Are we capable of being altruistic?

First Prize – Postgraduate Category

PANAYIOTIS AGISILAOU[†]

AND FREDERICK WANDSCHNEIDER[‡]

"*Man is neither angel nor brute, and the unfortunate thing is that he who would act the angel acts the brute."* Blaise Pascal (1670), Pensées #358

1. Altruism and Society

Understanding human behaviour has been the oldest concern and tradition of economic thinking, though its focus has shifted over the centuries and its explanations have been subject to the *fashionable methodology* of its time. Starting at the beginning of the last century and still being predominant in contemporary studies, neoclassical economics aimed to explain how human beings should behave, using the analytical fiction of a *Homo Economicus*[1]. Models based on this heuristic device perform well most of the time and are tractable too; yet observed phenomena sometimes contradict the conventional wisdom that human beings are intrinsically selfish, rational and exhibit stable preferences. In the second half of the last century, the new methodology of experimental economics created a range of evidence regarding choices made in the interest of other-regarding preferences.[2]

In an experimental study of the prisoners' dilemma game Kreps et al. [1982] show that a small probability of playing against an altruistic subject can trigger subjects to cooperate rather than

† PhD Economics

‡ PhD Economics

1 '*Homo Economicus*' is a term originally coined by John Stuart Mill to describe a narrowly self-interested individual. See Mill [1836, 1844].

2 For a comprehensive survey of altruism in experimental studies see Andreoni et al. [2007].

playing the dominant strategy of defection. It is challenging to directly test for non-selfish motivation, hence economists strived to eliminate all other possible sources, such as reputation, for the observed behaviour. Nonetheless, Andreoni and Miller [1993] show that 20% of the subjects need to be altruistic to support the equilibrium finding of cooperation in prisoners' dilemma games. Another line of research by Andreoni [1995] and others examines the role of altruism in the willingness of people to contribute to charities or public goods. Despite the fact that kindness and confusion play a significant role, experimental subjects choose to donate or contribute albeit they were aware of the opportunity to free-ride. Prominent in the experimental literature is the finding that proposers in dictator games give away on average 25% of their endowments (e.g. Forsythe et al. [1994]), while unfair offers in the extended ultimatum game are rejected (e.g. Güth [1982]).[3] Evidence from a field experiment by Uptorn [1973] suggests that money banish non-selfish behaviour, as the number of blood donators' drop when they are being paid.

The range of apparent contradictions leads to a renaissance in economic thinking, with scholars re-examining the premises of economic theory. At the heart of this development is the question why human beings behave like they do. Several theories have been developed to elucidate the observed phenomena within the realm of rational choice theory and utility maximisation. A starting point of this progression is the assumption that human beings have a genuine concern about the welfare of others, and, ceteris paribus, prefer outcomes whereby another person enjoys greater welfare. The "*willingness to act in the consideration of the interests of other persons, without the need of ulterior motives*"[4] is the core of human altruism.

Much of the ongoing debate on whether altruism or selfishness, or a mixture of the two, predominates as a driver of human behaviour has been the topic of economic discord since the early work of Adam Smith. The praise of both altruistic and selfish motives in Smith's work come to a contradiction. While in *The Wealth of Nations* Smith maintains that human beings are pursuing their own interest without any intention to promote that of the society, and therefore all benefits come as a result of '*unintended consequences*', in his later work on the *Theory of Moral Sentiments* he states: "*How selfish soever man be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness*

[3] Zizzo [2012] questions whether giving rates in Dictator games really reflect social preferences and are not confounded by experimenter demand effects.
[4] Nagel [1978, p. 79].

necessary to him, though he derives nothing from it, except the pleasure of seeing it." (Smith [1976, p. 9]).

Khalil [2001] claims that the theories which contemporary economists have developed over the last decades to explicate altruistic giving resemble the three theories of altruism which had already existed during Smith's time. A first explanation for altruism is that it ensures cooperative giving.[5] A second explanation is that altruism increases one's own utility if ones utility includes the utility of the potential recipient.[6] A third explanation maintains that altruism arises from a canon of morality.[7] So while individuals can act selfishly to take advantage of efficient mutual exchange in the tradition of Smith's *invisible hand*, there seem to exist several other reasons why human beings can act altruistically (including Smith's own interpretation, that Altruism is grounded on the natural tendency of people to feel sympathy).

The linkage between efficient selfish exchange and benevolent altruistic behaviour is well established in economic thinking.[8] According to Kolm [2006] the main advocates of altruism are those classical economic thinkers who contrive our understanding of the efficiency of selfish exchange (i.e. Smith, Mill, Edgeworth, Walras, Pareto). Kolm argues that a population of altruists could jeopardize the efficient functioning of a market (prices could not sent the correct signal to market participants if you would frequently pay more for an exchange than required, and in return receive an additional gift from your partner!), while on the other hand an exchange with a population of purely selfish agents could collapse (imagine your trading partner constantly tries to fool you!).

It seems reasonable to postulate that both altruism and selfishness coexists and both play their role in economic exchange systems. As Wicksteed [1933, p. 174] puts it: "*What makes it an economic transaction is that I am not considering you except as a link in the chain, or considering your desires except as the means by which I may gratify those of someone else-not necessarily myself. The economic relation does not exclude from my mind every one but me, it potentially*

[5] The idea that human beings act strategically and reciprocate has most famously been expressed by Sugden [1984] and Axelrod [1984].
[6] Becker's [1974] theory of social interactions understands altruism as an enhancement of others to increase one's own utility, in a utility-based setting where $U_i = U_i(U_{-i}, x_i)$.
[7] Etzioni [1986] among others argues that morals and norms dictate altruistic giving.
[8] The idea that altruism can enhance efficiency has been present in the economic literature since the seminal contributions by Barro [1974] and Becker [1974]. The original idea of cooperation is due to Axelrod and Hamilton [1981] and Axelrod [1984].

includes every one but you". Hence, altruism and selfishness do not contradict or exclude each other, but are rather voluntary choices in a state of smooth symbiosis. However, if we establish that both types of preferences exist, we need to ask why (sometimes case dependent) some people choose to act selfishly, some to act altruistically. In what follows, we explore in a framework of a parametric centipede game whether human beings are capable of being altruistic, that is, whether the choice to adopt altruistic preferences is evolutionary stable. Only if altruism does not lead to a reduced fitness, some human beings will be capable of acting altruistically.

2. Is altruism evolutionary stable?

The centipede game is a finite sequential move game with perfect information, wherein two players alternatively get a chance to take the larger proportion of a continuously growing pile of money or pass the pile to the other player. Passing decreases a player's payoff if the opponent player takes the larger proportion on the next move. If the opponent also passes, the players are presented with the same choice situation, however with increased payoffs.

Backward induction reasoning implies that the unique subgame perfect equilibrium of the game is for both players to take at every time they have to make a move in the game. Despite the very lucrative opportunity the game offers to players, the equilibrium prediction is that the player who makes the very first move takes the pile.

However, various experimental studies evidence systematic departures from the equilibrium prediction, inferring among others that subjects may have other-regarding preferences.[9] One of the suggested explanations for the pronounced discrepancy between the theoretical prediction and the experimental behavior upholds that the pool of subjects contains two types of players: altruistic and egoistic.

Based on the idea that a specific preference is more likely to be maintained and to flourish if it proves to be economically successful, we study, in the framework of the centipede game, a dynamic process in which preferences for those types can change over time, building on the indirect evolutionary approach developed by Güth and Yaari [1992] The key question is whether individuals of the altruistic type are evolutionarily successful in the sense that their population share increases over time. To provide continuous dynamics of the evolution process we look at

[9] See for instance, McKelvey and Palfrey [1992], Fey et al. [1996] and Nagel and Tang [1998].

an explicit model of evolutionary dynamics, the replicator dynamics, which is due to Taylor and Jonker [1978].

2.1. Description of the game

We consider a parametric class of centipede games with $K+1$ nodes, numbered by k, where $k \in \Lambda = [0, K]$. The player deciding at the last node (node K) is labeled L. The other player is labeled O. At each node players can either *pass* or *take*. As soon as one player take the game is over. If a player *pass*, then it is the other player's turn to decide. The payoff structure is as follows. *Take* at node $k=0$ gives (0,0) payoffs. From then on, any pass adds to both players' *accounts* a payoff of 2. The player who eventually *take* gets whatever is in his account plus 1; the other player gets whatever is in his account minus 3. If player L *passes* at the last node, both players get exactly what has been accumulated in their account so far, i.e. $2K$. Let $\pi_i: \Lambda \to R$ be the material payoff function of each player $i, i = \{L, O\}$. We consider two types of players, egoists, denoted by E, who maximize their own payoff and altruistic types, denoted by A, who maximize the aggregate payoffs (i.e. the payoffs of both players). Let $m_i \in \{A, E\}$ denote player i's preference type. The (subjective) utility function of an altruistic player is $u_{\{A\}} = \pi_i + \pi_j,\ i \neq j$ and for an egoistic player is $u_E = \pi_i,\ i = \{L, O\}$. Given that players are randomly matched the expected (subjective) utility of each player is $V(m_i, m_j) = \left(\frac{1}{2}\right)\pi_L(m_i, m_j) + \left(\frac{1}{2}\right)\pi_O(m_i, m_j)$, where $\pi_L(m_i, m_j)$ is the material payoff of a type m_i player when a player L meets a player of type m_j and $\pi_O(m_i, m_j)$ is the material payoff to a player of type m_i when a player O meets a player of type m_j.

To begin with, if all players are egoists (subjective and material payoffs are identical), the unique subgame perfect Nash equilibrium (henceforth "SPE") prescribes for both players to *take* at every node they have to make a move. Therefore, the equilibrium predicts that the game ends at $k=1$. At the SPE both players receive a (material) payoff of 0 (see Figure 1).

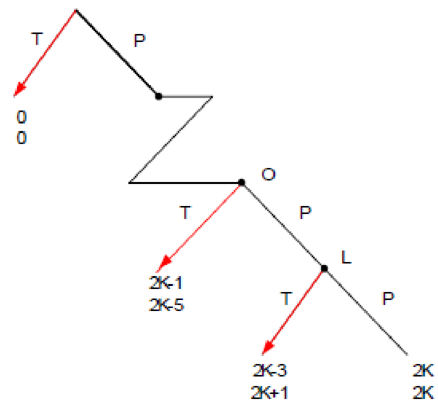

Figure 1: two egoists.

Now suppose that two altruists meet. In this case, given that decisions are based on subjective utility maximization players *pass* every time they have to make a move in the game. Thus, at the SPE each player receives a material payoff of $2K$. Remarkably, this outcome is (first best) efficient as both players obtain the maximum feasible material payoff (see Figure 2).

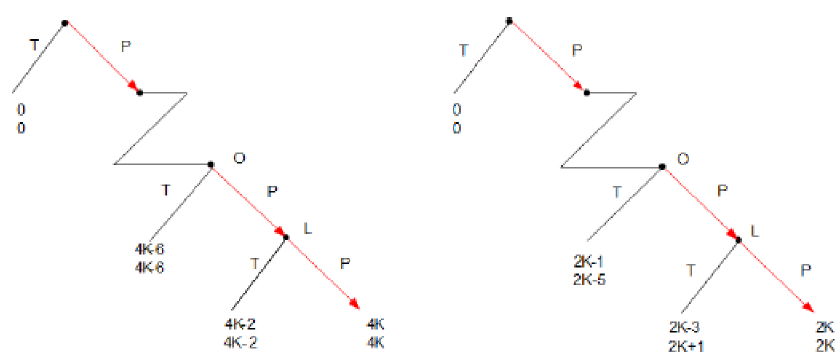

Figure 2: Both players are altruists. The tree to the left (right) depicts the subjective (material) payoffs.

Consider now the more interesting scenario where the game is played by an altruist and an egoist. If player L is altruist, and thus *pass* at all nodes, player O also *pass* at all nodes. Therefore, at the unique SPE each player receives a material payoff equal to $2K$. The outcome is (first best) efficient (see Figure 3).

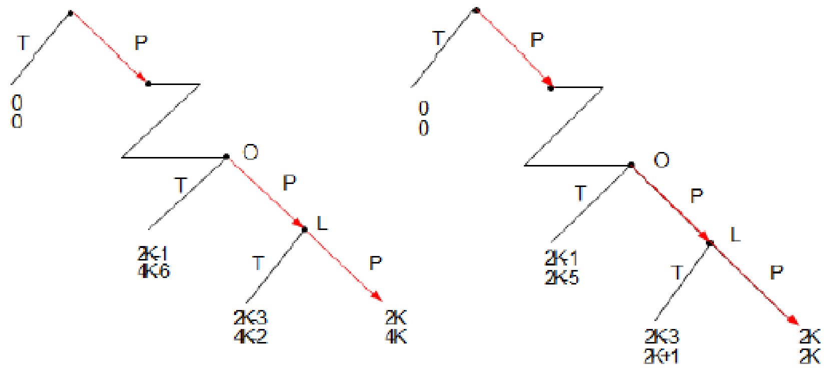

Figure 3: L is altruist and O is egoist. The left (right) graph depicts the subjective (material) payoffs.

The outcome is second best efficient if player L is an egoist. Given that player O always *pass* player L mimics the behaviour of player O at all nodes but the terminal one where he *takes*. At the SPE player L receives a material payoff equal to $2K + 1$, while player O receives $2K - 3$ (see Figure 4).

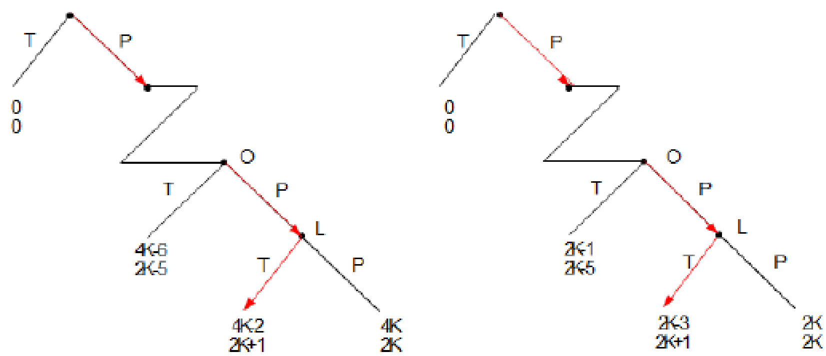

Figure 4: L is egoist and O is altruist. The left (right) graph depicts the subjective (material) payoffs.

The material payoff (fitness) matrix for the emerging evolutionary game where preferences are endogenous and compete with each other for survival is:

		σ_A	$1-\sigma_A$
L/O		Altruist	Egoist
σ_A	Altruist	$2K, 2K$	$2K-\frac{3}{2}, 2K+\frac{1}{2}$
$1-\sigma_A$	Egoist	$2K+\frac{1}{2}, 2K-\frac{3}{2}$	0,0

The evolutionary game has three Nash equilibria: two are in (asymmetric) pure strategies and one is in mixed strategies: i) $(\sigma_A = 1, \sigma_A = 0)$, ii) $(\sigma_A = 0, \sigma_A = 1)$ and iii) $(\sigma_A = \frac{4K-3}{4K-2}, 1-\sigma_A = \frac{1}{4K-2})$. It is straightforward to show that the symmetric Nash equilibrium in mixed strategies where players randomize between their pure strategies (here preferences) is the unique evolutionary stable equilibrium[10]. This strategy is robust to evolutionary pressure and can resist the onslaught of mutants with different preferences.

The evolutionary stable population distribution is mixed and in such proportion that just balances the advantages and disadvantages of being of either type. The evolutionary stable state is determined by the following trade off. On the one hand, an altruist reduces his success by choosing an action that reflects some concern for the other player's success. On the other hand, his preference type restricts him to pass at every node he has to make a move, which triggers a favorable (strategic) reaction by the other player.

To explore the properties of the evolutionary process we look at the replicator dynamics, which for our game is governed by $\frac{\dot{x}}{x} = (\frac{1}{2})(x-1)[2x(2K-1)-4K+3]$. In words, the population share of individuals programmed to a certain pure strategy (altruists or egoists) grows in the replicator dynamics if and only if the strategy earns a material payoff above the average material payoff of the current population state, which is represented by the mixed strategy. Note that only the material performance matters for the growth of players of a given

[10] The concept of evolutionary stable strategies was initially introduced by Smith and Price [1973].

preference. However, preferences matter indirectly, since they (together with the population's preference configuration) determine players' material performance.

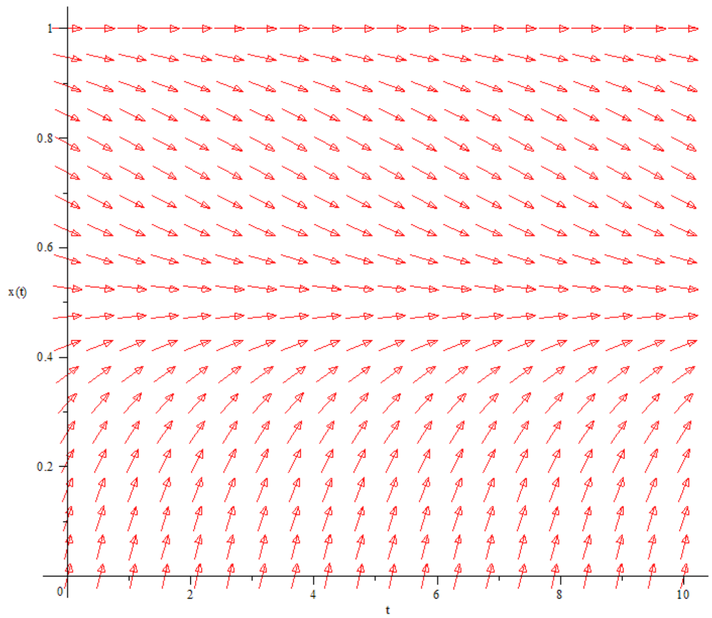

Figure 5: Evolution of altruism under the continuous replicator dynamics for $K = 1$ (the trust game). At the steady state 50% of the population are altruists.

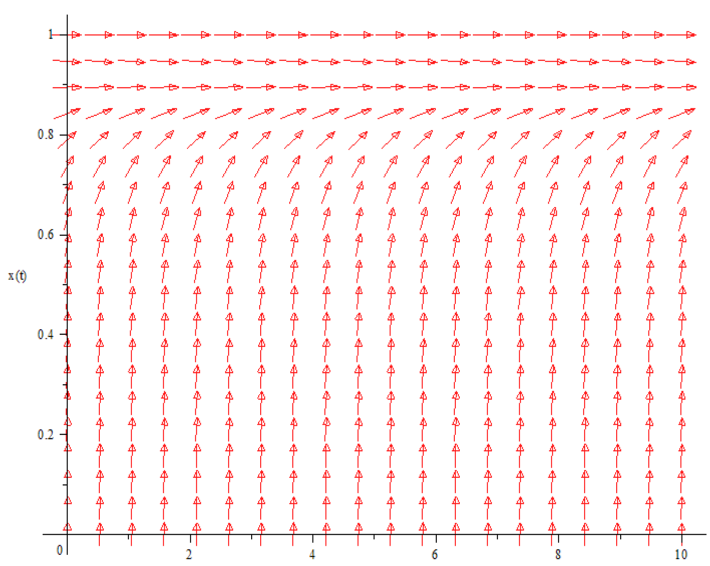

Figure 6: Evolution of altruism under the continuous replicator dynamics for $K = 3$. At the steady state 90% of the population are altruists.

It is straightforward to see that the exact composition of preferences in the population that emerges crucially depends on the length of the game as represented by parameter K. In particular, as K increases the frequency of altruists in the population increases, however at a diminishing rate $\frac{\partial x}{\partial k} = \frac{1}{(2K-1)^2} > 0$ and $\frac{\partial^2 x}{\partial k^2} = \frac{-2}{(2K-1)^3} < 0$. Intuitively, the longer the length of the game, the greater the efficiency gains ensuing to the population from cooperative behavior. In the limit where K tends to infinity x tends to 1, i.e. unmixed population consisting only of altruists.

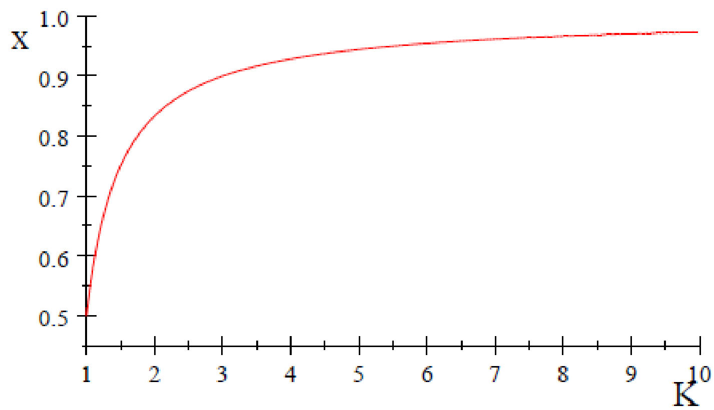

Figure 7: Frequency of altruists in the stable equilibrium as a function of K.

3. Conclusions

A glimpse to the inquiries of the classical economic thinkers unequivocally reveals that the idea of apprehending altruism and selfishness as being complementary, rather than mutually exclusive, has been a long-standing convention in economics. With the promise of behavioural economics, scholars have recently reviewed the premises of rational selfish behaviour. This development has enriched our understanding of various phenomena previously thought to contradict the conventional wisdom of the narrow minded Homo Economicus.

In this essay we explore, in the framework of a centipede game, whether altruistic types can survive over the evolutionary process. By committing to altruistic preferences an individual constraints his behaviour, thus altruism has a commitment value for the society. By linking opponent's success with his own success the altruistically motivated individual discourages his egoist opponent from actions that would lower his material payoff because egoists' economic success would also decline. These beneficial indirect effects on the behavior of others may dominate the direct disadvantages of being altruistic. Without departing from the realm of rational choice theory our analysis concludes that individuals are indeed capable of being altruists. Altruism is perceived as a mechanism to improve the welfare of a population/society.

References

Andreoni, J. (1995), Cooperation in public–goods experiments: kindness or confusion, *The American Economic Review*, 85(4): 891-904.

Andreoni, J., Harbaugh, W., and Vesterlund, L. (2007). Altruism in experiments, *The New Palgrave Dictionary of Economics,* 2.

Andreoni, J. and Miller, J. (1993), Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence, *The Economic Journal*, 103(418): 570-585.

Axelrod, R. and Hamilton, W. (1981). The evolution of cooperation, *Science,* 211(4489): 1390-1396.

Barro, R. (1974). Are government bonds net wealth? *Journal of Political Economy,* 82(6): 1095-1117.

Becker, G. (1974). A theory of social interactions. *Journal of Political Economy,* 82(6): 1063-93.

Etzioni, A. (1986). The case for a multiple-utility conception. *Economies and Philosophy,* 2: 159-183.

Fey, M., McKelvey, R., and Palfrey, T. (1996). An experimental study of constant-sum centipede games. *International Journal of Game Theory,* 25(3): 269-87.

Forsythe, R., Horowitz, J., Savin, N., and Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic behaviour,* 6(3): 347-369.

Güth, W., Schmittberger, R., and Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic behaviour and Organization,* 3(4): 367-388.

Khalil, E. (2001). Adam Smith and three theories of altruism. *Researches economiques de Louvain,* 67(4): 421-435.

Kolm, S. (2006). Introduction to the economics of giving, altruism and reciprocity. *Handbook in the Economics of Giving, Reciprocity and Altruism,* 1:1-122.

Kreps, D., Milgrom, P., and Roberts, J. (1982). Rational cooperation in the finitely repeated prisoner's dilemma. *Journal of Economic Theory,* 27(2): 245-252.

McKelvey, R. and Palfrey, T. (1992). An experimental study of the centipede game, *Econometrica,* 60(4): 803-836.

Mill, J. (1836). On the definition of political economy, and on the method of investigation proper to it. *London and Westminster Review.*

Mill, J. (1844). Essays on some unsettled questions of political economy. *London: Longmans, Green, Reader Dyer.*

Nagel, R. and Tang, F.-F. (1998). Experimental results on the centipede game in normal form: an investigation on learning. *Journal of Mathematical Phycology,* 42: 356:384.

Nagel, T. (1978). The possibility of altruism. *Princeton University Press.*

Smith, A. (1976). The theory of moral sentiments. *Oxford, England: Clarendon Press,* (original work published in 1790), sixth edition.

Smith, M. and Price, G. (1973). The logic of animal conflict. *Nature,* 246: 15-18.

Sugden, R. (1984), Reciprocity: the supply of public goods through voluntary contributions. *The Economic Journal,* 94(376): 772-787.

Taylor, P. and Jonker, L. (1978). Evolutionary stable strategies and game dynamics. *Mathematical Biosciences,* 40: 145-156.

Uptorn, W. (1973). Altruism, attribution, and intrinsic motivation in the recruitment of blood donors. *PhD thesis,* Cornell University.

Wicksteed, P. (1933). The common sense of political economy. London: Routledge and Kegan Paul Limited, (first published in 1910).

Zizzo, D. (2012). Do dictator games measure altruism? *Handbook on the Economics of Philanthropy, Reciprocity and Social Enterprise,* Edward Elgar (forthcoming).