Trust, Communication, and Inequality

Joanna Bryson (bryson@conjugateprior.org)
Department of Computer Science, University of Bath
Bath, BA2 7AY, United Kingdom
The Center for Information Technology Policy, Princeton University
303 Sherrerd Hall, Princeton, NJ 08544, USA

Paul Rauwolf (paulrauwolf@gmail.com)
The Institute for New Economic Thinking, Oxford University
Eagle House, Walton Well Rd, Oxford OX2 6ED, United Kingdom

Abstract

Inequality in wealth is a pressing concern in many contemporary societies, where it has been show to co-occur with political polarization and policy volatility, however its causes are unclear. Here we demonstrate in a simple model where social behavior spreads through learning that inequality can covary reliably with other cooperative behavior, despite a lack of exogenous cause or deliberate coordination. In the context of simulated cultural evolution selecting for trust and cooperative exchange, we find both cooperation and inequality to be more prevalent in contexts where the same agents play both the roles of the trusting investor and the trusted investee, in contrast to the condition where these roles are divided between disjoint populations. Cooperation is more likely in contexts of high transparency about potential partners and with a high amount of partner choice; while inequality is more likely with high information but no choice in partners for those that want to invest. While not yet a full model of contemporary society, our approach holds promise for examining the causality and social contexts underlying shifts in income inequality.

Keywords: trust; cooperation; inequality; behavioral ecology; agent-based modeling; social learning; cultural evolution

Introduction

Wealth inequality is an issue of considerable social concern. Striking the appropriate balance between egalitarian and meritocratic (for various definitions of *merit*) distributions of wealth and power has been a dominant political concern since at least the time of Marx (1867). In recent decades, the world as a whole has seen wealth inequality decrease, concomitant with a substantial drop in extreme poverty (Olinto et al., 2013). However in the wealthy countries of the OECD, inequality has been increasing. The present increase in inequality has sometimes been attributed to information technology, particularly artificial intelligence (Brynjolfsson & McAfee, 2014; Ford, 2015, for reviews). Although some argue that inequality *per se* is not so much of a concern as the wealth of a society's poorest members, others argue that radical differences in income lead to a lack of shared identity and political disenfranchisement (Plato, 380BC; Atkinson, 2015). While the causality is not yet established, there is good evidence that greater inequality increases political polarization (McCarty & Shor, 2016), leads to political volatility (Dutt & Mitra, 2008), and damages economic growth (Ostry et al., 2014). Establishing causal connections between these correlates is not only of scientific interest, but would have substantial consequences for public policy. However, establishing causality via interventions is both costly and potentially ethically problematic.

Here we propose that our understanding of the causes and consequences of inequality may be advanced if we leverage the more widely explored models of the dynamics of cooperative behavior. Cooperation is now understood to be as fundamental a biological behavior as competition (Marshall, 2015). Cooperation is a strategy by which 'selfish' genes can propagate themselves, as one vector of their transmission may pay a cost to increase the probability that other vectors survive and flourish. In some circumstances, such strategies generate absolute increases in overall advantage for a collection of cooperators; however, in all circumstances at least some individuals must focus sufficiently on their own survival to persist. Consequently, increasing and decreasing social investment may be an appropriate response to changing contexts (MacLean et al., 2010; Bryson et al., 2014); though oscillations in the dominance of social vs. independent strategies have also been shown to emerge without an exogenous environmental cause (Cavaliere et al., 2012).

The work presented here demonstrates that levels of inequality can also vary and emerge with no environmental trigger and no change discrete change in policy. We extend here a previous account of trust and cooperation in human economic transactions. This previous account used an evolutionary model to demonstrate two human-like behaviors: blind trust in unfamiliar strangers, and the costly rejection of unfair but profitable offers. Here we replicate and extend that model in the context of social learning rather than biological evolution. We also show that the results hold in a more biologically plausible spatial context (where interactions are only with neighbors), then examine the dynamics of inequality under both conditions. We find persistent inequality emerging only in limited circumstances, but transient inequality emerging as part of periodic collapses in cooperative behavior. These are transient because the context of trust required fades until the behaviors resulting in inequality are again suppressed.

Background and Previous Work

The Trust Game

In this paper we explore the dynamics of trust, exchange, and inequality in the context of an established model from the behavioral economics literature: the trust game. The trust game has two players playing anonymously, and an experimenter mediating their cooperation. The first player, **the investor**, is

given money by the experimenter, then offered the option of keeping the money, or entrusting it to an anonymous other: **the investee**. Both players know that the experimenter will not simply give the money to the investee, but also multiply it by a factor b. The investee can now keep all the money, or give any amount of it back to investor. In the present paper, we focus on one-shot games, where this is the sum total of all interactions between the two anonymous players. In such a context, there is no obvious motivation for the investee to return any money, but generally they do. Similarly, there is no reason for the investor to have blind faith in this return, but many do make the investment, though often that trust is misplaced (Berg et al., 1995; Güth et al., 1997).

Rauwolf & Bryson (2016) demonstrate that evolutionary pressure can be maintained for trusting in this game in some contexts, as described below. This work extends from that of a number of other theoreticians, who have been exploring whether experimentally-observed levels of trust might be explained if we assumed that investors have either knowledge of the investee's rate of return due to reputation, or at least a population-level expectation based on experience of behavior such as might be captured by evolution (Bear & Rand, 2016). Manapat et al. (2012) have shown that selective pressure for trust can arise when there is sufficient chance of knowing a trustee's return rate. This is true even when information is delayed and inconsistent (Manapat & Rand, 2012).

Tarnita (2015) has similarly shown trust should emerge in this context, however she also exploits the mechanism of a structured environment: that is, that interactions are more likely to occur with some individuals than others. Although sometimes overlooked in mathematical theory (Sober & Wilson, 1998, for a history), in practice it takes time to traverse physical space, so nearly all interactions between organisms in the real world have this property. Further, both because offspring tend to be born near their mothers and bear family resemblance, and also social species tend to learn behavior from each other, those nearer to you are more likely to behave as you do. These properties have been shown to support both the genetic and cultural evolution of cooperation, and their ubiquity may explain the similar ubiquity of cooperation in nature (Hamilton, 1964; Marshall, 2015).

Experimental researchers have also examined the effects of offering investors in the Trust Game information about the trustees' return rate before they decide whether to invest. Investors often reject offers that would be profitable ($r > 1/b$), implying an implicit demand for fairer returns (i.e. $\approx 1/2$ Manapat et al., 2012). Rejecting an offer even when it provides a net gain for the investor again seems irrational and maladaptive, but here again Rauwolf & Bryson (2016) show that it can be advantageous. If there is sufficient competition between potential investees, then a strategy of withholding benefit from those who offer too little return on investment can pay off in the long term.

Previous Work: Methods and Results

Rauwolf & Bryson (2016) use an evolutionary agent-based model to determine whether it might be adaptive to blindly trust someone whose rate of return is unknown, even preferring them to someone who has offered a rate that is beneficial but unfair. We do this in a context of partner choice. We assume that on any given round of the game, a trustee may interact with only a subset of the population. The number of potential partners (k) ranges from 1–5. The other independent variable in the experiment is the probability of knowing (q) the the return rate (r) of any one partner. Return rates are the sole adaptive feature of the population of investees. The population of investors have two adaptive characteristics: their probability of preferring to trust a stranger (t) over keeping their money to themselves when there is no available partner with a known suitable r, and the level of r they demand (d) before entering a game with a player with known r.

Throughout this paper, the benefit of investing b is set to 3, so the investee always receives triple the money invested. We also focus on the binary version of the game: the options available to the investor are only all or nothing (as per Güth et al., 1997). The decision rule of the investor therefore is the following: An investor knows r for j out of k potential partner trustees, where $j/k = q$, the probability of knowing r. An investor invests with trustee i who has return rate r_i with probability:

$$p(r_i) = \begin{cases} 1 & i \leq j;\ \max\limits_{1 \leq x \leq j} r_x = r_i \geq d \\ 0 & i \leq j;\ \max\limits_{1 \leq x \leq j} r_x \neq r_i \\ \frac{t}{k-j} & i > j;\ \max\limits_{1 \leq x \leq j} r_x < d \end{cases} \quad (1)$$

An entrusted trustee deterministically returns the investment $\times br$ to the investor, and keeps $b(1-r)$.

A population is initially seeded with random values of r, d and t drawn from the range $[0,1]$. The entire population is born simultaneously, and all live for 500 rounds. There are 500 agents of each type. Each round every investor is offered the opportunity to invest with one of k randomly selected trustees. After 500 rounds, a new generation of investors and trustees are selected, with the probability of a set of variables persisting into the next generation being directly correlated to an agents' wealth. In line with Manapat et al. (2012), each agent is selected for the next generation using the pairwise comparison process. When an attribute is added to the next generation it is slightly mutated over a Gaussian distribution with $\mu = 0$ and $\sigma = 0.01$. After 1000 generations, we observed the populations' averages for the three dependent variables (traits.) We found that the rate of return tended to depend on both information and competition between partners. Nearly all the investment is returned under conditions where all return values are known ($q \approx 1$) and there are many partners ($k \approx 5$), whereas return rates approached 0 as q did, or when $k = 1$. In contrast, trust and demand peaked not at the extreme limits of the parameter space, but at intermediate values. Trust was only selected for in contexts of partial information, and came to very high levels when both $0.2 < q < 0.7$

and $2 \leq k \leq 5$. For higher levels of q, trust drifts neutrally because r is driven high by direct competition. As the number of partners increases, the maximum information rate q for maintaining high t declined further, since effectively more partners provide more knowledge of return rate. The shape of the curve for demand was similar, although demand was always in the range $0.333 < d < 0.6$. The floor is set by b; any return lower than b means investments lose money. Again, where r was driven high by competition (high q), or where there is no information (low q), d drifts unused. Selection for the costly rejection of partners only occurred in the same situations as trust.

Methods: Replication and Extensions

We have produced a new ABM to replicate and extend our previous results. In the present paper, we wished to examine four further questions:

1. *Do the results hold for social learning?* Here we assumed a stable but not particularly cognitive population that learned socially, rather than evolved. Rather than learning from its own experience (which may be useless in a constantly shifting environment) the agents learn from each other. After each round of the game, agents copy the parameters of their potential trading partners if those partners have more money than they do.

2. *Does spatial structure alter the results?* We added location in an x,y grid as a parameter of the agents. In the spatial, local condition, partners were drawn from adjacent neighbors. We also ran a condition without spatial structure to confirm the replication. As reviewed above, space is expected to facilitate cooperation, and also diversity.

3. *What is the impact of investment castes?* Previous models had investors and investees in disjoint populations. We ran conditions assuming the same, and alternative conditions assuming that all individuals could invest, and any individual might be called upon to be an investee. We expected the unified population to show greater, more stable cooperation due to mutualism, as per Estrela et al. (2015).

4. *In what conditions if any will inequality emerge, and will it be stable?* This was our principle motivation for this study. We anticipated that extremely high rates of return (r) or demand (d) might generate inequality.

Simulations were run in NetLogo on a 33x33 grid, so with 1089 locations, each with one either a single agent, or in the situation where investors and investees were different castes, two agents, but these agents did not interact. The environment provided each agent with one unit of currency per round. The agents invested all of their wealth if they chose to invest, so the results could be cumulative; however this rapidly exceeded the arithmetic capacity of NetLogo, so all agents were also taxed at a flat rate every round. For spatial models, the world was assumed to be toroidal, so all agents had 8 neighbors, of which k would be drawn randomly on each round as potential investees.

Results

Our results show the following:

1. The dynamics of learning are very much like the dynamics of evolution—the outcomes are certainly comparable. This is assuming that social learning occurs implicitly or at least automatically. Fig. 1, row 1 replicates Rauwolf & Bryson (2016) most closely in both experimental condition and outcome.

2. Spatial assortment also had surprisingly little impact (Fig. 1, top two rows vs. bottom two).

3. The most striking result of our replication was the difference between a unified population of investors and investees (rows 2 and 4 of Fig. 1) vs. the original disjoint castes (rows 1 and 3). The unified population had a far greater range of values for which the return rate r was high. This result also clarified the contexts affording high levels of demand. Where there is high information (q), r is driven by pure competition and demand (d) drifts sufficiently low that it has not impact. When the populations are separate, for low q, r collapses and d again has no impact. But in the case where the traits are bound in a single population, there is sufficient selective pressure on d to keep costly punishment high even under low information. Interestingly, trust (t) and d seem to complement each other at least when r is sufficiently high to merit trust.

4. The only conditions that afforded substantial inequality were those with the single population (Fig. 2). The top 1% of the population ordinarily had between 1–1.5% of the wealth at least among the investors, although the investee-only caste suffered greater differentiation where there was either high information and no choice or low (but not zero) information and choice. But the greatest inequality, where the top 1% of the population had as much as 50% of the wealth, occurred only when investors and investees were the same individuals, and then only in conditions of high information ($q > 0.5$) and no partner choice ($k = 1$).

5. In the conditions best replicating human performance on the trust game, inequality surges periodically correlated with a drop in return rate. There is no stable equilibrium for cooperation in any of these models, but rather the dynamics are such that cooperation periodically erodes but then recovers. In our model, inequality was a characteristic result of these periods of collapse (Figure 3).

Discussion and Future Work

There is a growing body of evidence has shown that inequality and polarization have tracked each other closely in the USA since at least the 1880s, well before the onset of IT or

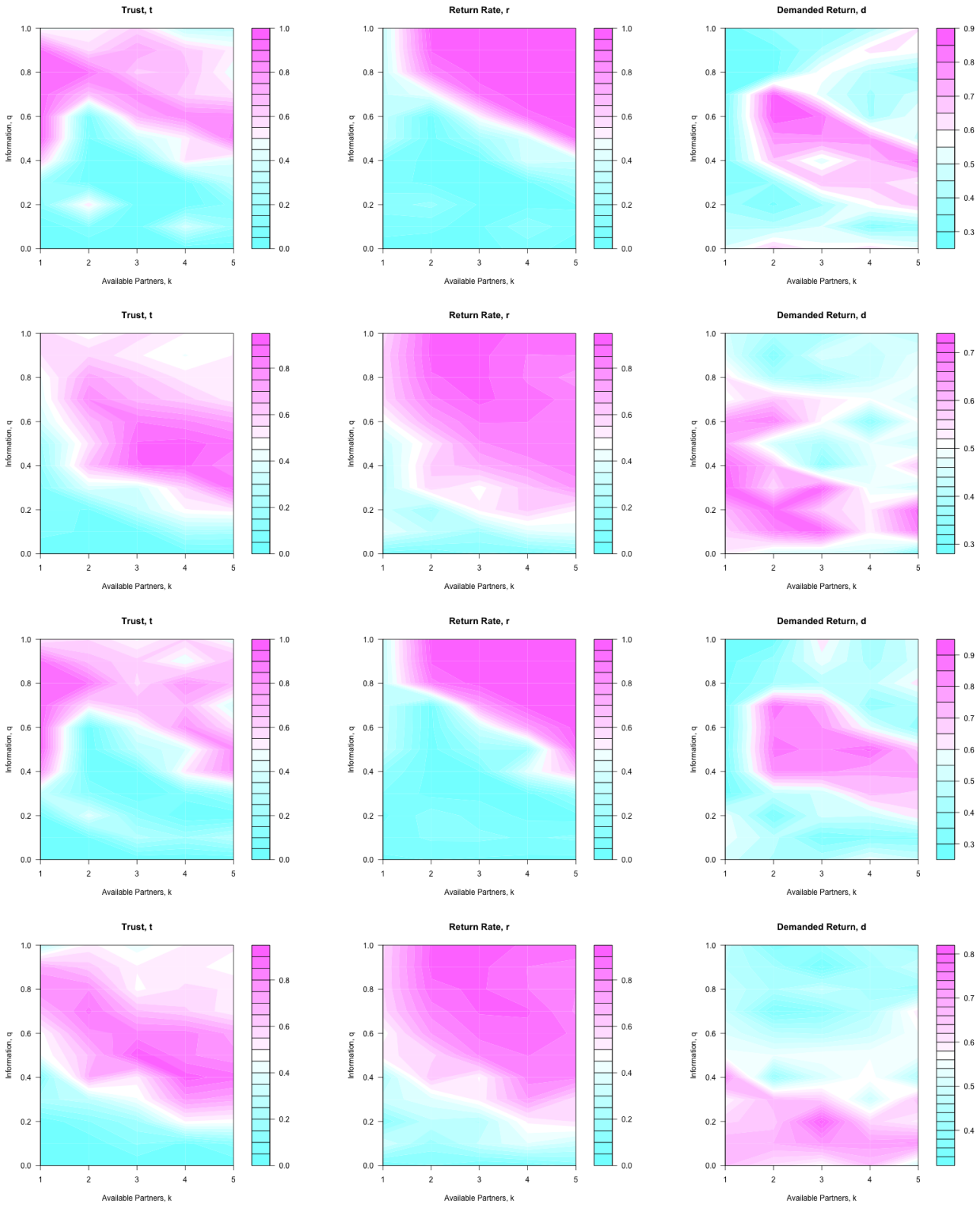

Figure 1: Trust, rate of return, and demand for the four conditions. The top two rows partners are randomly selected from the full population, the bottom two rows they are only selected from adjacent neighbors. The top and third row the populations of investors and investees are disjoint, the second and bottom row all agents play both roles. Values represent averages over $N = 1089$ agents for the final 300 time cycles (of 1200), for each of 5 runs. Note that index colors do *not* indicate consistent ranges, but rather vary by subfigure—see keys for values. See main text for interpretation.

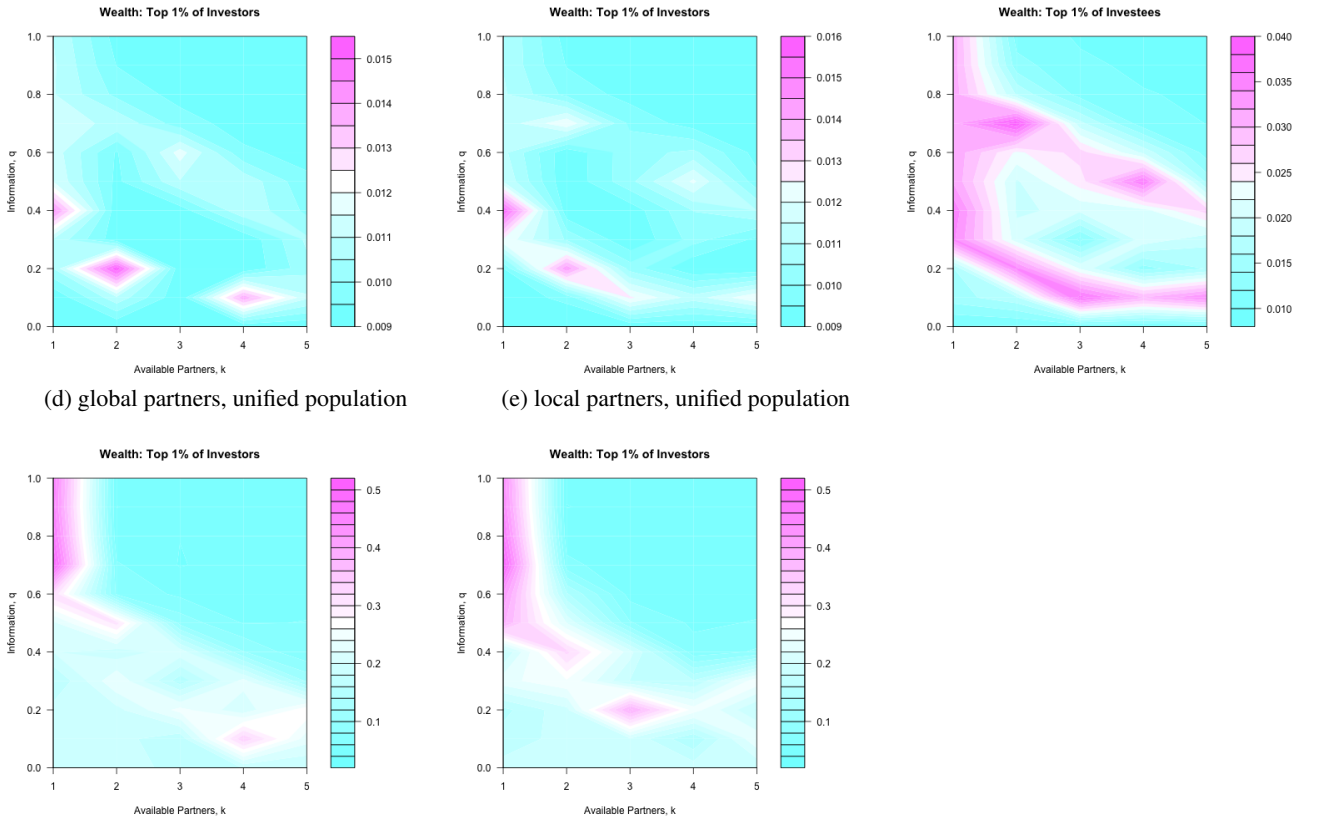

Figure 2: Inequality as measured by the proportion of wealth held by the top 1% of investors for each of the four conditions; values averaged as per Figure 1. Note that index color ranges are *not* consistent across subfigures, see keys. The only condition of extreme inequality occurs when investors and investees are drawn from the same population, there is no partner choice (although agents can still refuse to invest at all), and relatively high levels of information are available about potential investees. Note that localised investment reduces the required level of information to generate inequality.

AI. Both inequality and political polarization peaked immediately before and after World War I, then plummeted during the Great Depression (McCarty & Shor, 2016). Though now viewed as a norm, the long flat trough of both inequality and polarization between WWII and the 1980s may actually be the aberration, as ordinary oscillations in altruism such as were described by Cavaliere et al. (2012) were held in check by financial policy (Bryson & McCarty, 2016).

The results shown here are preliminary, but open the way to time-series analysis which may afford a better understanding of causality in this model, which can then be checked for match to the existing political and economic data. Certainly the dynamics of these models fluctuate greatly: there is no stable equilibrium, but rather tendencies for cooperation which occasionally are compromised for local profit. Creating better correlation to human society will probably also require modeling the stickiness of institutions.

Acknowledgments

Thanks to Will Lowe for assistance with the figures, and Nolan McCarty for discussions on inequality and polarisation. Rauwolf would like to acknowledge partial funding support through a University of Bath fees-waived PhD studentship. Bryson would like to acknowledge the Universities of Bath and Princeton for sabbatical support.

References

Atkinson, A. B. (2015). *Inequality: What can be done?* Harvard University Press.

Bear, A., & Rand, D. G. (2016). Intuition, deliberation, and the evolution of cooperation. *Proceedings of the National Academy of Sciences*, *113*(4), 936–941.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 122–142.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company.

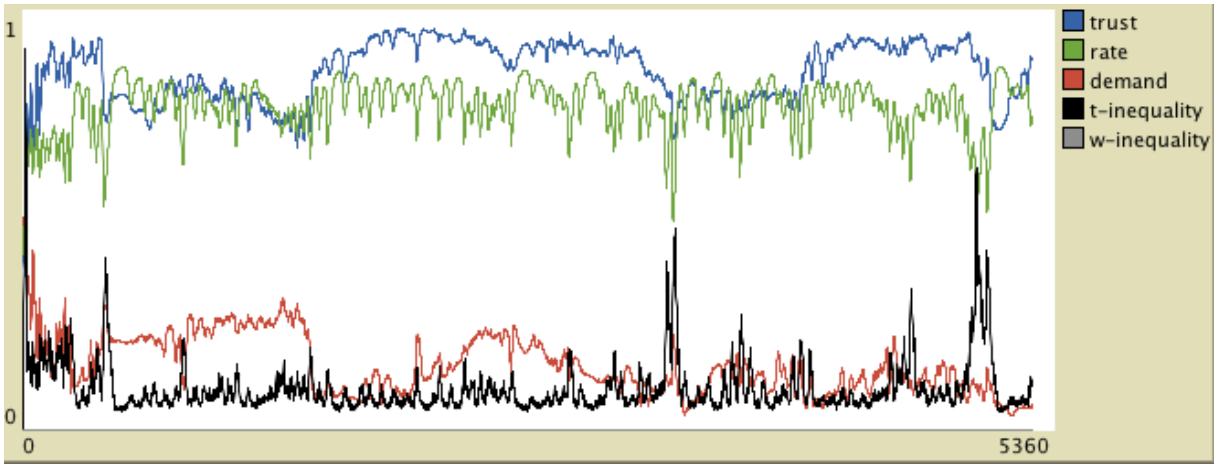

Figure 3: These simulations do not reach stable equilibria, but rather evolve dynamics that persistently converge in ways determined by the socio-economic context. In this exemplar, chosen because of the correspondence to human levels of trust and demand, the context is: the probability $q = 0.5$ of knowing a potential partner's return rate (r), and the number of available partners to choose between per game $k = 3$; investees are drawn from the same population as the investors (no castes); and they are always drawn from the possible 8 immediate neighbors (spatially localized exchanges). The measured values are investee return rate r, investor demand d and trust t—these are population averages ($N = 1089$); inequality is the proportion of overall wealth held by the richest 1% of agents.

Bryson, J. J., & McCarty, N. (2016). *Polarization and inequality: Towards a mechanistic account.* (in prep.)

Bryson, J. J., Mitchell, J., Powers, S. T., & Sylwester, K. (2014). Understanding and addressing cultural variation in costly antisocial punishment. In M. A. Gibson & D. W. Lawson (Eds.), *Applied evolutionary anthropology: Darwinian approaches to contemporary world issues* (pp. 201–222). Heidelberg: Springer.

Cavaliere, M., Sedwards, S., Tarnita, C. E., Nowak, M. A., & Csikász-Nagy, A. (2012). Prosperity is associated with instability in dynamical networks. *Journal of Theoretical Biology*, *299*(0), 126–138. doi: 10.1016/j.jtbi.2011.09.005

Dutt, P., & Mitra, D. (2008). Inequality and the instability of polity and policy. *The Economic Journal*, *118*(531), 1285–1314. doi: 10.1111/j.1468-0297.2008.02170.x

Estrela, S., Morris, J. J., & Kerr, B. (2015). Private benefits and metabolic conflicts shape the emergence of microbial interdependencies. *Environmental Microbiology*, n/a–n/a. doi: 10.1111/1462-2920.13028

Ford, M. (2015). *Rise of the robots: Technology and the threat of a jobless future*. Basic Books.

Güth, W., Ockenfels, P., & Wendel, M. (1997). Cooperation based on trust. an experimental investigation. *Journal of Economic Psychology*, *18*(1), 15–43.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. *Journal of Theoretical Biology*, *7*, 1–52.

MacLean, R. C., Fuentes-Hernandez, A., Greig, D., Hurst, L. D., & Gudelj, I. (2010, 09). A mixture of "cheats" and "co-operators" can enable maximal group benefit. *PLoS Biol*, *8*(9), e1000486. doi: 10.1371/journal.pbio.1000486

Manapat, M. L., Nowak, M. A., & Rand, D. G. (2012). Information, irrationality, and the evolution of trust. *Journal of Economic Behavior & Organization*.

Manapat, M. L., & Rand, D. G. (2012). Delayed and inconsistent information and the evolution of trust. *Dynamic Games and Applications*, *2*(4), 401–410.

Marshall, J. A. R. (2015). *Social evolution and inclusive fitness theory: An introduction*. Princeton University Press.

Marx, K. (1867). *Das kapital: Kritik der politischen oekonomie* (Vol. I). Hamburg: Otto Meissner.

McCarty, N., & Shor, B. (2016). *Partisan polarization in the united states: Diagnoses and avenues for reform.* (Available at SSRN 2714013)

Olinto, P., Beegle, K., Sobrado, C., & Uematsu, H. (2013). The state of the poor: Where are the poor, where is extreme poverty harder to end, and what is the current profile of the world's poor? *World Bank: Economic Premise*(125), 1–8.

Ostry, J. D., Berg, A., & Tsangarides, C. G. (2014). *Redistribution, inequality, and growth*. International Monetary Fund.

Plato. (380BC). *Republic*. Thrasyllus of Mendes.

Rauwolf, P., & Bryson, J. J. (2016). *Trust mediates costly punishment in environments of partial information and partner choice.* (under review)

Sober, E., & Wilson, D. S. (1998). *Unto others: The evolution and psychology of unselfish behavior*. Cambridge, MA: Harvard University Press.

Tarnita, C. E. (2015). Fairness and trust in structured populations. *Games*, *6*(3), 214. doi: 10.3390/g6030214