Building trust by wasting time

Carl Bergstrom
Department of Biology
University of Washington
Seattle WA

Ben Kerr
Department of Biology
University of Washington
Seattle, WA

Michael Lachmann
Max Planck Inst. for Evol. Anthropology
Leipzig, Germany

October 2, 2006

> Time waste differs from material waste in that there can be no salvage.
>
> Henry Ford[1]

1 Introduction

Much of neoclassical economics is founded upon models in which wholly self-interested agents interact — but an overwhelming preponderance of empirical evidence suggests that human individuals are not purely self-interested.

[1]From *My Life and Work* (with Samuel Crowther, 1922) quoted in Anon, ed. (2003) *The Big Book of Business Quotations*, New York, Basic Books.

Rather, they express other-regarding preferences such as fairness, trustworthiness, or generosity that we collectively label *values*. While these values surely facilitate social and economic exchange and with them the concomitant gains to communication and trade (Zak this volume, Kimbrough et al. this volume: [1, 2]) , we should not leap too hastily from this observation to the conclusion these values are part of our genetic nature — and that without them there, modern commerce would be impossible.

Rather than try to justify the prevalence of other-regarding behavior in economic exchange by means of a model in which economic structure is jury-rigged on top of a system of values that were perhaps adaptive in small-group interpersonal interactions, we invert the causal chain. We argue that markets need not *rest upon* values that arose before them; instead, markets may *create* the values which allow them to function effectively. Markets are human constructions; in creating them, the participants engage in a process of mechanism design, selecting the rules of the strategic games in which they will be involved. These rule choices give rise to conventions of behavior — and where such conventions are granted normative force, they may appear to us as values.

To illustrate this process, we provide an example in which other-regarding values of fidelity and loyalty emerge within a system of primarily self-regarding individuals, as mechanisms to alter the structure of the games so as to facilitate and stabilize beneficial outcomes. These are important values in our society; when surveyed about attributes that are highly valued in a potential mate, men and women rank the values of loyalty and fidelty as among the most important [3, 4]. By way of illustrating how these values can arise from conventions of behavior, we will address an unlikely question: why do we build trust by wasting time? We look at how conventions of lengthy courtship and values of trustworthiness or faithfulness might arise as optimal strategies for purely self-interested agents.

2 Wasting time

Animals court prospective mates for days or weeks, even when the breeding season is precariously narrow. Adolescents "hang out" with their friends, killing time in shopping malls and on street corners when they could be earning money or mingling with the opposite sex or developing their prowess in any number of academic or athletic pursuits. Firms invest heavily in lengthy and expensive processes of contract negotiation before initiating cooperative relations, even when these contracts may be largely insufficient to provide adequate legal recompense in the event of a unilateral defection.

Economists [5, 6] and biologists [7, 8, 9] commonly interpret these time costs either as as attempts acquire information by assessing the quality of potential partners, or as efforts to transmit information via costly signalling, with signal costs paid in a currency of time. While these may be important components of time-wasting behavior, here we propose an alternative hypothesis that may operate instead of in addition: we argue that wasting time functions to change the incentive structure in such a way that long-term cooperative behavior becomes strategically stable. This allows us to bring waste to bear upon another problem that faces those working in biological, economic, and legal fields alike: that of maintaining cooperation in the face of individual incentives for defection. While the stand-by explanations of kin selection or reciprocity may be adequate to explain some cases of cooperation, the conditions for these mechanisms seem too restrictive to explain the majority of cooperative interactions [10]. Recent authors have started to replace kin-selection and reciprocity with explicit market analogies, affording a key role to partner choice, reputation, and mechanism design [11, 12]. The present proposal rests very much within that line of explanation.

3 A simple model of courtship

We develop a simple evolutionary game as model of courtship[2] . Because we are interested in the cost of wasting time, we will keep track both of payoffs Π from the game and the amount of time T taken to play the game, and we will look at the metagame of maximizing E[payoff]/E[elapsed time] as in forgaging theory [14, 15] . This is equivalent to maximizing the weighted average E^* of payoff rates E^*[payoff/elapsed time] where the rates are weighted by the amount of time spent at earning each rate.

The basic interaction

Two individuals engage in a basic trust game akin to that treated by Ostrom and Schwab in this volume [16]. The intereaction begins with an opt-in / opt-out decision by the first player whom we call the recipient. If the recipient opts out (declines), the game is over; if the recipient opts in (accepts), a simple dictator game follows with the second player, who we call the actor, in the role of dictator. The actor chooses whether to cooperate (C), yielding a payoff of 2 to each party, or exploit (E), yielding a payoff of 3 to himself

[2] T. Bergstrom and Ponti independently developed a closely related model in an unpublishing 1998 working paper [13]; the remarkable similarities between the present analysis and that previous one are enough to make one believe in genetics, vertical cultural inheritance, or both.

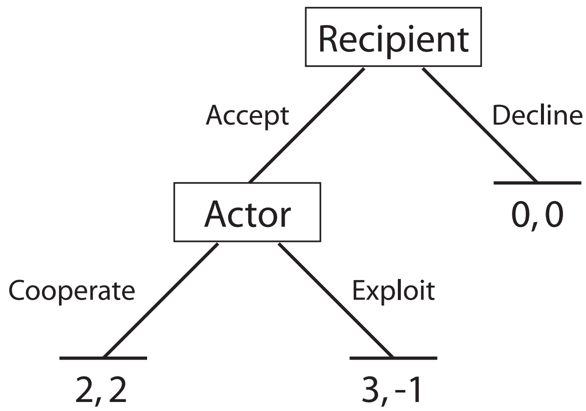

Figure 1: Trust game in extended form

and a payoff of -1 to the recipient. Either way, the game takes 1 time unit. Figure 1 shows this game in extended form.

We treat this as a one-shot game with anonymous partners drawn from a large population. Individuals do not recognize other individuals, nor do they play again with the same partner once the partnership is terminated. This game has a single subgame-perfect equilibrium (Decline, Exploit) and under replicator dynamics, a population would spend almost all its time on a neutral component along which the recipient plays Decline always and the actor plays Exploit with at least 1/3 probability (see e.g. [17].)

Iterated play and payoff rates

Now we extend the game to consider the possibility of repeated play between partners. As illustrated in Figure 2, each round of play takes some duration of time t_b for the receiver to make a decision and an additional duration t_c for the actor to make a decision if granted the opportunity. If the recipient declines the partnership, each individual is randomly assigned a new partner

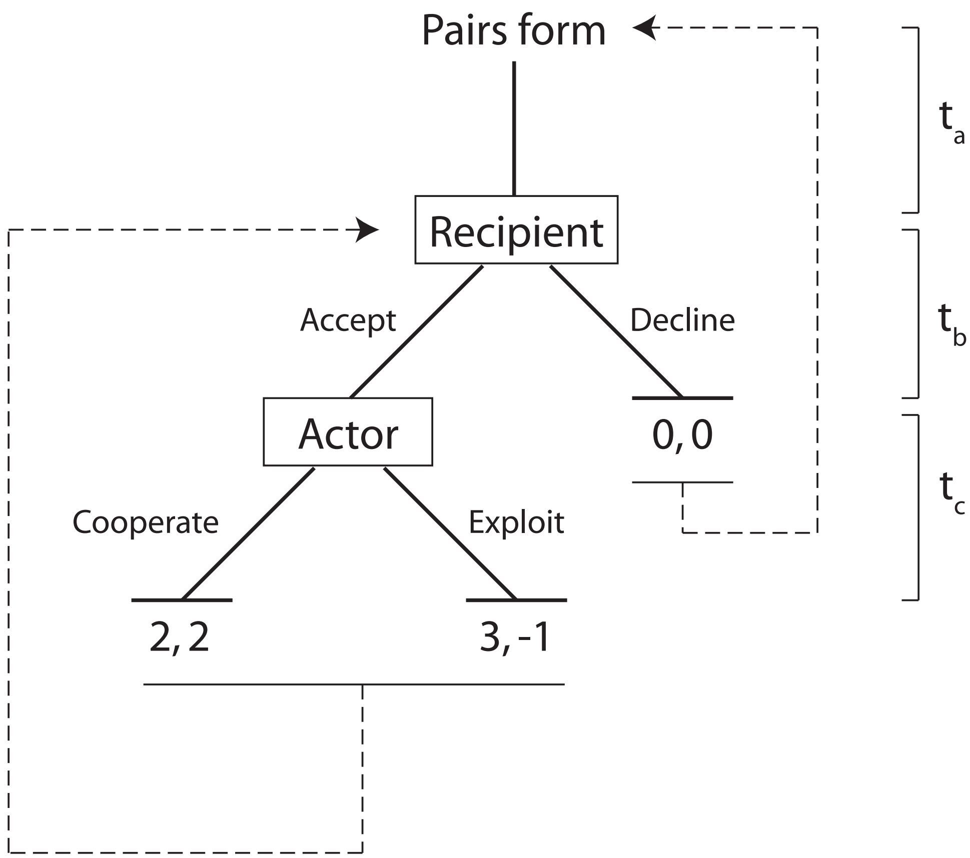

Figure 2: Repeated version of the choice game with time.

and role, and this pair formation process takes some duration of time t_a. To keep the algebra maximally simple, we assume that play can continue indefinitely without temporal discounting or accidental disruption of the partnership. While some of the details would change if we relaxed this assumption, the basic conclusions would continue to hold.

To further simplify the algebra, we will proscribe all strategies that require the ability to count or more than one round of memory on the part of the players. The basic points that we are making here will survive the extension to a more complex set of memory-based strategies, but that extension comes at a heavy algebraic cost.

Actors cannot count rounds, and they only get to move if receivers opt in. Thus the actors' only strategy choice is to play a possibly mixed strategy with probability $1 - p$ of cooperating and p of exploiting in each round. Recipients can recall only the most recent round and we restrict them to pure strategies; thus their strategy choices are limited to DA, decline always; DE, decline if and only if exploited; or DN; decline never[3]. Defining the *time cost of re-pairing* as $t_1 = t_a + t_b$ and the *time cost of playing* as $t_2 = t_b + t_c$, the payoff rates are then as follows (see Appendix 1):

$$\begin{aligned}
\Pi(\mathrm{DA}, p) &= (0, 0) \\
\Pi(\mathrm{DE}, p) &= \left(\frac{2 - 3p}{p\,t_1 + t_2}, \frac{2 + p}{p\,t_1 + t_2}\right) \\
\Pi(\mathrm{DN}, p) &= \left(\frac{2 - 3p}{t_2}, \frac{2 + p}{t_2}\right)
\end{aligned} \tag{1}$$

Now we observe a few things. First, if the receiver plays DN, then the actor's payoff increases monotonically with p, and therefore the actors should always exploit, yielding a negative payoff to the receiver. Thus the receiver does better with DA than with DN. So we will not expect to see receivers playing DN at equilibrium.

Next, what is the optimal strategy for the actor if the receiver plays DE? We differentiate the actor's payoff from $\Pi(\mathrm{DE}, p)$ with respect to actor strategy p to get $(t_2 - 2t_1)/(p\,t_1 + t_2)^2$. When $t_1 < t_2/2$, this is strictly positive for p on $[0, 1]$ and thus we get an edge solution $p = 1$. When

[3] A fourth possibility, Decline if and only if not exploited, breaks up favorable partnerships while retaining unfavorable ones, and will not be considered here.

$t_1 > t_2/2$, this is strictly negative for p on $[0, 1]$ and thus we get an edge solution $p = 0$.

As a result, when t_1 is small, actors' best response to DE is to exploit always and receivers therefore do better playing DA than DE. When t_1 is large, actors' best response to DE is to exploit never, and there DE outperforms DA.

Therefore we expect the following phenotypic equilibria of the replicator dynamics: When $t_1 < t_2/2$, we cannot maintain cooperation and instead receivers will decline to interact at all. When $t_1 > t_2/2$, receivers will accept interactions and actors will cooperate. Thus when t_1 is sufficiently large, there is no problem maintaining cooperation. But what can the participants do if t_1 is too small?

Incorporating courtship

To answer this, assume $t_1 < t_2/2$. We extend the game slightly, adding a convention in which actors must first "court" recipients for a period of duration t_d before the recipient decides whether to accept or decline the interaction. This courtship period offers no direct payoff return to either player; it is simply "wasted time". This game is shown in figure 3.

Now we take what we learned above. By adding a courtship period, we have effectively extended the time cost of re-pairing from $t_1 = t_a + t_b$ to $t_{1*} = t_a + t_d + t_b$. When the time cost of re-pairing t_{1*} is less than $t_2/2$, actors will do best to exploit always and in this setting recipients can do no better than to decline always (DA) for payoff rates $\Pi = (0, 0)$. When the time cost of re-pairing exceeds $t_2/2$, actors will do best to cooperate always. With a courtship time t_d just beyond that needed to induce the

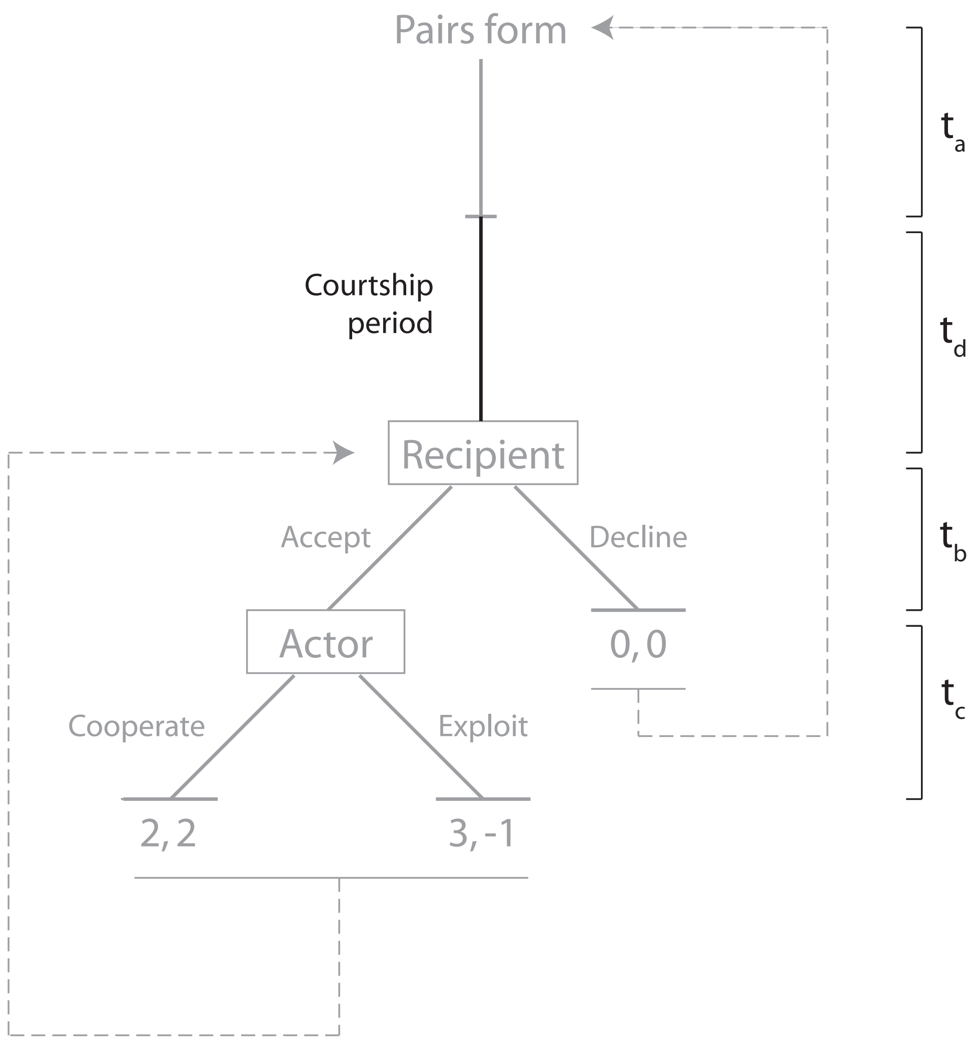

Figure 3: Extended version of the choice game with time. The Actor now courts the recipient for some time t_d before the participants play the game treated previously (shown here in grey).

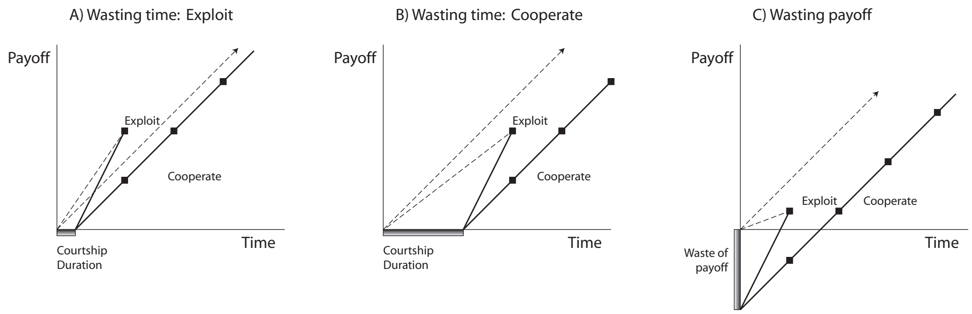

Figure 4: Stabilizing cooperation. The actor does best to choose the strategy which yields the highest payoff rate over time. In each figure payoff rate for exploiting and then finding a new partner is shown by the dashed line from the origin to the point labelled "exploit." The payoff rate for cooperating repeatedly is shown by the dashed line through the origin, drawn parallel to the slope of the cooperate trajectory. Panel A) When courtship times are short, exploit offers the higher payoff rate to the actor. Panel B) When courtship times are long, cooperate offers the highest payoff rate to the actor. Panel C) Wasting payoff can stabilize cooperation as well. Given a payoff waste as shown, cooperation offers a higher payoff rate then exploitation.

actor to cooperate, i.e., $t_d = t_2/2 - t_1 + \epsilon = (t_c - t_b)/2 - t_a + \epsilon$, we will have an equilibrium in this game in the players court for time t_d, receivers agree to play, and actors cooperate always. This yields asymptotic payoffs $\Pi = (2/t_2, 2/t_2)$. By instituting the additional courtship stage of the game, which is purely wasteful in any direct sense given that it uses time and confers no direct payoff, we have stabilized the socially efficient outcome which otherwise would have been unobtainable. Thus the need to allow effective cooperative exchange has lead to the development of a convention — wasteful courtship and subsequent cooperation — for the game.

Figure 4, panels A and B provide a graphical illustration of the courtship duration necessary to stabilize cooperation by the actor, given that the receiver is playing DE. The dashed line indicates the average payoff rate should the actor exploit always ($p = 1$); the solid line indicates the average payoff rate should the actor cooperate always ($p = 0$). When the courtship period is short, exploiting and then finding a new partner provides a higher payoff rate. When the courtship period is long, repeated cooperation offers a higher payoff rate. This explanation closely parallels Charnov's marginal value theorem used to compute the optimal duration of patch use [18].

4 Other kinds of waste

In the present model, we are concerned with payoff rate defined as payoff over time; we have considered models in which the participants initially reduce payoff rate by increasing the denominator, wasting time at the beginning of their interaction. Alternatively, individuals could reduce payoff rates by decreasing the numerator, and this wil have a similar strategic influence on the game. For example, if payoff is accrued in monetary units, the players could waste money instead of time. Panel C of Figure 4 illustrates this graphically.

Economists have treated this latter case in considerable detail. Carmichael and MacLeod [19] develop a mathematical model of gift exchange, in which the institution of gift exchange stabilizes cooperation among agents playing a prisoner's dilemma. Their logic is closely analogous to that which we have presented here: "Under this convention, an exchange of gifts at the beginning of a new match will break down mistrust and allow cooperation to start immediately. The reason is simple. A parasite in a gift-giving society will have to buy a succession of gifts, while incumbent honest types need buy the gift only once. The gift giving custom lowers the relative return to being a parasite." (p.486)

Bolle [20] develops model of gift-giving even more closely related to that in the present paper. Bolle extends the prisoner's dilemma approach of Carmichael and MacLeod to the trust game treated here; this allows Bolle to deal with asymmetric games (such as pairing between males and females) in which the two players have different preferences. Bolle also notes that the time costs of courtship can serve a gift-exchange function: "Evaluated by their respective wages, the time spent together [by a courting pair] is a large scale exchange of gifts. Both may enjoy this time — but from an 'objective' point of view there is hardly any utility from this extensive being together." (p.1). The major differences between Bolle's model and ours result from our focus on time-wasting: to handle this in depth, we separate time from payoff, treating time explicitly within the model and maximizing the *payoff rate* rather than the expected payoff.

The theory of non-salvagable assets [21, 22] puts forth a similar explanation of expenditures on advertising, custom furnishings, lavish showrooms, and other highly firm-specific investments that neither directly generate revenue nor are recoverable should the firm be dissolved. By this explanation, advertising is effective precisely because it is expensive. Advertisements convey no direct information about the quality of the product. Rather, through the expense of advertisement a firm signals that it has sunk appreciable cost into its reputation, and thus indirectly signals its intention to engage in long-term cooperative relationships rather than short-term exploitative ones. While we treat investment in time here, we notice that time is, after all, an utterly non-salvagable asset.

Time differs from many other non-salvagable assets in that — unlike advertisements or custom furnishings, the costs of wasted time accrue directly to both parties. Similarly, time investment in courtship differs from unidirectional gift-giving in its reciprocal nature. The *reciprocal exposure* [23] inherent in wasting time thereby both protects the recipient from exploitation by the actor, and protects the actor from having a recipient opt out subsequent to the courtship period. Models suggest that gift-giving (whether as a signal or as a way of changing the game payoffs) may work most efficiently with gifts that are costly to produce but have low value to the recipient, because low-value gifts avoid the risk to the actor of a recipient who simply collects gifts and then opts out [20, 9]. As Spence [5] and later Bolle [20] point out, time wasted (alone or better yet, together) serves beautifully in this respect.

5 Discussion

In the model presented above, we have shown how a group of purely self-interested agents can establish a courtship convention and standards of subsequent behavior that resemble values such as generosity and trustworthiness, particularly should they come to be imbued with normative force. Here potential gains through exchange or trade, as modeled by a trust game, are realized not because the players come to the game with generous tendencies or placing value on faithfulness. Rather, this opportunity for gain through trade provides all participants with an incentive to expand the game itself, as in Figure 3, so as to generate within the new game the incentives for non-exploitative behavior and extended partnership fidelity.

We believe that a vital step in the evolution of smoothly functioning social and financial institutions is the process by which the players themselves choose or modify the rules of the game (e.g. Goodenough, this volume: [24]). By turning simple games, often with non-cooperative solutions, into mechanism design problems, agents can expand the possibilities for stable prosocial interaction, thereby facilitating exchange and generating values.

Legal structure surrounding contract law serves a similar function; O'Hara (this volume: [25]) describes a number of ways by which "the law's commitment to enforce contracts can, at the margin, provide added assurances that the risk to dealing with strangers in minimized." Interestingly, it be inefficient for the legal system to offer too complete of protection. As O'Hara points out, full legal compensation can generate a moral hazard problem in which interacting parties lack any incentive to gather relevant information about the past performance and likely future intent of their partners.

Our model here bears a close relationship to models of punishment. First of all, the duration of of courtship necessary to stabilize cooperation (or the payoff wasted, if we are wasting payoff currency) is exactly equal to the amount that would have to be extracted as a punishment for exploitation, in order to deter exploitation by payoff-maximizing players. Once the initial courtship becomes an established custom, we can think of the initial courtship period as a sort of advance punishment in the sense that if all recipients require courtship and all recipients break off a partnership after exploitation, the cost of exploitation will be the immediate need to go through a courtship period. In this way, the recipient's decision to break of the repeated interaction imposes a *de facto* cost equal to the duration of courtship, and this cost serves as a sufficient punishment to deter exploitation.

One the major problems with most models of punishment is that if there is any cost to the act of punishing, then punishing itself becomes a sort of altruistic act. A wronged individual typically does better to cut his or her losses and move on than to invest in costly post-hoc punishment. When the timing of the punishment shifts from after the fact (and conditioned on misbehavior) to before the interaction (and necessary to initiate any interaction), the individual incentives not to punish disappear. Indeed, every individual has to go through the full courtship period so as to avoid providing the actor with an immediate incentive to exploit. Thus the act of imposing the punishment costs necessary for cooperation — which are the same whether they occur at the start of each partnership or after each transgression — shifts from an altruistic act to a self-interested one. This mechanism offers another solution to the second-order public goods problem associated with altruistic punishment [26].

Acknowledgements

The authors thank Mark Grady for bringing our attention to this problem, Ted Bergstrom for sharing his unpublished manuscript on this problem, Oliver Goodenough for the epigraph, and the many participants of the Gruter Institute program on Free Enterprise: Values in Action who provided us with numerous provocative discussions and helpful suggestions.

Appendix 1

To compute the payoff $\Pi(\mathrm{DE}, p)$, we use first-step analysis [27], with the census point at the recipient's decision point immediately after pair formation or retention; notice that a recipient who players DE always accepts upon reaching the census point. Let T_n be the expected time elapsed after n plays of the subgame in which the actor actually reaches his decision point (i.e., after n plays in which the recipient accepts either because the partnership is a new one, or because the recipient has cooperated on the previous round). If the actor cooperates, then the players return to the census point after one decision interval t_b by the recipient and one t_c by the actor for a total of $t_b + t_c$. If the actor exploits, $t_b + t_c$ elapses as before, but now an additional t_b elapses during which the recipient chooses to decline and an additional t_a elapses during the formation of new pairs. This gives a total elapsed time of $t_a + 2t_b + t_c$. We can express T_n recursively:

$$\begin{aligned} T_n &= p(T_{n-1} + t_a + 2t_b + t_c) + (1-p)(T_{n-1} + t_b + t_c) \\ &= T_{n-1} + p(t_a + t_b) + t_b + t_c \end{aligned} \tag{2}$$

Similarly, the expected payoffs P_n and Q_n to the Actor and Recipient respectively are

$$\begin{aligned} P_n &= 2 + p + P_{n-1} \\ Q_n &= 2 - 3p + Q_{n-1} \end{aligned} \tag{3}$$

Our initial census point occurs after the first pairs are formed, before any points have been scored, so $T_0 = t_a$, $P_0 = 0$, and $Q_0 = 0$. Thus we can write the recursions (3) in explicit form:

$$\begin{aligned} T_n &= n(p\,(t_a + t_b) + t_b + t_c) + t_a \\ P_n &= n(2 + p) \\ Q_n &= n(2 - 3p) \end{aligned} \tag{4}$$

The expected payoff rates for actor and recipient are then $\lim_{n\to\infty} P_n/T_n = (2+p)/(p\,(t_a+t_b)+t_b+t_c)$ and $\lim_{n\to\infty} Q_n/T_n = (2-3p)/(p\,(t_a+t_b)+t_b+t_c)$ respectively.

References

[1] P. Zak. Value based economies. In *Free Enterprise: Values in Action (tenetive)*. TBA, 2007.

[2] E. Kimbrough, V. L. Smith, and B. J. Wilson. Building a market: From personal to impersonal exchange. In *Free Enterprise: Values in Action (tenetive)*. TBA, 2007.

[3] D. M. Buss and M. Barnes. Preferences in human mate selection. *Journal of Personality and Social Psychology*, 50:559–570, 1986.

[4] J. A. Simpson and S. W. Gangestad. Sociosexuality and romantic partner choice. *Journal of Personality*, 60:31–51, 1992.

[5] A. M. Spence. Time and communication in economic and social interaction. *Quarterly Journal of Economics*, pages 651–660, 1973.

[6] A. Farmer and A. W. Horowitz. The engagement game. *Journal of Population Economics*, 17:627–644, 2004.

[7] R. Dawkins. *The Selfish Gene*. Oxford University Press, 1976.

[8] C.-A. Wachmeister and M. Enquist. The evolution of female coyness: Trading time for information. *Ethology*, 105:983–992, 1999.

[9] P. D. Sozou and R. M. Seymour. Costly but worthless gifts facilitate courtship. *Proceedings of the Royal Society of London Series B*, 272:1877–1884, 2005.

[10] P. Hammerstein. Understanding cooperation: An interdisciplinary challenge. In S. Bowles and P. Hammerstein, editors, *Dahlem Conference Report: Genetic and Cultural Evolution of Cooperation*. MIT Press, Cambridge, MA, 2003.

[11] R. Noë and P. Hammerstein. Biological markets. *Trends in ecology and evolution*, 10:336–339, 1995.

[12] P. Hammerstein, editor. *Genetic and Cultural Evolution of Cooperation*. MIT Press, Cambridge, MA, 2003.

[13] T. C. Bergstrom and G. Ponti. Long-term partnership and desertion. Working paper, November 1998.

[14] D. W. Stephens and J. R. Krebs. *Foraging Theory*. Princeton University Press, Princeton, 1986.

[15] D. W. Stephens, B. Kerr, and E. Fernández-Juricic. Impulsiveness without discounting: the ecological rationality hypothesis. *Proceedings of the Royal Society of London Series B*, 271:2459–2465, 2004.

[16] D. Schwab and E. Ostrom. The vital role of norms and rules in maintaining open public and private economies. In *Free Enterprise: Values in Action (tenetive)*. TBA, 2007.

[17] L. Samuelson. *Evolutionary Games and Equilibrium Selection*. MIT Press, Cambridge, Massachusetts, 1997.

[18] E. L. Charnov. Optimal foraging; the marginal value theorem. *Theoretical Population Biology*, 9:129–136, 1976.

[19] H. L. Carmichael and W. B. MacLeod. Gift giving and the evolution of cooperation. *International Economic Review*, 38:485–509, 1997.

[20] F. Bolle. When to buy your darling flowers: on cooperation and exploitation. *Theory and Decision*, 50:1–28, 2001.

[21] P. Nelson. Advertising as information. *Journal of Political Economy*, pages 729–754, 1974.

[22] B. Klein and K. B. Leffler. The role of market forces in assuring contractual performance. *The Journal of Political Economy*, 89:615–641, 1981.

[23] O. E. Williamson. Credible commitments: Using hostages to support exchange. *American Economic Review*, 73:519–540, 1983.

[24] O. Goodenough. Tba. In *Free Enterprise: Values in Action (tenetive)*. TBA, 2007.

[25] E. A. O'Hara. Trustworthiness and contract. In *Free Enterprise: Values in Action (tenetive)*. TBA, 2007.

[26] R. Boyd, H. Gintis, S. Bowles, and P. J. Richerson. The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences USA*, 100:3531–3535, 2003.

[27] H. M. Taylor and S. Karlin. *An Introduction to Stochastic Modeling*. Academic Press, New York, third edition, 1998.