Denzau et al. (2014, p. 5)[3]

This proposition applies to the actor who chooses in the context of particular social or market interactions, and to the observer/theorist who models the actors with the objective of explaining the roots of their action and/or the consequences of the action for society.

This perspective and interpretation in modeling agent actions and in modelling the consequences of those actions, is what distinguishes my remarks here from my discussions of the two kinds of markets and of two-person games in previous publications and lectures that I cite. In the quotation by Alexander (1933), the two modelling perspectives bifurcate on the impulse origins of action and that which issues from the action.[4] Both perspectives are important, but our scholarly emphasis in economics is, understandably, almost entirely on the consequential issues not the causal roots of action which probe into matters that overlap the disciplines of philosophy, psychology, social psychology and sociology. The upshot in experiments is that we either fail outright, or our success is tarnished by incompleteness arising from not understanding why we get the results we observe, both when they are confirming and when they are not confirming. Hence, our underestimate of the efficacy of agent action in private information markets for non-durables; our overestimate of the efficacy of rational

[3] Henri Bertoft expressed a challenging form of this insight as a proposition in the philosophy of science: “Science believes itself to be objective, but is in essence subjective because the witness is compelled to answer questions *which the scientist himself has formulated*. Scientists never notice the circularity in this because they hear the voice of “nature” speaking, not realizing it is the transposed echo of their own voice. Modern positivist science can only approach the whole as if it were a thing among things. Thus the scientist tries to grasp the whole as an object for interrogation. So it is that science today, by virtue of the method that is its hallmark is left with a fragmented world of things which it must then try to put together.” (Bortoft, 1996, p 17)

[4] See also Haig (2011) for a similar perspective by an evolutionary biologist.

action in complete information asset markets; and our complete failure to anticipate action in simple trust and ultimatum games by focusing on individual utility maximization rather than the rule-governed socializing and socialized individuals we study.

Neoclassical Marginal Analysis: Max-U, the Mental Model of Individual Economic Action, and Source of System Rationality.

The neo-classical economic model of all action by an individual is represented by a mapping from agent action into an outcome that yields personal subjective utility value to the agent, U[outcome (action)]. Given the set of alternative actions, A, an action, a_i (.)[5], in A by individual i is chosen so that U_i is maximized. Thus a_i^* = arg-Max U_i (a_i). Max-U is a generic mental model of best self-interested individual action in any system of interacting self-interested agents. Best action, a_i^*, may solve an unconstrained maximization problem; e.g., a monopoly seller of a commodity to buyers who are assumed to act non-strategically to reveal their demand at every price.[6] Alternatively, Max-U may solve the consumer choice problem in which a_i^* is an N-tuple of commodity quantities, purchased at fixed prices, subject to the consumer's income, or budget feasibility constraint. In a Nash equilibrium of a normal form game of strategy, an individual's utility outcome depends on the actions of all others. Subject to the condition that all not-I others choose an equilibrium action (a_i^*), then a_i^* represents the

[5] The notation a_i (.) refers to action as depending on the parameters or circumstances that characterize the individual. For each seller or buyer in an isolated market, (.) includes the subjective monetary value of units of the good; similarly, in an auction (.) is the value of the item to a bidder, with each bidder choosing his equilibrium bid (value) function; or each player in a two-person trust or ultimatum game chooses to maximize his or her own payoff conditional on the other player choosing to do likewise.

[6] For monopoly experiments see Smith (1981) where strategic under revelation of demand by buyers can limit monopolists from finding the monopoly price.

equilibrium best action for individual i to choose. Equilibrium via Nash is reduced to a Max-U "game against nature" of the same form as in the previous examples.

SUPPLY AND DEMAND (S&D)

Max-U Applied to the Supply and Demand for Non-Durable Goods: Complete information is not necessary.

Consider the market for a good or service that is consumed on demand. In the macro accounts these are non-durable goods and services. The 1870's intellectual break-through in neoclassical economics was to apply Max-U to characterize market price as an equilibrium between S&D.[7] (Jevons, 1862, 1871) One hundred years later the first experimental markets tested the ability of S&D equilibrium theory, Max-U, to predict the price and volume exchanged in such a market. (Smith, 1962; Davis and Holt, 1993) The market results provided unexpected support for the equilibrium version of Max-U--"unexpected" because the experiments involved strictly private decentralized information on individual utilitarian outcomes.[8] Thus, each

[7] The market outcome maximized Marshall's total buyer plus seller profit. However, Smith (WN) had achieved the same rational market results without relying or requiring individual rationality. (See Inoua, 2018)

[8] There is a large literature on "no trade" theorems that leave unexplained why people trade under conditions rationally inimical to it. For a laboratory investigation see Magnani, J. and Oprea, R. "Why Do People Violate No-Trade Theorems? A Diagnostic Experiment" at http://www.ryanoprea.com/. The authors formally model individuals that are overconfident as to their private information, of limited strategic sophistication, or noisy best responders under weak incentives, and precisely identify ways in which the rational modelling of outcomes fail to explain what people do. Consistent with the theme of this paper, I would propose that no individual could self-

individual had complete information on his or her own value(s) but had no information on the value(s) of any other person in the market. Economists at the time believed and taught that the competitive equilibrium was an abstract ideal, an unattainable state unless market agents each had complete information on the market S&D, and thereby knew the equilibrium-clearing price. Large numbers of buyers and sellers were believed also to be necessary. Finally, it was necessary that no participant could influence price with all participants acting as price-takers, with the very concept of S&D derived conditional on actors facing fixed alternative prices. None of these conditions were present in the first experiments.

An outpouring of experimental literature, after mid-20th century, falsified the belief that complete information is necessary for achieving a competitive equilibrium.[9] The early and many subsequent experiments used the bid/ask trading rules of the "double auction." Originally conducted as oral auctions they have been replaced for the most part by computerized versions. However, market performance was evaluated in studies that were extended to other price-making institutional procedures: Posted offer, posted bid and uniform price sealed bid-

identify with the representations offered here that are embedded in a framework defined by the theorist in advance. There is something in humans, chimps, even rats—call it curiosity—that leads them to try things, to explore and test the limits of their environment. When they get responses, enough of them deemed advantageous to make it a bettering strategy, they may repeat it. We exist because of these characteristics of adaptation to our environment.

[9] Svorenčík (2015) identifies this episode with "the experimental turn in economics."

offer auctions.[10] (Smith, 1982) Moreover, in very asymmetric S&D environments, replicable comparisons established that convergence was slower under complete information than incomplete. (Smith, 1965, 1980)

Jevons is usually credited with the "theorem" that had ruled predominantly in economics through the 1950s: "A market…is theoretically perfect only when all traders have perfect knowledge of the conditions of supply and demand, and the consequent ratio of exchange (price)…"[11] (W.S. Jevons, 1871, pp 86-7).

Mental Models of the Theorist or of Market Agents? Jevons vs Traders

Jevons the theorist needed complete information on the set of buyer values to specify market demand, and similarly, for supply. The conjunction of market S&D determined the competitive market clearing price, publically unknown in the absence of this distributed information. He had no explicit mental model of the traders, nor a model of traders' mental model of each other, operating in a market where, knowing only their own values, they determined a bid to buy or ask to sell for a unit.[12] Hence, attaining an equilibrium is surely beyond the reach of any market in the absence of complete information on S&D. Traders, however, in order to function effectively must formulate bid-ask strategies that—in effect—mentally model other traders, although not explicit or formal. Subjects in experiments,

[10] A "designer market"—a real time uniform price double auction—has also been used to evaluate market performance. (McCabe et al., 1993)

[11] He believed, however, that brokers on commodity exchanges somehow were able to infer that information.

[12] Gjerstad (2013) provides the first dynamic model of equilibrium, and tests of its convergence properties.

however, spontaneously interact, each responding to the bids, asks and contract acceptances of others. They "know" what to do, and do it, but have difficulty articulating a descriptive account of what they know and do.

Do Experiments Rehabilitate the Classical Model of Supply and Demand?

Inoua (2018) makes a compelling argument that the market experiments support classical economic theory, properly interpreted, not neoclassical theory, as we who did the research tended to believe.[13] The classical school "is mostly caricatured today: it is often said to have ignored the demand side…. But from Smith to Dupuit, demand is consistently represented by the concept of willingness to pay…classical theory is in fact rigorous and nearly complete *even in its original form*.[14] (Inoua, 2108, p 1; italics added)

Classical economists, beginning with WN and Adam Smith's followers through much of the 19th century, conceptualized demand as "value in use", which in modern language is maximum willingness-to-pay, measured by a schedule of market reservation prices. Exchange value is the market price. Demand is smooth for aggregates of individuals whose allocations are random, discrete, and not predictable as with markets as a whole. This is long reflected in

[13] This section is a late-draft addition. In earlier drafts, I had argued that the neoclassical diversion to Max-U equilibrium economics had lost the dynamic price-specialization-discovery process prominent in WN and classical economics, then rediscovered in experiments. Never-the-less—I thought—Max-U survived nicely in specifying static equilibria in market supply and demand, which in turn justified the induced utility value methodology for implementing Max-U neoclassical and game-theoretic models in the laboratory. Where it failed—so I thought—was in trust and other small group games. Inoua (2018) corrects my limited thought perspective in a way that includes neoclassical Max-U as a special, and empirically falsified, special case. The classical model simply begins with the observational distinction between "use" value and market price value, where use value corresponds precisely to modern notions of willingness-to-pay, or reservation prices, but with no required commitment to the individual rationality of these self-imposed limits by imperfectly informed, error-prone, individuals, a prominent theme in TMS and WN. I had been answering (in the negative) the question: "Is utility theory a theory of everything?" In substance, Inoua rightfully asks: "Is utility theory a theory of anything?"

[14] Inoua (2018, Section 3) offers an elegant "large market" completion in terms of V(p), a Liapunov function.

experiment results that approximate market equilibrium predictions taken as a whole far better than the individual outcomes that have their own error structure within the whole represented by aggregate S&D.

ASSET MARKETS

Mental Models of Asset Market Bubbles where Traders Are Given Complete Information on Fundamental Value

The euphoria that ensued, upon finding that S&D theory predicted price-quantity outcomes under far weaker information conditions than originally expected, was dampened by the discovery that markets for durable, and hence re-trade-able, goods under complete common information converged only very slowly across three within-subject replications. The delay took the form of price "bubbles"—systematic mispricing—that had direct application to housing and securities markets in the economy.[15]

I want to examine three different responses in the literature, each with distinct perspectives that illustrate how different mental models of subject trader behavior underlie different understanding of the patterns observed in these asset trading experiments.

In one prominent response, the price deviations from publically advertised fundamental values are irrational, a judgement based on mental models of theorists and experimentalists that focus on the outcome of the actions and their welfare implications (Lei, et al., 2001). Market rationality originates and depends on individual rationality. Thus, anyone acting in

[15] Smth et al., 1988; Gjerstad and Smith, 2014, chapter 2.

his/her own interest should easily see and understand that no one should be willing to buy a unit at a price above fundamental value. Likewise, no one should be willing to sell a unit below fundamental value. In either case to do so is to lose money. Anyone who avoids such actions can expect to collect a larger sum of money at the end of the experiment. Thus, irrational actors explain the experimental results. "Irrational", however, means that the subjects have a false mental model only in the sense that they do not think about their task in the way the experimenter/economist thinks about it. They do not use the information provided to perform reasoned calculations for informing the actions taken.

Caginalp and Balenovich (1999) propose an entirely different microeconomic mental model of agent behavior in asset trading. (Also see Cagilalp, et al., 2001) Their model of bubbles is based on the hypothesis that the market is composed of two kinds of investor-traders, each of whom are not irrational, but act in their own best interest according to how they perceive the market; i.e., each pursues his or her own interest in their own way. *Fundamental investors* buy (sell) shares in proportion to the discount (premium) between fundamental value, FV, and observed price in the market. Hence, given FV, if P is price, and $FV - P > 0$ they are active buyers; if $FV - P < 0$ they are active sellers; in each case activity is in proportion to the difference, and is rational in the same manner that economists analyze the situation. *Momentum investors* buy in proportion to the percentage rate of change in the current price regardless of FV. Hence, they are active buyers if $dP/Pdt > 0$, and active sellers if $dP/Pdt < 0$. Depending on their relative weights in the population of traders, the interaction between these two investment sentiments yields a rich variety of different bubble price patterns measured relative to FV. As an example, consider the housing market. The fundamentalist buys (sells) by

comparing housing prices with housing rents, where the latter is an indicator of FV. The momentum-ist buys (sells) entirely on the bases of current price changes. In this construction, bitcoin is an example of a market exclusively driven by momentum, as the items have no identifiable intrinsic worth.[16] (Caginalp and Caginalp, 2018)

A third perspective on the bubble literature is that of Sunder (1995, p 474) who observes that these asset market experiments eventually converge if the same subject group returns for a second and third session. Since equilibrium convergence in experimental markets commonly requires learning, asset market "bubble" phenomena can be interpreted as a pattern regularity in that learning process. This observation combined with the insights in Inoua (2018) suggest that asset bubbles are a variation on the classical economics that distinguishes use value from exchange value, which for durables incudes resale value. The latter reflects expectations of future trading value. Inexperienced subjects form diverse individual expectations of their WTP for future asset value. Some subjects sell for too little—prices below FV—others pay too much (buy above FV). Across sessions, they learn to correct these errors, and prices converge to the rational expectations FV of the asst.

MODELLING MORAL SENTIMENTS: PRINCIPLES BY WHICH WE EACH LEARN TO JUDGE THE CONDUCT AND CHARACTER OF OUR NEIGHBORS AND THEN OF OURSELVES

[16] However, I believe this construction is incomplete as it omits the intrinsic worth of bitcoin for its services in transactions only. A buyer of an item can convert from dollars into bitcoin, and in time interval c purchase the item; similarly, a seller of the item can convert to dollars. If c is small relative to the volatility of bitcoin, and the conversion transactions cost is low, then bitcoin will yield intrinsic fundamental value from its transactions services though it be a poor store of value.

The most serious experimental blow to the generality of Max-U as a predictive theory of all action occurred in the 1980s and 1990s. Over these two decades, experimentalists examined the choices of anonymously matched players in two-person ultimatum and trust games. The replicable and remarkably robust findings of this work documented a massive failure of the traditional Max-U model to predict subject choices. In this section, I want to apply the TMS model and three of its propositions to these games to show the model's relevance for comprehending the power of a non-utilitarian focus on modelling human social relationships. I will return to neoclassical market theory and experiments at the close, with some comments on these and the market experiments from the perspective of WN.

David Hume's Mental Model of Human Sociability Was Utilitarian; Adam Smith's Was Not

David Hume, in synchrony with the subsequent neoclassical Max-U mental model of all action, reduces all virtue and vice to their consequent outcome utilities. Adam Smith quotes Hume, then comments himself: "No qualities of the mind, he (Hume) observes, are approved of as virtuous, but such as are useful or agreeable either to the person himself or to others; and no qualities are disapproved of as vicious but such as have a contrary tendency." (TMS, p 188; Hume, 1740, p 591) Although Smith in no way denies the association of virtue with pleasure and of vice with pain, he resolutely states "…still I affirm that it is not the view of this utility or hurtfulness which is either the first or principal source of our approbation and disapprobation. These sentiments are no doubt enhanced and enlivened by the perception of the beauty or deformity which results from this utility or hurtfulness. But still, I say, they are originally and essentially different from this perception." (TMS, p 188)

For Hume (and the modern utilitarian), "'tis evident" that there is a one-to-one correspondence between virtue (vice) and utility (disutility). But what is the source of this association? Why is it not alone a utility matter? What in Adam Smith's mental model of action causes him to break with his friend Hume in interpreting the very foundations of the human capacity for mutual sympathetic fellow feeling?

From the Mirror of Society to Propriety

Smith's mental model of the sources of human action begins with a mental experiment. He asks us to imagine a member of our species growing up in complete isolation from any other member. That person cannot know any more about what it might mean to have a deformed mind than to have a deformed face for he has no natural means for looking at these things, no "mirror" that enables him to see these demarcating features of himself. "Bring him into society" and you give him the mirror he must have to become social and acquire knowledge of all these things. (TMS, p 110) Others always flag our actions with their stamp of approval or disapproval ("approbation or dis-approbation" in Smith's more precise articulation). Because we are part of society we learn that others react to our expressions of joy or sorrow, and this gives rise to new experiences of joy and sorrow in us. Thus do our feelings, and our capacity for sympathy, gradually become the basis for mutual fellow feeling.

The Great School of Self Command

Smith notes that young children are without self-command. Parents, quite properly indulgent of the child's social ignorance, intervene only to ensure minimum safety. Upon encountering playmates or entering school, the child finds that others have no such "indulgent

partiality." Wishing for favor and to avoid contempt, as children we initiate the process of moderating our passions that others will be pleased with us.[17] Thus, the child "enters into the great school of self-command." (TMS, p 145)

In that great school, a person does not lose his or her self-interested nature. On the contrary, common knowledge that all individuals are self-interested is how we know that an action hurts person A by taking something good from him, and benefits person B by giving him more of a good thing. "Though it may be true, therefore, that every individual, in his own breast, naturally prefers himself to all mankind, yet he dares not look mankind in the face, and avow that he acts according to this principle." When he acts, "he must…humble the arrogance of his self-love, and bring it down to something which other men can go along with." (TMS, p 83)

All humans are strictly self-interested, but go through a maturation process in which they learn to conduct themselves according to general rules. The process is evolutionary, and

[17] Perner et al. (1989) found that children before about age 4 ½ lack awareness of mental phenomena in others, and have as yet no "mindreading" ability; i.e., a natural ability to infer mental states or representations in others from their actions and words observed. An experiment that supports this proposition is the following. Smarties is a UK candy brand. In the experiment a child is asked, "What is in this box?" The reply, "Candy or Smarties." The box is opened and shown to the child who sees that the candy has been removed, and the box contains only pencils. The child is then asked what will be the response of the next child who come into the room when asked what is in the box. Below about age 4 ½ the answer will be "pencils." But older children are able to comprehend the concept of a "false belief" and say that the next child will respond with "candy." Simon Baron-Cohen (1995, Chapters 4 and 5) report other experiments that illustrate the development of mindreading and its connections with autism.

does not stem originally from reason. Actions are not approved or condemned because they have been examined and found to conform or not with certain general rules. “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.” The “general rules of morality…are ultimately founded upon experience…” (TMS, p 158-9)

The General Rules We Follow: Propositions on Beneficence and Justice

The rules germane to this paper fall into only two broad categories, beneficence and justice.

> Beneficence Proposition 1 concerns actions that do intentional good for other(s); Smith assures us that “these alone seem” to require a reward response based on the feeling of gratitude invoked by the action. Moreover the greater the benefit done the higher will tend to be the reward. (TMS, p 78, 81)

Contemporary behavioral-experimental-ists may object that the observations are due, simply, to “reciprocity.” However, Smith’s mental model is not circular. We observe reciprocal sequential actions, but how do we explain the actions? The observations, declared to be due to reciprocity, cannot also serve as an explanation. Beneficence and the calculus of gratitude-reward is the underlying explanation of positive reciprocity; only much later does Smith announce his reciprocity theorem: “Kindness is the parent of kindness,” (TMS, p 225)

The flip side of beneficence leads to Justice Proposition 1 in which intentional actions of a hurtful tendency provoke a punishment response because of the feelings of resentment they

cause in the injured party.[18] Moreover, "[A]s the greater and more irreparable the evil that is done, the resentment of the sufferer runs naturally the higher…" and hence to greater punishment. (TMS, p 78, 83-4) This proposition is the foundation of Smith's theory of property in jurisprudence in which rules of *propriety* morph into rules of *property*.

> "The most sacred laws of justice, therefore, those whose violation seems to call loudest for vengeance and punishment, are the laws which guard the life and person of our neighbour; the next are those which guard his property and possessions; and last of all come those which guard what are called his personal rights, or what is due to him from the promises of others." (TMS, p 84)

Punishment is dis-utilitarian, and tends to deter hurtful actions, but again Smith, the astute observer, is careful to make clear that this is *not the original reason that society punishes.* Note also that he uses the term "vengeance" in the above quotation. Smith's perspective is particularly well stated in his *Lectures on Jurisprudence*:

> "It is to be observed that our first approbation of punishment is not founded upon the regard to public utility which is commonly taken to be the foundation of it. It is our sympathy with the resentment of the sufferer which is the real principle." (Smith, 1978, p 475)

[18] In previous incarnations of our papers on Adam Smith, we called this an Injustice Proposition, because it is about punishing injustice. However, as I show shortly, Smith's theory of jurisprudence established justice by limiting and discouraging injustice. In that sense, "Injustice Proposition" is the appropriate direct-cause label, but the "Justice Proposition" is the label appropriate for capturing Smith's consequentialist analysis. Either label is correct. See Smith (2012); Smith and Wilson (2017).

Though it is widely believed that society has evolved punishment as a means of deterring criminal acts, this is not the original principle from which it arose. Contrastingly, we are informed by Smith, that early state jurisprudence took the form of allowing the family and friends of the victim of a crime like murder to choose the avenging, not deterring, penalty—it's too late to deter, but not too late to avenge the feelings of the victim.[19] The state intervenes to keep the peace by defusing an outburst of violence by the victim's family and friends; otherwise there are perpetual problems like the "Hatfield's and the McCoy's." Smith explains that this is why, when the British made the export of wool a capital crime, it was impossible to put together a local jury willing to level any such penalty; there was no "victim" to avenge in the eyes of the citizens!

In general, we cannot sympathize with a murderer or a robber, who is no part of our experience, but we can easily imagine the victim's distress due to the loss of a loved one, or of her property, caused by the perpetrator of the crime.

Returning to the quotation above on the "sacred laws of justice," Smith notes that theft and robbery carry larger penalties than violation of promises (contracts). This difference, he explains, is due to the asymmetry between gains and losses: "To be deprived of that which we are possessed of, is a greater evil than to be disappointed of what we have only the expectation. Breach of property, therefore, theft and robbery, which take from us what we are

[19] Note carefully that Smith sees rules as backward looking, arising out of common experience and thereby accepted by common-knowledge consent; not forward looking, subject to greater uncertainty of affect and the wind shear of unintended consequences.

possessed of, are greater crimes than breach of contract, which only disappoints us of what we expected." (TMS, p 84) [20]

In summary, crime leads to hurt, leads to resentment, leads to punishment, which deters, but that is not the origin of laws against murder, theft and robbery, and contract violation. That origin is in community sympathy for the victim, and the emotional need to avenge that suffering. Thus does punishment enter the rule of law, later become a crime against "the public," and only then put to work as a deterrent. Smith does not waver in the clarity with which he distinguishes between the roots of action and their issues, and this, in turn, carries over into distinctions between the origins of norms (rules of law) and the work they do for society.

Buttressing these propositions are two propositions that clarify the conditions under which they apply. Thus, under extortion, or threat of it, the calculus of gratitude-reward does not apply, for Beneficence Proposition 2 states 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. It may disappoint of the good which might reasonably have been expected, and upon that account it may justly excite dislike and

[20] Smith derived the asymmetry between gains and losses from the more fundamental psychological principle of asymmetry between our joy and our sorrow. Hence, Smith's theorem predicted the behavioral regularities found by Kahneman and Tversky (1979) which they rationalized in the light of fact in the form of "prospect theory." To recognize Smith by saying he "anticipated" the modern findings is a bit too generous toward modern discoveries. Smith's fully developed mental model of agent action predicted modern findings, and we were too ill informed to notice. Ashraf, et al., (2005) call attention to several such "anticipations" by Smith, but fall short of appreciating his superiority in comprehensively modelling agent action from the perspective of the agent, and only then examining its external and societal consequences.

disapprobation: it cannot, however, provoke any resentment which mankind will go along with." (TMS, p 78)

Symmetrically, Justice Proposition 2 states that: "Though the breach of justice…exposes to punishment, the observance of the rules of that virtue seems scarce to deserve any reward." (TMS, p 81-2) Society does not reward people for not disturbing their neighbor or for obeying the traffic laws. In the classical liberal tradition good things result from discouraging and limiting bad things; the achievement of societal good is through the large residue of freedom left over after applying rules for punishing and limiting the bad things that, based on our common experience, we agree deserve to be avenged.

We do not punish actions that fail to show beneficence nor do we reward actions that fail to violate law. Liberty implies individual responsibility that is decentralized in the voluntary (self-command) choice to do, or not, good things for others; for to require it would destroy the spontaneous expression of gratitude, and undermine human sociability. Similarly, there are no rewards for law-abiding conduct, which is your duty as a responsible citizen and facilitates solidarity.

Connecting with *The Wealth of Nations*

Property rights, derived from Justice Proposition 1, are a necessary condition for wealth creation through specialization, which in turn in "limited by the extent of the market." (WN, p 31) In TMS we learn that the rules of propriety in local communities ancient and modern morph into rules of property in the civil order and thus enable wealth to be created across the range of unknown and unknowable others in global communities. A sufficient condition for wealth

creation is Smith's fundamental axiom of discovery, "the propensity to truck, barter and exchange." (WN, p 25) This axiom in WN is directly derive-able from Beneficence Proposition 1 as an application in the mutual simultaneous exchange of the preferment in every trade wherein each provides gratitude to other, and each at once rewards other, with third party enforcement of property reducing dependence on mutual trust.

Extensive-Form Trust Games and Mental Models of the Results

I want to summarize various old and new trust games results, published in Smith and Wilson (2017; 2018), but examine them here in the context of Adam Smith's mental models of action, with commentary on the contrasting mental models of neoclassical game theory and experimental-behavioral economics.

Baseline Trust Game

In the baseline game of "pure trust" shown in Figure 1, Player 1s can either opt out by moving right, yielding payoffs (P1, P2) = ($12, $12), or move down, passing to their paired Player 2 counterparts; Player 2 can then "cooperate" (move right) yielding more for both, (P1, P2) = ($18, $30) or defect, move down, yielding (P1, P2) = ($6, $42). The results: Of 49 pairs, 45% of Player 1's opt out; 55% pass to Player 2; and 67% of Payer 2's cooperate, 33% defect. Thus we replicate the large number of similar such game results reported in the 1990s.

The neo-classical mental model, Max-U (own), predicts that Player 1 will not pass to Player 2 because Player 2 will move down in her own strict self-interest. Obviously, that model fails by a huge two to one margin.

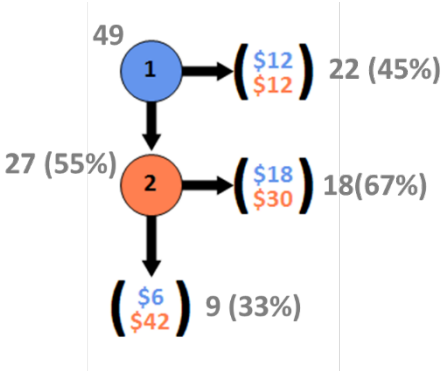

Figure 1. Baseline Trust Game

How do we modify our model of the actors in this game to better account for their actions? It is sufficient for preserving the utilitarian mental model if we give up the hypothesis that only one's own payoff matters. That is, imagine that action is still justified if and only if utility is maximized, but we interpret the negative evidence as merely falsifying the hypothesis that the utility function depends only on own payoff. Behavioral and experimental economists, thoroughly trained in the neo-classical Max-U tradition, have overwhelmingly chosen this route of postulating a utility representation of the form, U (own, other). With the specification of a just-so utility function of this interdependent form, behavioral and experimental economists thereby have rescued the calculus of Max-U.

Methodologically, however, this adaptation is flawed, raising many unanswered questions. Preferences, now dubbed "social," are at the mercy of empirical findings for fleshing out their fuller meaning. Early in the empirical exploration of factors effecting trust game cooperation, experiments demonstrated that "intentions" mattered. (McCabe et al., 2000; McCabe et al., 2003; Falk et al., 2008) Hence, "…equity models exclusively based on preferences

over the distribution of material payoffs cannot capture reciprocal behavior. Models that take players' fairness intentions and distributional preferences into account are consistent with our data, while models that focus exclusively on intentions or on the distribution of material payoffs are not." (Falk et al., 2008, p 287)

Consequently, it followed that if equity preferences models are to be properly "social" (including distributional) we must include both the joint material payoffs, and agent intentions in the utilitarian model of choice. Other scientific traditions reject this Ptolemaic procedure of adding observed new parameters to the utility function, in effect adding circles within circles to capture each new empirical discovery. Karl Popper's student, Imre Lakatose (1978) classified these cases as "degenerating programs" that commit to perpetually follow, rather than lead, each new empirical finding. Badly missing is a general coherent theoretical framework that implies these findings and derives new testable implications.

Applying Beneficence Proposition 1 to the Baseline Trust Game

In contrast with the traditional Max-U(own) mental model of action, the results in Figure 1 are consistent with Adam Smith's first proposition on beneficence. Two thirds of Player 2s, knowing that their Player 1s have passed to them, respond cooperatively. This is consistent with feelings of gratitude toward their Player 1 counterparts for having passed to them, and also self-command, the ability to resist the temptation to defect for the higher own reward. Moreover, Smith's model points to several sources of error indicating why one-third of Player 2s defect on the offer to cooperate: Player 2 may fail in

- reading intentions into Player 1's decision, reflecting an inability to put themselves into the circumstance of Player 1 and infer intent;
- feeling gratitude and rewarding the action;
- feeling enough gratitude to overcome the payoff foregone from cooperating;
- exercising self-command given the temptation to gain a higher reward.

These various sources of prediction error provide guidelines for further experimental designs and measurement. If, as suggested by Smith, the impulse to reward is proportionate to (increasing in) the gratitude felt, these elements can be expected to be responsive to variations in the payoffs.

The model also helps us to understand why more Player 2s respond cooperatively (67%) than Player 1s offer cooperation (55%); random assignment to the two positions implies each would act as the other if their role were to be reversed. The difference is that each Player 2 knows for certain the action of their Player 1 counterpart, whereas all Player 1s are uncertain as to the response of their Player 2s. The difference (12%) should vary predictably with what the players know about each other, and as we vary the context changing how the players read each other actions or expectations.

Trust Game with Punishment Option

Figure 2 shows a punishment version of the baseline game in Figure 1. The neo-classical Max-U (own) model prediction is the same as in the baseline trust game of Figure 2. Player 1 should move right at the top, and this model predicts none of the action we observe in the rest of the tree. Smith's Justice Proposition 1 above predicts that the Player 1s who pass to Player

2s, who in turn defect on the offer to cooperate, will feel resentment for this hurtful response to Player 1s well-intetideoned offer to cooperate. That resentment will tend to provoke a punishment response. To be credible, the feelings should be strong enough that Player 1 is

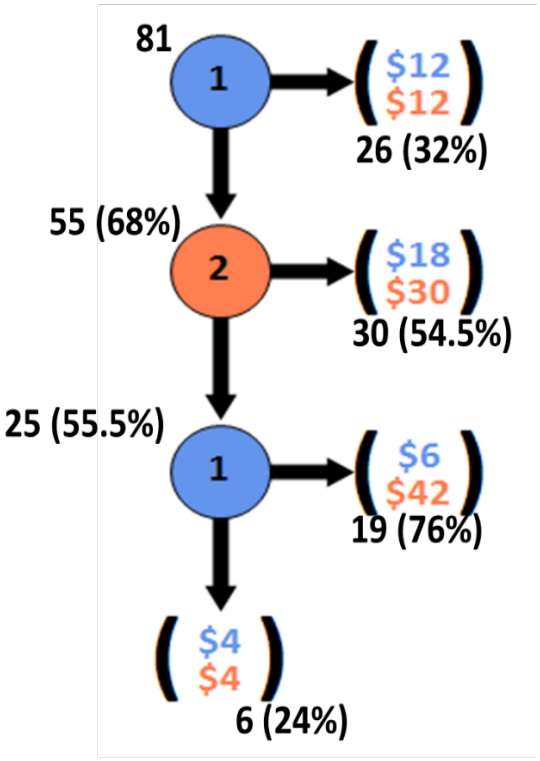

Figure 2. Trust Game: Punishment version

willing to incur a cost to punish Player 2. The parameterization in Figure 2 provides for a low cost (Player 1's payoff drops from $6 to $4), but very high punishment of Player 2 (whose payoff falls from $42 to $4). In conformity with Justice Proposition 1, 24% of Player 1s invoke the punishment. Why do less than a one-quarter of the Player 1s invoke this low cost, high punishment option? This, we suggest, is because in Smith's model people expect the punishment to fit the crime, neither excessive nor inadequate, for "the impartial

spectator…never, even in thought, attempts any greater vengeance, nor desires to inflict any greater punishment, than what every indifferent person would rejoice to see executed." (TMS, p 24)

Comparing Figure 2 with Figure 1, however, we note that adding the new punishment node changes the frequency of choice at both of the other nodes. We learn repeatedly, you might say incessantly, from Smith, that circumstances or context (that means all decision nodes and payoffs) matter; adding a node, even if it is payoff dominated (the player choosing it is strictly worse off) changes how the players read actions as signals. Thus, down moves by Player 1 increase from 55% in the baseline "pure trust" game to 68%. Why? The answer is in our discussion above of the greater implicit uncertainty, and its effect, that Player 1 faces in not knowing what Player 2 will choose. Because Player 1 now has the option to punish defection in Figure 2, some Player 1s, who would play right in the base line game of Figure 1, are induced to play down in Figure 2. But that reactive choice fails to anticipate how Player 2s read the move. Comparing the frequencies of choice by Player 2s, defection increases from 33% to a whopping 55.5%. Here, Player 1s badly misread their counterparts, an "anomaly" that cries out for further study; Player 1s read Player 2s much better in the baseline pure trust game.

Experimental economics is about nothing if it does not include the study of sources of error—the theme for all science in Mayo (1996).

Trust Game Tests of Beneficence Proposition 2

According to this proposition, because beneficence is always freely given, the failure to take deliberate action to benefit another is not an action that others resent, and feel impelled

to punish in response—no “real evil” is done. Your right to forego such an option, without retribution, is recognized.

We apply this proposition to the trust game of Figure 1 by adding a node; if Player 1 moves right to select the equilibrium of the game ($12, $12), play passes to Player 2 who chooses this option or at a cost to herself punishes Player 1 with a lower payoff. Figure 3 shows the implementation we test, where Player 2 chooses between ($12, $12) and ($10, $10).

We first ran sessions totaling 25 pairs; 15 Player 1s moved right, but not a single Player 2 chose to punish the action. (Smith and Wilson, 2017) The outcome seemed not credible. Theory never predicts that well! We decided to increase simple size to 38 by running another 13 pairs, as in Figure 3. Of 38 pairs, 23 Player 1s moved right, but no Player 2 punished the action.

However, comparing the results in Figure 1 and Figure 3, we see that adding the option to punish failure to show beneficence reduces Player 1s frequency of offers to cooperate from 55% to 39%. Good decision, because Player 2s’ cooperative response declines from 67% to 47%. The option to punish a right move at the top, changes how Player 2s read a down more by Player 1; the move conveys less trust, inducing more Player 2 defections.

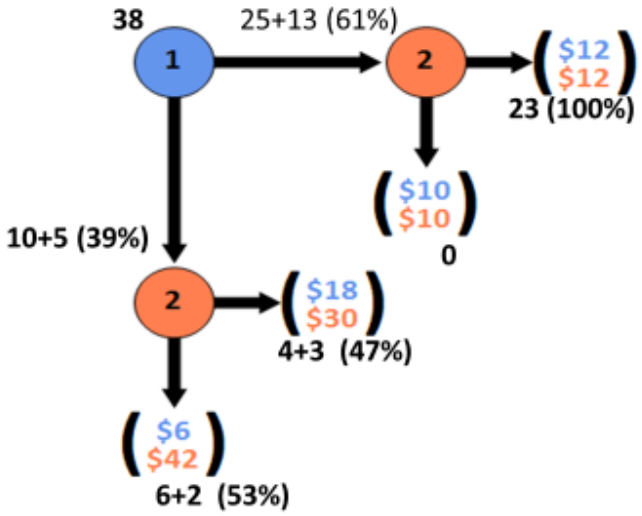

Figure 3. Trust Game with Option to Punish Want of Beneficence

Implications of Beneficence Proposition 2 for the Ultimatum Game: The Voluntary Ultimatum Game

In the ultimatum game (UG) the Proposer offers a split of M one-dollar bills to the Responder, M—X for the proposer and X for the Responder. The Responder either accepts the offer and the offer amounts are paid, or the responder rejects the offer in which case each player receives zero dollars. If it is common knowledge that both players are strictly self-interested, and locally non-satiated, then each prefers more money, dis-prefers less, and the Proposer should offer X = 1 to the responder and keep M—1 for herself. Rational Max-U means that $1 for the Responder is better than zero, and the offer is predicted to be accepted. Hoffman et al. (1994, 1996) report data on offers and rejection rates under a variety of different instructional contexts for the

division of $10, and extend the comparisons for stakes of $100. Proposers tend to offer far more than $1. The mean offer is about 45% of the total, whereas the median and mode is 50%. A binary choice version of the UG is shown in Table 1 by Falk et al. (2003) where the highly replicable results are typical across a large literature.

The explanation for the strong tendency toward equal-split of M is "fairness" in the outcome sense. But the argument is circular: Equal division is "fair" behavior. But what is "fair"? Equal division. Moreover, as Smith and Wilson (2018) show, "fair" choice behavior is indistinguishable from "non-envious" behavior, recalling the challenge to science of Henri Bertoft (footnote 2).

Understanding the pattern of results in Table 1 has constituted one of the major challenges for behavioral and experimental economists. Thus, in row one, the experimenter

Table 1

Ultimatum Game Offers and Rejections Reported in Falk et al. (2003)

Alternative to (8, 2) that Proposer can Offer	Frequency at which Proposers Offer (8, 2)	Proportion of (8, 2) Offers Rejected
(5,5)	0.31	0.444
(10, 0)	1.00	0.089
(2, 8)	0.73	0.267
(8,2); no alternative	1.00	0.18

requires the Proposer to offer either (5, 5) or (8, 2). Thirty-one percent offer (8, 2) and 44.4% of these offers are rejected (69% offer equal division and are all are accepted). On the face of it, in a world where we think of preferences as driving the choice of actions, and actions determining outcomes, this seems like an unambiguous expression of preference for equality of outcome or "fairness" (no scholar wants to call it "non-enviousness" although formally equivalent). Thus, "Simple games test game-theoretic principles in the clearest possible way. " "Since equilibria are so simple to compute…the ultimatum game is a crisp way to measure social preferences rather than a deep test of strategic thinking..." (Camerer, 2003, p 9). Note that these strong and unambiguous interpretations and views are driven by thinking of the form: action→outcome (M—x, x)→ Preference (M—x, x). But this, as we have seen, is not the

thought process in *Sentiments*, where the *conjunction of circumstances and outcomes matter in determining action*.

Viewed in this way, the data in Table 1 are a fountain of puzzlements. Consider row 2. Given a choice between offering (10, 0) and (8, 2), every Proposer offers (8, 2), but 8.9% are rejected! Responders at the rate of 8.9% do not like this outcome. Why? Well, it is said to reflect very strong fairness attitudes. Again, when the alternative is (2, 8), 73% offer (8, 2) which is rejected by 26.7%. And when only (8, 2) can be offered, 18% are rejected. But how is it fair to punish the proposer who does his best to be other-regarding when the equal-split outcome is unavailable, as imposed by the experimenter? Surely these responses are messages for the experimenter as much as for the proposer subject, where the responder is expressing dissatisfaction with the circumstances of the game.

Indeed, many studies have followed this path of interpretation and shown that responders are expressing their emotions of anger. Xiao and Houser (2005) show that when responders are given the option of expressing their anger toward the proposer, as an alternative to rejecting the proposers offer, they use the option and accept the offer. Other studies providing evidence for UG behavior as an emotional response include Palatal and Murnighan (1996), O'Connor et al. (2002), Sanfey et al. (2003), and Van't Wout et al. (2006).

As we have indicated, Beneficence Proposition 2 states that want of beneficence provokes no resentment, and hence no desire to punish. But the fact that beneficence is always free, and cannot be extorted, also implies that, in the presence of extortion, the calculus of benefit-reward does not apply. Where action is involuntarily imposed, the ordinary rules and

calculus cannot be expected to apply. Hence, if the proposer offers more than the equilibrium amount to responder, it is not out of the goodness of his heart, but under the duress of veto power. Knowing this, it is not possible for the responder to feel gratitude and a requirement to reward the action by accepting it. If more is offered, and it is accepted, these cannot be the motives. Indeed, both parties are involuntary players whose resentment is like that of reluctant dualists, except that the circumstances are imposed by the experimenter, not societal norms.

So what sort of calculus does apply? Based on TMS, I think we do not know; we only know what does not apply, namely that of benefit-reward. Accordingly, we begin by asking if it matters whether the responder has the option to choose, or not, to enter the UG along with his/her paired proposer. If a person does so choose, then Beneficence Proposition 2 implies that the context is coercive (or can be so interpreted) and we have some empirical proof of concept.

A test of the hypothesis that it matters if the responder can voluntarily choose or not to enter the UG, is shown in Figure 4 for the division of the fixed sum of $24. (Smith and Wilson, 2018) Observe that, contingent on entry, the proposer chooses between an equal split of $24 and an 11 to 1 split—far more unequal than the options available in Table 1, or in any data set that we know of. Most (94%) Player 1 choose to pass to the Player 2 proposer, 40% offer the equilibrium amount, and 61% accept. This data supports the equilibrium outcome and at rates far in excess of any reported in the literature for the population represented.

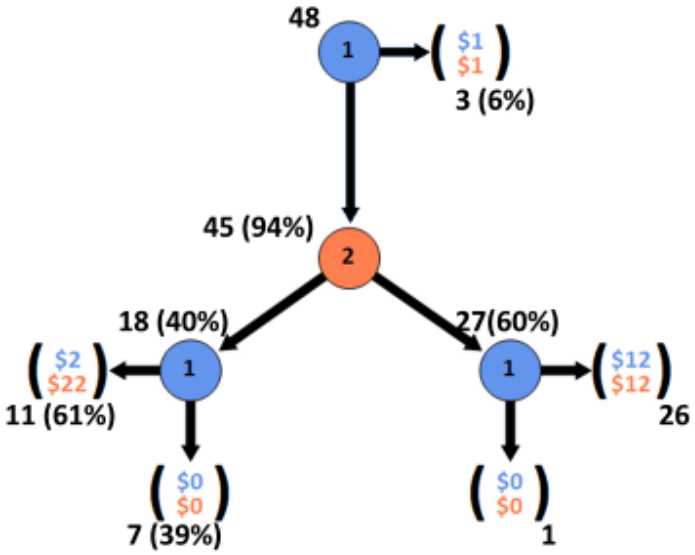

Figure 4. The Voluntary Ultimatum Game for the Division of $24

These results are consistent with "anomalous" results reported by Yamagishi et al. (2012). Gintis (2000) had introduced the concept of "strong reciprocity," a "prosocial" behavioral hypothesis.[21] Yamagishi et al. (2012, p 20366) ask whether individuals who reject "unfair" offers in the UG, exhibit similar prosocial behavior in other games, e.g., as trustees or trustors in trust games. "A strong reciprocator is characterized as having a disposition toward both positive and negative reciprocity, so it follows that strong reciprocators who reject unfair offers

[21] Without the challenges to the over simplifications of the homo economicus model by Herb Gintis, I am not sure when we would have been led to discussions like that in this section. But I believe his valiant attempt to rescue the neo-classical Bentham-Jevons framework with modern equilibrium notions has failed, because of the inadequacy of that framework.

in the UG should behave in a fair manner and should reciprocate positively in other games. This prediction was clearly rejected by our findings." Yamagishi et al. (2012, p 20366) Adam Smith's propositions clarify the results obtained in these two different game circumstances, and are consistent with these findings, but not with the model of prosocial behavior.

SUMMARY

Art Denzau and his co-authors provided the impetus in this paper for re-examining alternative mental models of action in socio-economics.

Modeling individual actions in social and economic contexts can either take the perspective of their external consequences, with the intent of deriving the implications for society or economy, or that of the actor with the intent of identifying the sources of their motivation to action. Economists in the neoclassical and modern tradition almost universally take a consequentialist perspective that focuses on outcomes, which is identical to the source of motivation provided that its basis in utility maximization is also the actor's only motivation.

The classical tradition as expressed in TMS and WN did not follow that pathway. In TMS, although all actors are naturally and rightly self-interested, not all action is driven by self-interested choice, for we all begin learning around age six that some of our actions lead to gratitude in others, and a requirement for them to reward the actor; but other actions lead to resentment and a desire to punish the actor. All such actions, however, depend on common knowledge that all are self-interested, otherwise we have no way of knowing who benefits or is hurt by an action we select. And in WN, both property rights as a necessary condition, and the propensity to trade as a sufficient condition for wealth creation can be interpreted as based on

the beneficence and justice propositions articulated in TMS. In this vision, neoclassical economics was a diversion from TMS, a work that was in turn central to a full and effective understanding of WN. In both works Smith modelled the actor first, then the consequences of their actions. If what I say here is somewhat more than you can prove was said in these works that is because Smith's mental model, when combined with my own explorations, has carried me further down the intellectual branch of socio-economics that he pioneered with his friend, David Hume.[22] It is they, however, that made such developments possible.

Reference

Alexander, Samuel . 1933. *Beauty and other Forms of value.* Collected Works of Samuel Alexander, Volume 4. London: Macmillan.

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

before The British Association for the Advancement of Science. Reported in Howey, R. S. (1989, pp 16-18) *The Rise of the Marginal Utility School* New York: Columbia University Press.

[22] These works were sources of insightful answers to experimental game-theoretic questions that had long bugged me, and that I sought to generalize. Non-experimentalists are less likely to have made such a connection. Hence, I am sure that I am seeing more in Smith than can be sustained by Smith scholars dedicated to text analysis. None of that changes my theme that socio-economics suffered a huge loss in its race to create a new and better program to replace the classical tradition. The power of that tradition is evident in the fact that so much of its foundation has re-emerged in modern experimental scholarship and research.

Baron-Cohen, Simon (1995) *Mindblindness an essay on autism and theory of mind.* Cambridge Mass.: MIT Press.

Bertoft, Henri. 1996. *The Wholeness of Nature: Goethe's Way of Science*. Glasgow: Floris Books.

Bronfman, C., K. McCabe, D. Porter, S. Rassenti and V. Smiyh (1996) "An Experimental Examination of the Walrasian Tatonnement Mechanism," *Rand Journal of Economics*, 27 pp. 681-699.

Caginalp G and D. Balenovich (1999) "Asset flow and momentum: Deterministic and stochastic equations." Philosophical Transactions Royal Society of London A 357, 2119–2133.

Caginalp G., D. Porter D and V. Smith (2001) "Financial bubbles: Excess cash, momentum, and incomplete information." J. Psychological Finance and Markes, 2:80–99.

Caginalp, Cary and Gunduz Caginalp (2018) "Valuation, liquidity price, and stability of Cryptocurrencies." *Opinion, Proceedings National Academy of Science*, February 6, vol. 115 (6) 1131–1134.

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

Chamberlin, Edward (1948) "An Experimental Imperfect market." Journal of Political Economy, 56 (2) pp 95-108.

Davis, Douglas D. and Charles A. Holt. 1993. *Experimental Economics.* Princeton: Princeton University Press.

Denzau, A.T. and North, D.C. (1994) "Shared mental models: ideologies and institutions," *Kyklos* 47.

Denzau, A.T. and Roy, R.K. (2005) “Mental models and game theory: cognitive constructions of multiple Nash equilibria,” Presented at Public Choice Society, March.

Denzau, Arthur T., Douglass C. North and Ravi K. Roy (2014) Shared Mental Models

Falk, Armin, Ernst Fehr and Urs Fischbacher. 2003. “On the Nature of Fair Behavior.” *Economic Inquiry* 41 (1): 20-26.

Falk, Armin, Ernst Fehr, Urs Fischbacher. 2008. “Testing Theories of Fairness—Intentions Matter,” *Games and Economic Behavior* 62(1): 287-303.

Gintis H (2000) “Strong reciprocity and human sociality.” *J. Theor. Biol.*, 206(2):169–179.

Gjerstad, Steven (2013) “Price dynamics in an exchange economy.” *Economic Theory,* 52:461–500. DOI 10.1007/s00199-011-0651-5

Haig, David (2011) “Sympathy with Adam Smith and reflexions on self,” *Journal of Economic Behavior & Organization* 77, pp 4–13.

Hoffman, Elizabeth, Kevin McCabe, Vernon L. Smith, and Keith Shachat. 1994. “Preferences, Property Rights and Anonymity in Bargaining Games,” *Games and Economic Behavior*, 7, pp. 346-380.

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

Hume, David. 1740. *A Treatise of Human Nature*, David Fate Norton and Mary J. Norton (eds.). Oxford University Press: New York, NY. (2000).

Inoua, Sabiou (2018) “A Rehabilitation of Classical Economics” Draft paper, private correspondence, July 22.

Jevons, W. S. (1862) “Notice of a General Mathematical Theory of Political Economy.”

Read before the British Association for the Advancement of Science, September. Reported by Richard S. Howey (1989) The Rise of the Marginal Utility School. New York: Columbia University Press, pp 16-18.

Jevons, W. S. 1871. *The Theory of Political Economy.* London: Macmillan, 3rd edition 1988.

Kahneman, Daniel and Amos Tversky. 1979. “Prospect Theory: An Analysis of Decision under Risk,” *Econometrica* 47(2): 263-291.

Lakatose, Imre (1978) *The Methodology of Scientific Research Programmes: Volume 1: Philosophical Papers.* Edited by John Worrall and Gregory Currie. Cambridge: Cambridge University Press.

Lei, Vivian, Charles N. Noussair, and Charles R. Plott. 2001. “Nonspeculative Bubbles in Experimental Asset Markets: Lack of Common Knowledge of Rationality vs. Actual Irrationality.” Econometrica, 69(4): 831–59

Magnani, J. and Oprea, R. "Why Do People Violate No-Trade Theorems? A Diagnostic Experiment" at http://www.ryanoprea.com/

Mayo, Debora G. (1996) *Error and the Growth of Experimental Knowledge.* Chicago: University of Chcago Press.

McCabe, K., and S. Rassenti and V.L. Smith. 1993. "Designing a Uniform Price Double Auction: An Experimental Evaluation," in D. Friedman and J. Rust, eds., *The Double Auction Market: Institutions, Theories and Evidence.* Reading: Addison-Wesley/SFI, pp. 307-332.

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.

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.

Marshall, Alfred (1890) *Principles of Economics.* London: MacMillan.

O'Connor, Kathleen M., Carsten K. W. De Dreu, Holly Schroth, Bruce Barry, Terri R. Lituchy, and Max H. Bazerman (2002) "What We Want to Do Versus What We Think We Should Do: An Empirical Investigation of Intrapersonal Conflict," *Journal of Behavioral Decision Making*, 15(5): 403-418.

Perner, J., U. Frith, A. Leslie and S. Leekham (1989) "Exploration of the Autistic Child's Theory of Mind: knowledge, belief and communication." *Child Development*, 60, 689-700.

Pillutla, Madan M. and J. Keith Murnighan (1996) "Unfairness, Anger, and Spite: Emotional Rejections of Ultimatum Offers," *Organizational Behavior and Human Decision Processes*, 68(3): 208-224.

Sanfey, Alan G., James K. Rilling, Jessica A. Aronson, Leigh E. Nystrom, and Jonathan D. Cohen (2003) "The Neural Basis of Economic Decision Making in the Ultimatum Game," *Science*, 300: 1755-1758.

Smith, Adam. 1759. *The Theory of Moral Sentiments.* Edited by D. D. Raphael and A. L. Macfie. Oxford: Oxford University press, 1976. The Dugald Stewart edition is available online Available online and in electronic formats at http://oll.libertyfund.org/titles/2620.

Smith, Adam (1776) *An Inquiry Into the Nature and Causes of the Wealth of Nations* (Cannan ed.), vol. 1. London: Methuen, 1904 http://oll.libertyfund.org/titles/smith-an-inquiry-into-the-nature-and-causes-of-the-wealth-of-nations-cannan-ed-vol-1

Smith, Adam (1982) *Lectures on Jurisprudence*, Edited by R. L. Meek, D. D. Raphael, and P. G. Stein. Oxford: Oxford University Press.

Smith, Vernon L. (1962) "An Experimental study of Competitive Market Behavior," *Journal of Political Economy, 70,* pp 111-137.

Smith, Vernon L. (1981) "An Empirical Study of Decentralized Institutions of Monopoly Restraint," *Economic Essays in Honor of E.R. Weiler*, Editors George Horwich and James P. Quirk, (Purdue University Press).

Smith, Vernon L. 1982. *Papers in Experimental Economics.* Cambridge: Cambridge University P.

Smith, Vernon L., G. Suchanek and Arlintom W. Williams (1988) "Bubbles, Crashes and Endogenous Expectations in Experimental Asset Trading Markets" *Econometrica,*

Smith, Vernon L. 2012. "Adam Smith on humanomic behavior," Edited transcript of keynote address, Academy of Behavioral Finance and Economics, UCLA. *Journal of Behavioral Finance & Economics,* 2, pp. 1-20.

Smith, Vernon L. and Bart J. Wilson. 2017. "*Sentiments*, Conduct, and Trust in the Laboratory," *Social Philosophy and Policy* 34(1): 25-55.

Smtih, Vernon L. and Bart J. Wilson (2018) *Humanomics: Moral Sentiments and the Wealth of Nations for the Twenty-First Century.* Cambridge: Cambridge University Press.

Sunder, Shyam. *1995. "Experimental Asset Markets," in J. Kagel and A. Roth, editors, The Handbook of Experimental Economics* .Princeton: Princeton University press.

Svorenčík, Andrej (2015) *The Experimental Turn in Economics.* PhD Dissertation

Van't Wout, Mascha, René S. Kahn, Alan G. Sanfey, and André Aleman (2006) "Affective state and decision-making in the Ultimatum Game," *Experimental Brain Research*, 169(4): 564-568.

Walras, Leon (1969) Elements of Pure Economics or The Theory Of Social Wealth Translated By William Jaff£. New York: A.M. Kelly.

Xiao, Erte and D. Houser (2005) "Emotion expression in human punishment behavior." *Proc Natl Acad Sci USA* 102(20):7398–7401.

Yamagishi, Toshio, Yutaka Horita, Nobuhiro Mifune, Hirofumi Hashimoto, Yang Li, Mizuho Shinada, Arisa Miura, Keigo Inukai, Haruto Takagishi, and Dora Simunovic (2012) "Rejection of unfair offers in the ultimatum game is no evidence of strong reciprocity." *Proc. Nat. Acad. Sc. USA,* December 11, 109 (50), 20364–20368.