OPEN

How Natural Selection Can Create Both Self- and Other-Regarding Preferences, and Networked Minds

Thomas Grund[1,2], Christian Waloszek[1] & Dirk Helbing[1]

[1]Department of Humanities, Social and Political Sciences, ETH Zurich, Switzerland, [2]Centre International de Criminologie Comparée, Université de Montréal, Canada.

Biological competition is widely believed to result in the evolution of selfish preferences. The related concept of the 'homo economicus' is at the core of mainstream economics. However, there is also experimental and empirical evidence for other-regarding preferences. Here we present a theory that explains both, self-regarding and other-regarding preferences. Assuming conditions promoting non-cooperative behaviour, we demonstrate that intergenerational migration determines whether evolutionary competition results in a 'homo economicus' (showing self-regarding preferences) or a 'homo socialis' (having other-regarding preferences). Our model assumes spatially interacting agents playing prisoner's dilemmas, who inherit a trait determining 'friendliness', but mutations tend to undermine it. Reproduction is ruled by fitness-based selection without a cultural modification of reproduction rates. Our model calls for a complementary economic theory for 'networked minds' (the 'homo socialis') and lays the foundations for an evolutionarily grounded theory of other-regarding agents, explaining individually different utility functions as well as conditional cooperation.

Many societal problems, such as pollution, global warming, overfishing, or tax evasion, result from social dilemmas. In these dilemmas, uniform cooperation would be good for everybody, but each individual can benefit from free-riding[1]. Although societies found ways to cope with so-called 'tragedies of the commons'[2], the evolution of other-regarding preferences under competitive selection pressure is still a challenging and topical scientific puzzle.
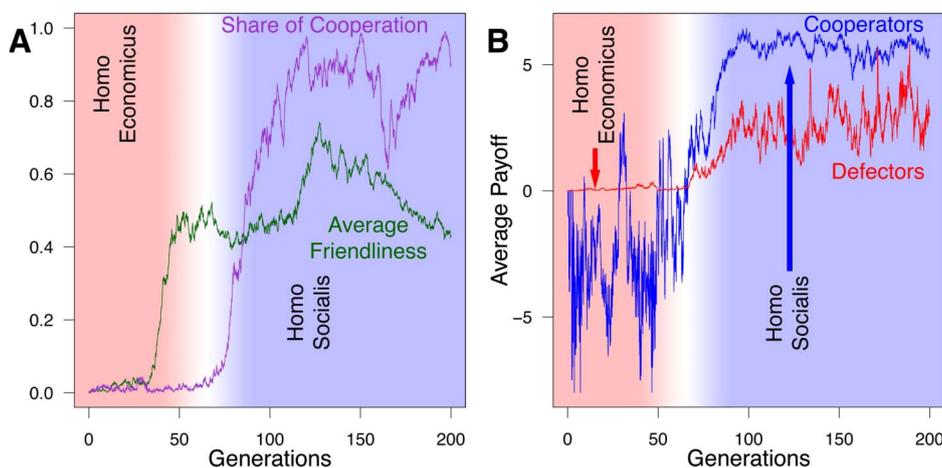
In social dilemma situations, caring about others can reduce individual success. While profit maximisation in single interactions would always demand non-cooperative behaviour, repeated interactions may sometimes support reciprocal altruism and result in human sociality[3,4]. But even in one-shot interactions, humans are not as selfish as theory suggests. A large body of experimental and field evidence indicates that people genuinely care about each other[5–8]. They tend to be not only concerned about individual success, but also about that of others[6,8,9].

Also in dictator and ultimatum games, the tendency to share is often attributed to other-regarding preferences[10,11]. But how did such other-regarding preferences evolve and spread? It was suggested that group selection would solve the puzzle[12–14] but it only works when groups do not mix[15]. Therefore, mechanisms not requiring kin or group selection have been looked for[16]. Some authors argue that humans intrinsically favour fairness[10,17]. But why do other-regarding preferences then vary geographically[6]?

To explain an innate sense of fairness, Gintis proposed a modification of reproduction rates by culture[18]. Other models studying the evolution of fairness preferences typically assume mechanisms in favour of pro-sociality, such as a 'shadow of the future'[4], costly punishment[19,20], reputation[21,22], genetic favouritism[23,24], genetic drift[25], or local interactions with an imitation of more successful neighbours[26–30]. However, the 'best response' rule is not favourable for the spreading of cooperation in social dilemma situations, where non-cooperative behaviour creates a higher payoff, no matter what the behavioural strategy of the interaction partner(s) is[31].

Results

Model. Our model does not require any of the previously mentioned social mechanisms and it even works when the best response rule is applied. For simplicity, we assume a spatial square lattice with periodic boundary conditions and $L \times L$ sites; 60% of the sites are occupied by one agent each, the other 40% are empty. Agents simultaneously interact with all other agents in their Moore neighbourhood—the eight sites surrounding their own site. In each time period, agents can choose to cooperate (C) or to defect (D). For all interactions with neighbours, agents get a

Figure 1 | A random spatial coincidence of friendly agents can lead to the sudden spreading of other-regarding preferences and a transition from a 'homo economics' to a 'homo socialis'. The graphs show representative simulation runs on a 30 × 30 spatial grid with periodic boundary conditions. 60% of all sites are occupied with agents who can either cooperate or defect. The payoff of interacting agents is determined as sum of payoffs from prisoner dilemma games with all Moore neighbours. The payoff parameters are: 'Temptation' $T = 1.1$, 'Reward' $R = 1$, 'Punishment' $P = 0$, and 'Sucker's Payoff' $S = -1$. The strategies (cooperation or defection) are updated simultaneously for all agents, applying the myopic best response rule to the utility function of each agent. It weights the payoffs of neighbours with the friendliness ρ_i and the own payoff with $(1 - \rho_i)$. Agents die at random with probability $\beta = 0.05$. To keep population size constant, surviving agents produce offsprings proportionally to their payoff in the previous round. Offspring move to the closest empty site ($v = 1$) and inherit attributes from the parent, here: the friendliness ρ_i. However, with probability $\mu = 0.05$, the friendliness of offsprings mutates. With probability 0.8 it is 'reset' to a uniformly distributed random value between 0 and the friendliness ρ_i of the parent, and with probability 0.2 it takes on a uniformly distributed value between ρ_i and 1. (A) Average of friendliness and share of cooperating agents as a function of time (one generation is $1/\beta$ periods). (B) Average payoffs of cooperators and defectors as a function of time. Initially, defectors are more successful than cooperators. However, when the sudden transition from a 'homo economics' to a 'homo socialis' occurs, the payoffs for defectors increases, but the payoffs for cooperators increases even more, which implies higher production rates of agents with other-regarding preferences.

payoff. If two interacting agents cooperate, each obtains the amount R ('Reward'); if both defect, each gets P ('Punishment'); and if one cooperates and the other one defects, the former one gets S ('Sucker's Payoff'), while the latter gets T ('Temptation'). The agents' reproductive fitness in time period t is given by the sum of all payoffs from interactions with neighbours (minus a value of $8|S|$ to ensure non-negative payoffs and avoid the reproduction of agents who are exploited by all their neighbours).

At the end of each period, individuals die with probability β. To keep population size constant, all agents who die are replaced by an offspring of one of the surviving agents. The likelihood of parents to create an offspring is strictly proportional to their actual payoff, i.e. their reproductive fitness. The offspring is born in one of the empty sites closest to the parent with probability v ('local reproduction'), while it occupies a random empty site irrespective of the distance to the parent with probability $1 - v$ ('random reproduction').
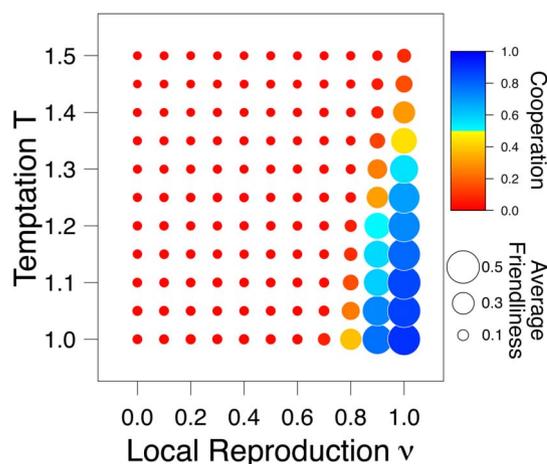
We assume a strict prisoner's dilemma with $T > R > P > S$. Although collective success is highest when everybody cooperates, defection is the payoff-maximising individual strategy in each single interaction, independently of the neighbours' strategies. In our model, individuals update their strategy (cooperation or defection) based on the myopic best response rule at the end of each period. However, rather than maximising their payoff P_i, we assume here that an individual i chooses the strategy that maximises the utility

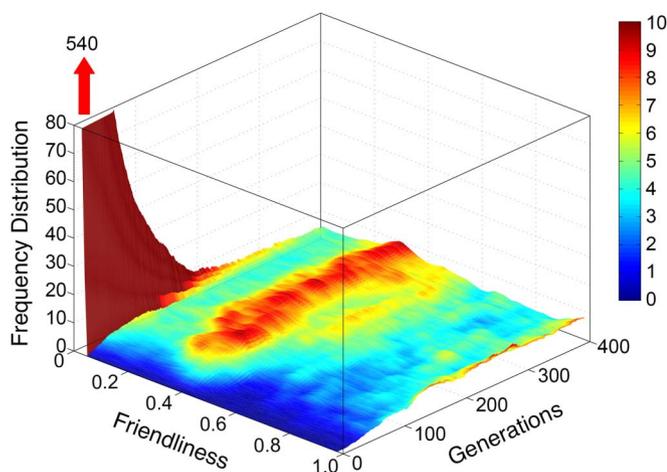$$U_i = (1 - \rho_i)P_i + \rho_i \bar{P}, \qquad (1)$$

where $\bar{P}$ denotes the average payoff of the interaction partners j. We do this because of previous studies and empirical evidence[10,29,32] suggesting that the utility is not just given by the own payoff P_i, but the payoff P_j of interaction partners j is also given a certain weight ρ_i. The variable $\rho_i \in [0, 1]$—the 'friendliness'—represents the degree of other-regarding preferences of agent i. A purely self-regarding individual with $\rho_i = 0$ only cares about the own payoff when choosing a strategy. An other-regarding individual gives the own payoff a weight

of $1 - \rho_i$ and the payoff of interaction partners a weight of ρ_i. Hence, strategy updates are assumed to be 'empathic', but reproduction is exclusively driven by individual payoff.

When selfishness is fixed ($\rho_i = 0$), best response behaviour promotes a 'tragedy of the commons'[1]. Instead, however, we assume a

Figure 2 | Local reproduction is crucial for the transition from a 'homo economicus' to a 'homo socialis'. The rate v of local reproduction determines the probability of an offspring to occupy the closest empty site to the parent. With probability $(1 - v)$, the offspring moves to an empty site that is randomly selected. All other parameters are specified as in Fig. 1. The circle size indicates average friendliness, while the circle colour represents the share of cooperators. The values are averages over 100 simulation runs between generation 100 and 500. Even at temptation levels around $T = 1.3$, the above phase diagram shows a sudden transition from self-regarding preferences (small dots) to other-regarding preferences (large circles), when the degree v of local reproduction is high enough.

Figure 3 | Evolution of the distribution of friendliness in the course of time for the parameter values used in Fig. 1. The plot shows an average over 100 runs, smoothed with MATLAB's local regression using weighted linear least squares and a 1st degree polynomial model. It is clearly visible that a broad distribution of individual utility functions results, even though everybody starts off with a purely self-regarding behaviour, for which the utility function agrees exactly with the payoff function.

(genetic or cultural) transmission of friendliness ρ_i from parent i to offspring j, which is subject to random mutation. In our model, mutation occurs with a constant probability μ that is independent of the strategies pursued in the neighbourhood. To avoid 'genetic drift', which would eventually promote friendliness scores of 0.5, the mutation of ρ_i is specified such that the offspring tends to be more self-regarding than the parent (if $\rho_i > 0.2$): With probability 0.8, ρ_j is set to a uniformly distributed random value between 0 and ρ_i, and with probability 0.2 it is set to a uniformly distributed value between ρ_i and 1.

Simulation results. Our computer simulations start in the most adverse condition for friendliness and cooperation. At time $t = 0$, all agents defect and nobody cares about the payoff of others ($\rho_i = 0$). However, mutations will eventually create higher levels of friendliness. According to the best response rule, an agent will

cooperate, if the utility $U_i(C)$ of cooperation is larger than the utility $U_i(D)$ of defection. The utility of cooperation is

$$U_i(C) = d[\rho_i T + (1 - \rho_i)S] + cR, \quad (2)$$

when surrounded by c co-operators and d defectors, and the utility of defection is

$$U_i(D) = dP + c[\rho_i S + (1 - \rho_i)T]. \quad (3)$$

Therefore, cooperation is expected to occur for

$$\frac{c}{d} > \frac{P - \rho_i T - (1 - \rho_i)S}{R - \rho_i S - (1 - \rho_i)T}. \quad (4)$$

That is, cooperativeness depends on the number of cooperative and defective neighbours, but it also depends on the level of friendliness ρ_i. We find that, for $\rho_i = 0$, agents *never* cooperate, while above a critical threshold of friendliness, namely for
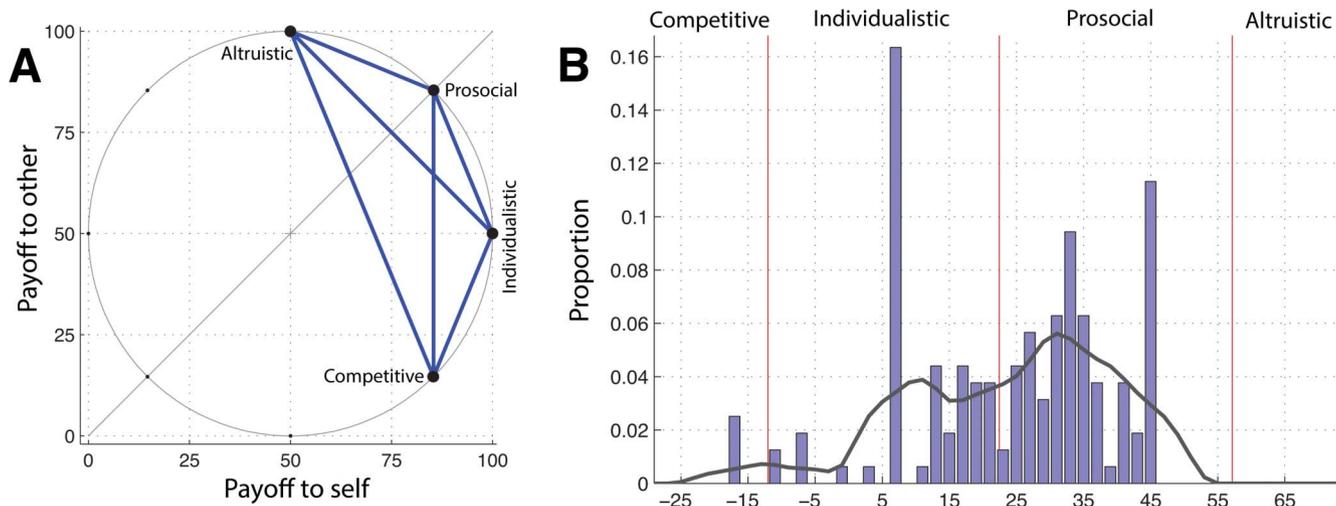
$$\rho_i > \frac{P - S}{T - S}, \quad (5)$$

they cooperate unconditionally. For values ρ_i of friendliness between $(T - R)/(T - S)$ and $(P - S)/(T - S)$, we find *conditional* cooperation[9], when enough neighbours cooperated in the previous round (note that, in our simulations, $S < 0$).

Hence, 'idealists' with a level of friendliness $\rho_i > (P - S)/(T - S)$ happen to cooperate even when they are surrounded and exploited by defectors. However, such idealists will normally get miserable payoffs and have very small reproduction rates. They tend to die without reproducing themselves. In fact, other-regarding preferences do not spread and selfishness thrives when offspring occupy randomly selected empty cells.

In contrast, when agents reproduce locally, other-regarding preferences suddenly emerge after some time (see Fig. 1A). How does this surprising, sudden transition from the 'homo economicus' to the 'homo socialis' take place? In principle, mutations could create a random co-location of mutation-borne 'idealists' by coincidence after a long time[28]. This would lead to the formation of a cluster of cooperators of 'supercritical' size. Such clustering would dramatically increase the relative fitness of other-regarding agents in the cluster and create sufficiently high reproduction rates to spread friendliness.

However, why does this transition happen in just a few generations (see Fig. 1B), i.e. much faster than expected? This relates to our distinction of preferences and behaviour. When an 'idealist' is born in a neighbourhood with friendliness levels supporting conditional

Figure 4 | Empirical distribution of other-regarding preferences (from R. Murphy et al. 2011[32], reproduction with kind permission of Ryan Murphy). *(A)* This figure shows the primary items from a 'Slider Measure' to determine Social Value Orientation. *(B)* Distribution of Social Value Orientation scores from the experimentally determined Slider Measure, as represented by angles. The dark line is a smoothed kernel density estimation.

cooperation, this can trigger off a cascade of changes from defective to cooperative behaviour. Under such conditions, a single 'idealist' may quickly turn a defective neighbourhood into a largely cooperative one. This implies higher payoffs and higher reproduction rates for both, idealists and conditional co-operators.

The intriguing phase transition from self-regarding to other-regarding preferences critically depends on the local reproduction rate (see Fig. 2). The clustering of friendly agents, which promotes other-regarding preferences, is not supported when offspring move away. Then, offspring are more likely to encounter defectors elsewhere and parents are not 'shielded' by their own friendly offspring anymore. In contrast, with local reproduction, offspring settle nearby, and a clustering of friendly agents is reinforced. Under such conditions, friendliness is evolutionary advantageous over selfishness.

Discussion

In conclusion, we offer an over-arching theoretical perspective that could help to overcome the historical controversy in the behavioural sciences between largely incompatible views about human nature. Both, self-regarding and other-regarding types of humans may result from the same evolutionary process. Whereas high levels of inter-generational migration promote the evolution of a 'homo economicus', low levels of intergenerational migration promote a 'homo socialis', even under 'Darwinian' conditions of a survival of the fittest and random mutations. The significance of local reproduction for the evolution of other-regarding preferences is striking and may explain why such preferences are more common in some parts of the world than in others[6].

Our modelling approach distinguishes between the evolution of individual preferences and behaviours. This makes cooperation conditional on the level of cooperation in the respective neigh-bourhood. Hence, when a few 'idealists' are born, who cooperate unconditionally, this can trigger off cooperation cascades, which can largely accelerate the spreading of cooperation[33]. Our model can also serve as a basis to develop an economic theory of other-regarding agents. The advantage is that it does not need to assume certain properties of boundedly rational agents—these properties rather result from an evolutionary process. In fact, our model naturally explains the evolution of individually different utility functions, as they are experimentally observed (see Figs. 3 + 4), and also the evolution of conditional cooperators[9,34].

A great share of economic literature is based on the assumption of the 'homo economicus', who takes decisions without considering the payoff or utility of others. In contrast to this traditional view, the 'homo socialis' never takes independent decisions, if the behaviour has external effects[35,36]. We might characterise this as a situation of 'networked minds', where everybody is trying to put himself or herself into other people's shoes, to take into account their utilities in the decision-making process. As a consequence, besides paying attention to networks of companies[37], economics should also consider networks of individual minds, i.e. social aspects. This is of particular relevance for information societies, in which individuals are increasingly connected via information and communication systems, such as social media[38,39]. A theory of networked minds could make a significant contribution to the convergence of the behavioural sciences[41], and it might also shed new light on social capital, power, reputation and value, and create a fundamentally new understanding of these[40]. We believe that this view can stimulate a huge and exciting field of research, and lead to a complementary theory to the one based on the 'homo economicus'. Due to the simplicity and fundamental character of the model proposed by us, we expect that it might serve as a starting point and basis for this new field.

1. Hardin, G. The tragedy of the commons. *Science* **162**, 1243–1248 (1968).
2. Ostrom, E. *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge University Press, Cambridge, 1990).
3. Trivers, R. L. The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35–57 (1971).
4. Axelrod, R. & Hamilton, W. D. The evolution of cooperation. *Science* **211**, 1390–1396 (1981).
5. Kahneman, D., Slovic, P. & Tversky, A. *Judgement Under Uncertainty-Heuristics and Biases* (Cambridge University Press, Cambridge, 1982).
6. Henrich, J. *et al.* In search of homo economicus: behavioral experiments in 15 small-scale societies. *Am. Econom. Rev.* **91**, 73–78. (2001).
7. Camerer, C. *Behavioral Game Theory—Experiments in Strategic Interaction* (Princeton University Press, Princeton, 2003).
8. Hoffman, E., McCabe, K. & Smith, V. Social distance and other-regarding behavior in dictator games. *Am. Econom. Rev.* **86**, 653–660 (1996).
9. Fehr, E. & Fischbacher, U. The nature of human altruism. *Nature* **425**, 785–791 (2003).
10. Tricomi, E., Rangel, A., Camerer, C. & O'Doherty, J. Neural evidence for inequality-averse social preferences. *Nature* **463**, 1089–1091 (2010).
11. Dufwenberg, M., Heidhues, P., Kirchsteiger, G., Riedel, F. & Sobel, J. Other-regarding preferences in general equilibrium. *Review of Economic Studies* **78**, 613–639 (2011).
12. Bowles, S. Group competition, reproductive levelling, and the evolution of human altruism. *Science* **314**, 1569–1572 (2006).
13. Traulsen, A. & Nowak, M. A. Evolution of cooperation by multilevel selection. *PNAS* **103**, 10952–10955 (2006).
14. Perc, M., Gómes-Gardeñes, J., Szolnoki, A., Floría, L. M. & Moreno, Y. Evolutionary dynamics of group interactions on structured populations: A review. *J. R. Soc. Interface* **10**, 20120997 (2013).
15. Sober, E. & Wilson, D. S. *Unto Others: The Evolution and Psychology of Unselfish Behavior* (Cambridge University Press, Cambridge, 1998).
16. Akay, E., Van Cleve, J., Feldman, M. W. & Roughgarden, J. A theory for the evolution of other-regard integrating proximate and ultimate perspectives. *PNAS* **106**, 19061–19066 (2009).
17. Fehr, E. & Schmidt, K. A theory of fairness, competition, and cooperation. *Q. J. Econ.* **114**, 817–868 (1999).
18. Gintis, H. Strong reciprocity and human sociality. *J. Theor. Biol.* **206**, 169–179 (2000).
19. Fehr, E. & Gächter, S. Altruistic punishment in humans. *Nature* **415**, 137–140 (2002).
20. Helbing, D., Szolnoki, A., Perc, M. & Szabó, G. Evolutionary establishment of moral and double moral standards through spatial interactions. *PLoS Comput. Biol.* **6** (4), e1000758 (2010).
21. Nowak, M. A. & Sigmund, K. Evolution of indirect reciprocity by image scoring. *Nature* **393**, 573–577 (1998).
22. Milinski, M., Semmann, D. & Krambeck, H. J. Reputation helps solve the tragedy of the commons. *Nature* **415**, 424–426 (2002).
23. Simon, H. A. Altruism and economics. *Am. Econom. Rev.* **83**, 156–161 (1993).
24. Wilson, D. S. & Dugatkin, L. A. Group selection and assortative interactions. *Am. Nat.* **149**, 336–351 (1997).
25. Leimar, O. & Hammerstein, P. Evolution of cooperation through indirect reciprocity. *Proc. R. Soc. Lond. B* **268**, 745–753 (2001).
26. Nowak, M. A. & May, R. M. Evolutionary games and spatial chaos. *Nature* **359**, 826–829 (1992).
27. Epstein, J. M. Zones of cooperation in demographic prisoner's dilemma. *Complexity* **4** (2), 36–48 (1998).
28. Helbing, D. & Yu, W. The outbreak of cooperation among success-driven individuals under noisy conditions. *PNAS* **106** (8), 3680–3685 (2009).
29. Szabó, G. & Szolnoki, A. Selfishness, fraternity, and other-regarding preference in spatial evolutionary games. *J. Theor. Biol.* **299**, 81–87 (2011).
30. Perc, M. & Szolnoki, A. Coevolutionary games – A mini review. *BioSystems* **99**, 109–125 (2010).
31. Roca, C., Cuesta, J. A. & Sanchez, A. Evolutionary game theory: temporal and spatial effects beyond replicator dynamics. *Phys. Life Rev.* **6** (4), 208–249 (2009).
32. Murphy, R. O., Ackermann, K. A. & Handgraaf, M. J. J. Measuring social value orientation. *Judgment and Decision Making* **6** (8), 771–781 (2011).
33. Fowler, J. H. & Christakis, N. Cooperative behavior cascades in human social networks. *PNAS* **107**, 5334–5338 (2010).
34. Gracia-Lazaro, C. *et al.* Heterogeneous networks do not promote cooperation when humans play a Prisoner's Dilemma. *PNAS* **109**, 12922–12926 (2012).
35. Christakis, N. A. & Fowler, J. H. *Connected* (Back Bay Books, Norwell, MA, 2011).
36. Bolton, G. E. & Ockenfels, A. ERC: A theory of equity, reciprocity, and competition. *The American Economic Review* **90** (1), 166–193 (2000).
37. Schweitzer, F. *et al.* Economic networks: The new challenge. *Science* **325**, 422–425 (2009).
38. Helbing, D. & Carbone, A. (eds.) Participatory science and computing for our complex world. *EPJ Special Topics* **214**, 1–666 (2012).
39. Helbing, D. Globally networked risks and how to respond. *Nature*, in print (2013).
40. Helbing, D. & Balietti, S. Fundamental and real-world challenges in economics. *Science and Culture* **76** (9–10), 399–417 (2010).

41. Gintis, H. *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences* (Princeton University Press, Princeton, 2009).

Acknowledgements

Author contributions

D.H. and T.G. developed the model and wrote the paper. T.G. and C.W. performed the computer simulations and the data analysis.

Additional information