Social Cooperation and the Apparent Tension between Personal and Impersonal Exchange*

Peter T. Leeson
Department of Economics
West Virginia University

Abstract

Experimental research provides evidence for both agents as noncooperatively self-interested *and* cooperatively other-regarding. Drawing on the work of F.A. Hayek, Vernon Smith attempts to resolve this apparent tension by distinguishing between two spheres of human interaction. According to V. Smith, in the personal sphere individuals are psychologically hardwired for cooperation and in the impersonal sphere they are hardwired for noncooperative behavior. Building on the solution offered by V. Smith, this paper offers a *rational choice* approach to seemingly anomalous experimentally observed behavior. Using the Ricardian Law of Association, I reconstitute trust game decision trees to resolve the apparent tension in experimental results.

* I am grateful to the F.A. Hayek Fellowship and the STICERD Center at the London School of Economics for generously providing support of this research. Additionally, I would like to thank Peter Boettke, Tyler Cowen, Chris Coyne, Steve Daley, Orson Francescone and Peter Jaworski for helpful comments and suggestions.

1. Introduction

The work of V. Smith (1998) points to an apparent tension in economic theory that originates in the writings of Adam Smith. This tension has long been recognized by economists—so much so that it took on a name on of its own—"das Adam Smith problem," reflecting its origin in debates regarding Smith by German social scientists in the 19th century. "Das Adam Smith problem" emerges out of two seemingly antithetical notions of what motivates individuals in Smith's writings. On the one hand Smith emphasizes noncooperative self-interest as the driving force of the market economy ([1776] 1976: 26-27). On the other hand he maintained that individuals are other-regarding and that cooperative behavior makes the market operate smoothly ([1759] 1976: 9).

V. Smith has pointed out that experimental economics offers support for both apparently antithetical assumptions. Trials of double auction markets offer evidence supporting the standard assumption that individuals are noncooperatively self-interested. The self-interested actions of anonymous individuals create competitive equilibrium just as theory predicts. As V. Smith reports, in these experiments, economic "[e]fficiency, as computed as a percentage of the consumer surplus that is realized by all subjects, always tends to approach 100%" (V. Smith 1998: 9).

This being the case, experimental trials of trust games should come to the same conclusion.

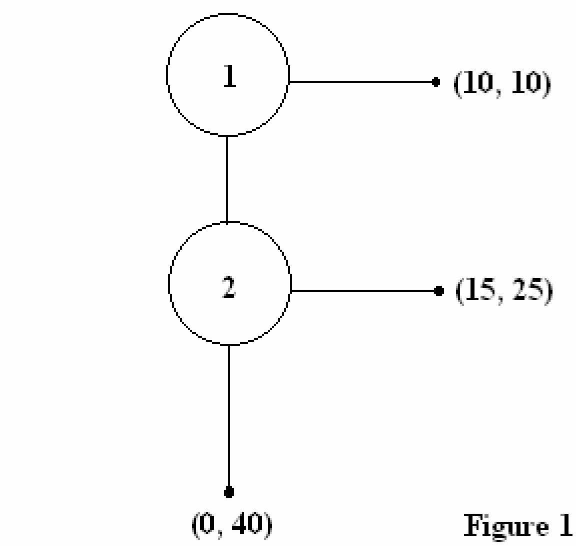


Figure 1.

The standard trust game decision tree is shown above in Figure 1. In these games Player 1 has the option of either passing decision power and the possibility of mutually higher payoffs to his partner (move from decision node 1 to decision node 2), or ending the game right there for a lower payoff (move right at decision node 1). If Player 1 passes to her partner, her partner can either reward Player 1 by giving both Player 1 and himself some payoff larger than the payoff Player 1 could get by not passing decision power to Player 2 (move right from decision node 2), or he can take an even larger payoff yet leaving Player 1 with nothing (move down from decision node 2). Thus, Player 1 must initially decide whether or not to trust Player 2 with the power to take advantage of the fact that she did not defect in round one, leaving both with the chance to earn more. The game theoretic prediction of this game is termination at the first payoff node by Player 1 where the payoff is (10, 10). Assuming both players are self-interested, backward induction dictates that the game should always end at payoff node 1, the subgame perfect Nash equilibrium.

Despite this, the majority of experimental trials of this game fly in the face the game theoretic prediction. Many trials show 50% of Players 1 handing the decision off to Players 2, and of those, 75% of Players 2 selecting the cooperative payoff (V. Smith, 1998).[1] Thus, there appears to be strong contradictory evidence to double auction trials that instead finds individuals to be other-regarding. This is depicted below in Figure 2.

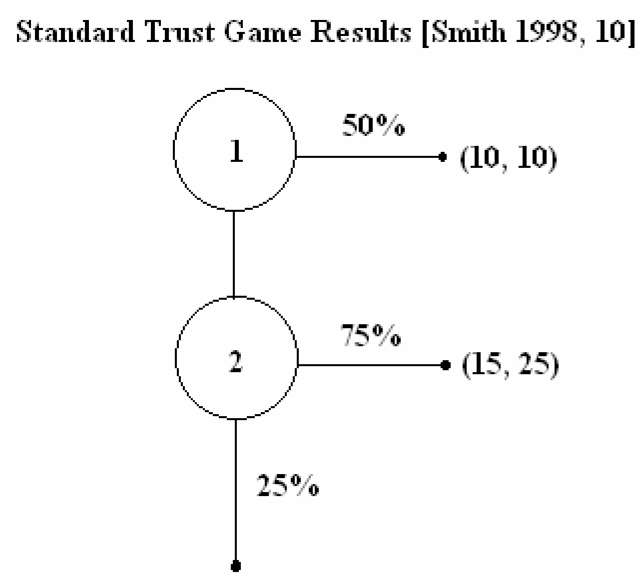


Figure 2.

Building on Hayek (1973), V. Smith resolves this apparent tension by contending that actors are "hardwired" for cooperative behavior in their personal interactions and non-cooperative behavior in their impersonal interactions. While this explanation addresses the apparent tension it fails to provide a rational choice explanation for the seemingly inconsistent behavior we observe. Because the peculiarity of humans is their ability to override instinctual impulses—i.e., their ability to rationally choose—a *rational choice explanation* seems both a more appropriate and more satisfying approach to our problem. Indeed, once it is acknowledged that agents override their impulsive

[1] Other experimental trials of this game show even higher levels of cooperation. See, for example, trials conducted by Kevin McCabe and Vernon Smith (2000).

psychological instincts every day, the troubling question emerges: why is it then that they overwhelmingly fail to do so in experimental trials of double auctions and trust games? Why, for instance, don't experimental trials provide significant evidence of agents defecting in trust games?[2] The most obvious answer is that agents choose to cooperate in trust games because it is in their interest to do so. But once this is admitted we are already departing from the psychological instinct notion and beginning down the path of a rational choice explanation. My project here can be understood as more fully articulating the rational choice explanation that lies at the heart of the answer to the questions posed above.

Employing a rational choice approach does not entail abandoning the "hardwired" idea completely. This paper builds on both Hayek and V. Smith by beginning with the "hardwired" approach, but departs from them in its emphasis on the particular hardwiring of man's mind. While Hayek and V. Smith focus largely on *psychological* hardwiring, I will focus on individuals' *rationality* hardwiring—the standard starting point of rational choice explanations. This paper's approach should also be distinguished from approaches like those of Frank (1987), which suggest that other-regarding preferences are selected because of their evolutionary superiority.[3] In contrast to this I will take the more standard economic approach that takes agents' utility functions as given and as consisting only of their own narrow self-interest. I hope to show that narrowly self-interested, rational actors often choose to cooperate because they understand that doing so promotes

---

[2] I say "significant evidence" because some percentage of subjects do defect. However, such cases are the minority. On this, see fn. 7.

[3] Frank's approach, as he notes, is in contrast "essentially the behavioral biologist's approach" (1987: 593). In this sense, it can be seen as a variant of V. Smith's psychological hardwiring explanation. See for example: (Hamilton 1964; Wilson 1975, 1978; Dawkins 1976; and Trivers 1985).

their own welfare. Using the Ricardian Law of Association (RLA) I reconstitute trust game decision trees to resolve the apparent tension in experimental results.[4]

My approach delivers a testable implication. Experimental subjects of trust game trials motivated by rationality hardwiring would characterize their cooperative behavior very differently from subjects motivated by other-regarding preferences or psychological hardwiring. Subjects asked to characterize their cooperative behavior or simply responding to the question, "Why did you cooperate?," along the lines of: "I cooperated because I would feel bad if I defected," would provide evidence for the other-regarding preference approach of Frank or the psychological hardwiring explanation of V. Smith.[5] On the other hand, subjects that responded to this question by stating that: "I cooperated because if we all cheated, we all would lose" or something similar would point to the validity of this paper's rationality hardwiring explanation. While duly noting the limitations of this line of questioning, it nonetheless should offer some data regarding the forces driving trust game participants' overwhelmingly cooperative behavior.

2. Personal and Impersonal Exchange

Hayek distinguishes between two types of relationships man has with others in two different spheres of his life—the "atavistic" sphere and the sphere of the "Great Society." The former is characterized by personalized relationships that necessitate cooperative, other-regarding behavior, while the latter consists of impersonal interactions and requires

[4] Although I limit my discussion to apparently anomalous cooperation in trust games, the solution found in the Ricardian Law of Association applies equally well to other experimental results that indicate higher than theoretically-predicted levels of cooperation. See, for instance, experiments concerning voluntary contributions to public goods (Andreoni 1988; Isaac, Walker, and Williams 1994; Palfrey and Prisbrey 1997).

[5] Frank seems to recognize that this is what experimental subjects cooperating for the reasons he describes would indeed respond. His approach implies that agents cooperate because "cheating simply makes them feel bad" (1987: 602).

noncooperative, self-interested behavior. Echoing this distinction, V. Smith has attempted to alleviate the tension in experimental results by distinguishing between two separate spheres in which individuals find themselves in society. In the sphere of personalized interaction, individuals behave cooperatively as trust game trials indicate. According to V. Smith, people act this way because their minds have evolved to instruct them to do so when confronted with a personalized setting. Over thousands of years of evolution, individuals learned the benefits of positive and negative reciprocation in personalized interaction and this behavior became hardwired as part of their response system. Consequently, “People are programmed for repeated social exchange. This is part of their natural instincts” (V. Smith 1998: 11). Similarly, the human mind evolved to instruct self-interested, non-cooperative behavior in impersonal settings. Thus, according to V. Smith, “Casual observation suggests the hypothesis that the same person can reciprocate toward individuals who are part of an ingroup to which the person belongs, but be noncooperative toward those in an outgroup for which the individual feels no affinity” (V. Smith 1998: 16).

Although experimental trials of trust games are one shot in nature, when individuals enter the realm of the lab to participate in the trust game experiment they cannot simply ‘check their past experiences at the door.’ Instead they bring with them all their personal experiences and this affects the way they behave in experiments. Specifically, individuals are habituated into the mechanics of repeated dealings and social cooperation in their real world personalized interactions where reciprocation plays an important role. As a result of these factors, one-shot experimental trials of trust games generate cooperative other-regarding behavior on the part of experimental subjects.

According to this view the "reciprocity instinct" is what leads individuals to behave differently in the different spheres of personal and impersonal exchange. As V. Smith puts it, "The key is to distinguish impersonal market exchange from personal social and economic exchange, and to understand that efficiency in the former is based on noncooperative behavior while efficiency in the latter requires reciprocity" (V. Smith, 1998: 8).

While V. Smith's hardwiring explanation addresses a solution to the apparent tension, its emphasis on psychological hardwiring as opposed to human rationality prevents a full realization of its insight. Unless we are to admit complete psychological determinism, it seems as though any answer we give to this question must come back to agents acting as they do because they find doing so in their interest. This fact strongly suggests that an approach that recognizes humans' rationality hardwiring—i.e., a rational choice explanation of experimental subjects' behavior is needed. By building on a variant of V. Smith's hardwired explanation that emphasizes individuals' rationality hardwiring, it is possible to resolve the apparent tension in experientially observed behavior.

3. The Tension Resolved: An Application of the RLA

Narrowly defined, the Ricardian Law of Association (RLA) deals only with so-called comparative advantage. It demonstrates the gains from specialization and trade even where one party has an absolute advantage in the production of everything (Ricardo [1817] 1821: par. 7.16). More broadly, however, the RLA can be seen as an explanation of the general benefits of widespread social cooperation based on the division of labor (Mises 1966: 160).

Fleshing out the idea that lies at the bottom of my answer to the question, 'why do so few individuals override their cooperative instinct in experimental trust games?,' the RLA reveals that ostensibly other-regarding behavior is really no such thing at all. Cooperation that appears to be other-regarding is merely a form of self-interested behavior. Specifically, an individual who acts in a seemingly other-regarding way is in fact acting in accordance with his enlightened self-interest. Just like individuals in the marketplace do not serve others out of benevolence but rather out of regard for their own self-interest, individuals in the marketplace do not refrain from cheating others out of benevolence but rather out of regard for their enlightened self-interest.[6] The result of such self-interested restraint is the same as the result of self-interested service—individuals are led through this process, as if by an invisible hand, to promote the interests of society. While in many instances it is surely possible to cheat and get away with it, individuals in the marketplace refrain from such activity because they have sufficient foresight to understand the longer-run consequences of their cheating.[7] Anti-social behavior undermines the system of social cooperation and destroys the benefits it confers on everyone. Cheating may generate a short-term gain, but this temporary gain comes at the expense of a permanent, much larger loss of potential wealth. Rational

[6] It is strange that so many economists have been quick to point out that self-interest, not benevolence, leads individuals to serve others in the marketplace and yet so few have pointed out that the same holds true for why indivduals refrain from harming others. Indeed, when it comes to explaining this behavior economists are prone to point to benevolence as the cause. It is not clear why this is the case, however, because the notion of individuals refraining from cheating their fellowmen from a regard for their self-interest is a perfectly natural extension of the notion that self-interest is what compels individuals to serve their fellowmen.

[7] There are, of course, exceptions to this tendency, just as there are exceptions to the tendency of subjects observed in experimental trials of trust games to behave in an other-regarding fashion. Some individuals are inherently anti-social in the sense that they are incapable of understanding the gains from social cooperation under the division of labor or lack the rationality hardwiring required to act on this understanding. Most individuals, however, are not this way but instead recognize the benefits from the social system of cooperation under the division of labor and act accordingly.

individuals acting in their enlightened self-interest recognize this fact and so abstain from anti-social activity.[8]

In terms of standard trust game results, the fact that large percentages of Players 1 pass decision power on to Players 2, and that the majority of Players 2 select the cooperative payoff, demonstrates that Players 1 and Players 2 at least implicitly understand the RLA and act in accordance with their enlightened self-interest. What has been traditionally characterized as other-regarding behavior is in fact quite the opposite. Players 1, rationally acting in their self-interest, pass to Players 2 with relative assurance that Players 2 will not cheat them because Players 1 understand that Players 2 are acting in accordance with their enlightened self-interest as well. When individuals act upon their enlightened self-interests, total wealth available to everyone expands. This is due to the RLA, which demonstrates that when social cooperation is enlarged greater wealth is generated for all.

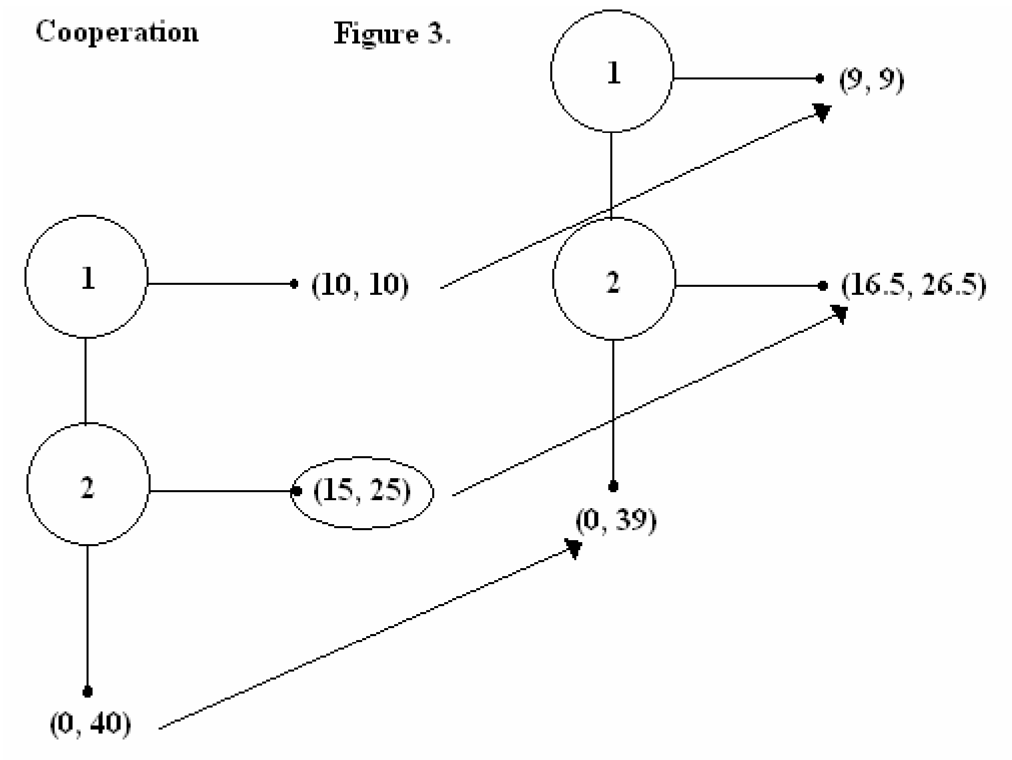
4. Modeling the RLA

While the trust game tested by experimental economists is a one shot game, experimental subjects invited to participate in the experiment are used to playing the trust game in the real world where interactions are not of a one shot nature. Since, as V. Smith points out, subjects are not able to simply 'check their experiences at the door,' they behave in the one shot trust game as they do in the real world trust games they actually find themselves in. Thus, to model the trust game as the lab subjects playing the game see it, multiple

[8] Ludwig von Mises stressed the fact that rational individuals refrain from anti-social behavior because it undermines social cooperation based on the division of labor and thus the wealth that this cooperation generates. He even maintained that individuals engage in the division of labor in the first place because they recognize the gains from doing so. According to Mises, "The division of labor is the outcome of man's conscious reaction to the multiplicity of natural conditions" (1966: 144).

decision trees must be modeled instead of one. It is worth noting that these real world trust games do not require repeated interaction between the same people to achieve the cooperative outcome. Individuals driven by their enlightened self-interests do not refrain from cheating others in the real world trust games they play because they fear punishment by others in repeated interaction. They refrain from cheating because whether they face the prospect of repeated interaction with the same people or not, they know that cheating will undermine their enlightened self-interest. In this way, individuals' behavior is not predicated on expectations of reciprocation. Individuals understand that cheating, whether they are likely to be punished or not, undermines the system of social cooperation that perpetually growing wealth requires. To cheat others would be to cheat oneself as the pie of available wealth shrinks with every anti-social action undertaken. The RLA also suggest that in addition to undermining social cooperation through defection, simply choosing not to engage in the system of social cooperation causes total wealth to shrink as well. The larger the number of individuals who choose self-sufficiency, the smaller the division of labor, and the less wealth that is generated. Like with cheating, this fact too is independent of individual expectations about reciprocation. The mere fact that cheating and self-sufficiency reduce the total wealth available to actors is enough to prevent anti-social behavior from permeating the market.

This information imparted by the RLA enables a reconstitution of the trust game decision tree as a series of decision trees as seen from the mind of the rational experimental subject. In terms of the trust game decision tree, cooperation causes the payoffs at the subsequent tree's cooperative node to increase, cheating (i.e., Player 2 playing down) causes *all* payoffs at the subsequent tree to decrease, and termination of the game at the first node causes *all* payoffs at the subsequent tree to decrease since this outcome represents autarky. The RLA-revised cooperative and cheating/autarkic decision trees and payoffs are shown below in Figures 3 and 4 respectively.



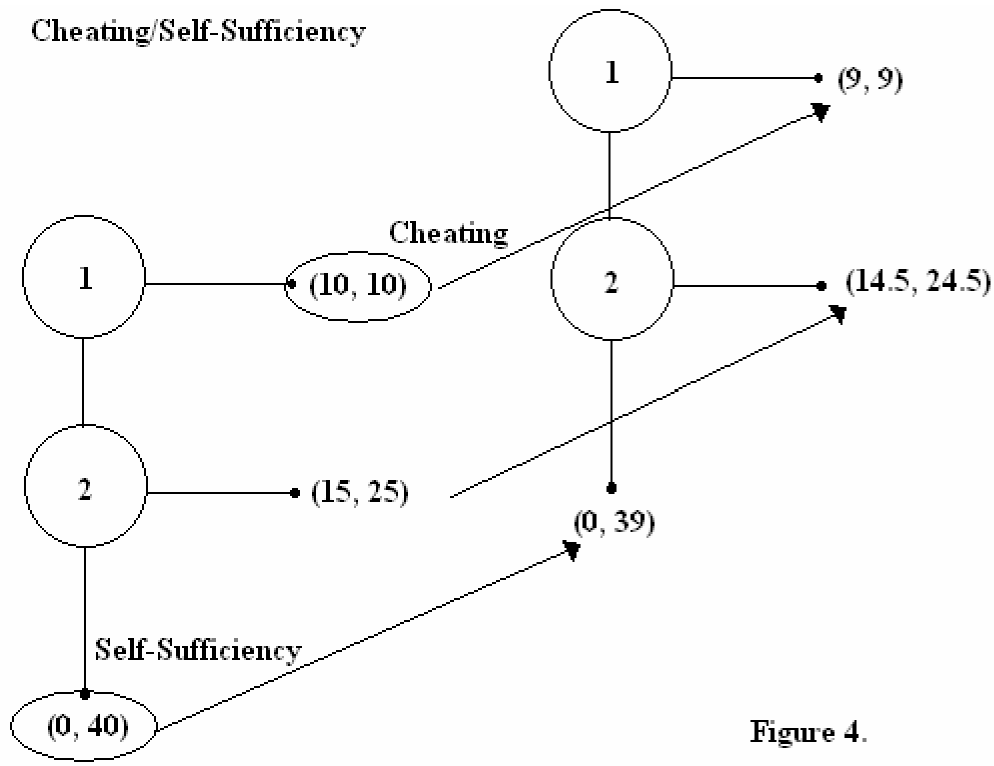


Figure 4.

Using the RLA, immediate game termination (the autarkic outcome) and defection by Players 2 (the social cooperation destroying outcome) represent fairly straightforward results—total wealth drops until cooperation is achieved, at which point total wealth rises until the autarkic or social cooperation destroying outcomes occur again. In other words, payoffs at *all* nodes drop with the autarkic or social cooperation destroying outcomes.

More interesting, however, is what happens to the payoffs at specific nodes using the RLA when cooperation is chosen. We know from the RLA that total wealth rises—so the payoffs at the cooperative node (payoff node 2) rise. But what happens to the payoffs at the autarkic and social cooperation destroying nodes (i.e., payoff nodes 1 and 3)? Using the RLA, the payoffs at these nodes should drop while the payoff at the cooperative node rises. This results from the fact that as social cooperation increases, the division of labor expands and individuals become increasingly reliant upon one another to satisfy their needs. This being the case, the costs of cheating and self-sufficiency rise. The more dependent individuals grow on others in society, the more destructive is undermining social cooperation through cheating, or exiting the system entirely, to their welfare (Mises, [1927] 1996: 25). As social cooperation expands, each individual's interest becomes more closely connected to the interests of all other members of society. Consequently, cheating others (or refusing to interact with them) comes closer and closer to cheating oneself. As cooperation continues, the payoff at the cooperative node continues to increase and the payoffs at the noncooperative and autarkic nodes continue to decrease.

If, after several rounds of cooperation, cheating occurs, a portion of the addition to social wealth accumulated over the previous rounds of cooperation is eroded. Thus, the payoffs faced at each node in the next round will be lower than in the previous one. If cheating occurred enough times following cooperation, it is possible to reduce the payoffs at each node to their level from the first round or even below be driven below this level. On the other hand, if after several rounds of cheating, individuals begin cooperating, the process of wealth destruction is reversed. Total wealth begins to climb again from its present, lower state. The payoffs at the cooperative node (node 2) in the subsequent round rise from the previous round, and those at the cheating/self-sufficiency nodes (nodes 3 and 1) decline. Over enough rounds of cooperation, the payoffs from this strategy can eventually surpass their levels from round one when the game began. The process of wealth creation and destruction over rounds of cooperation and defection is dynamic.

When viewed this way, the fact that we observe high levels of cooperative behavior in experimental trials of trust games ceases to be anomalous. Rationally self-interested agents serve their interests by making choices that expand the pie of social wealth. Since cooperation constitutes such a choice, it makes a great deal of sense that agents choose to cooperate. Furthermore, the RLA enables us to see that cooperative behavior in trust games is a form of self-interested behavior. Rather than antithetical to the behavior observed in double auction trials, cooperation in trust games is entirely consistent with the self-interested actions that characterize double auctions.

5. Conclusion

The apparent tension between agents' cooperative, seemingly other-regarding behavior and their noncooperative self-interested behavior goes back as far as Adam Smith. V. Smith's "hardwired" explanation used to address this tension rediscovered in experimental results over two hundred years later serves as a starting point for resolving the tension more fully. Following Hayek, however, his emphasis on *psychological* hardwiring of "natural instincts" fails to provide us with a rational choice explanation for the seeming tension experimental results provide. As soon as we recognize that humans' distinguishing feature is precisely their ability to override their "natural instincts," we are left with the unanswered question: "why then don't more experimental subjects defect in trust games?" The immediate answer to this question is that subjects choose to cooperate because they find this in their interest. This answer suggests agents' rationality hardwiring and highlights the importance of elaborating the rational choice explanation that ultimately lies at the heart of my description of why we observe the experimental behavior that we do.

The Ricardian Law of Association generally describes the growth in social wealth achieved through cooperation and the destruction of social wealth that occurs when agents cheat or act in isolation. The RLA thus provides a rational choice basis for evaluating the apparent tension in double auctions and trust games. What seems to be other-regarding behavior in trust game cooperation is actually self-interested behavior on the part of rational agents that recognize the implications for social wealth of cooperation vs. anti-social behavior as described in the RLA.

With the aid of the RLA, we are able to reconstitute the standard trust game decision tree confronted by experimental subjects as a series of decision trees with payoffs shaped by the implications of the RLA, as seen from the mind of rational experimental subjects. Doing this allows us to understand why far from anomalous, high levels of cooperation in trust game trials are reasonable and consistent with the self-interested behavior observed in double auction trials.

This approach offers a testable implication based on how trust game participants characterize their cooperative behavior or answer the question, “Why did you cooperate?” Respondents that indicate they cooperated because they would feel bad or guilty for defecting would support the approach of Frank or V. Smith. It is my contention, however, that respondents would indicate an understanding of the greater consequences for social wealth and well-being of cooperation vs. cheating. Lab data predicated on this testable implication have, to my knowledge, yet to be conducted, offering an avenue for future research.

References

Andreoni, James. "Why Free Ride? Strategies and Learning in Public Goods Experiments." *Journal of Public Economics* 1988, 291-304.

Dawkins, Richard. *The Selfish Gene*. New York: Oxford University Press, 1976.

Frank, Robert H. "If Homo Economicus Could Choose His Own Utility Function, Would He Want One With a Conscience?" *American Economic Review* 1987, 593-604.

Hamilton, W. D. "The Genetical Theory of Social Behavior." *Journal of Theoretical Biology* 1964, 1-32.

Hayek, F.A. *Law, Legislation and Liberty*. Chicago: University of Chicago Press. 1973.

Isaac, Mark, Walker, James and Williams, Arlington. "Group Size and the Voluntary Provision of Public Goods: Experimental Evidence Utilizing Large Groups." *Journal of Public Economics* 1994, 1-36.

McCabe, K. and Smith, V. "A Comparison of Naïve and Sophisticated Subject Behavior with Game Theoretic Predictions." *PNAS* 2000, 3777-3781.

Mises, L.v. *Liberalism*. Irvington-on-Hudson, NY: Foundation for Economic Education, 1996.

____________. *Human Action: A Treatise on Economics*. San Francisco: Fox & Wilkes, 1966.

Palfrey, Thomas and Pilfrey, Jeffrey. "Anomalous Behavior in Pubic Goods Experiments: How Much and Why?" *American Economic Review* 1997, 829-846.

Ricardo, D. *On the Principles of Political Economy and Taxation*. London: John Murray, Albemarle-Street, 1821.

Smith, A. *An Inquiry Into the Nature and Causes of the Wealth of Nations*. Oxford: Oxford University Press, 1976.

____________. *The Theory of Moral Sentiments*. Oxford: Oxford University Press, 1976.

Smith, V. "The Two Faces of Adam Smith, Distinguished Guest Lecture." *Southern Economic Journal* 1998, 1-19.

Trivers, Robert L. "The Evolution of Reciprocal Altruism." *Quarterly Review of Biology* 1971, 35-57.

____________. *Social Evolution*. Menlo Park: Benjamin/Cummings, 1985.

Wilson, Edward. *Sociobiology*. Cambridge: Harvard University Press, 1975.