Social Context and the Evolution of Empathy

Abstract

Human societies differ in the social context of their economic interactions: whereas hunter-gatherers depend primarily on familiars for their subsistence, people in market societies depend primarily on strangers. This variation raises a critical question: How does social context influence the evolution and behavioral expression of empathy? Experimental evidence consistently shows that empathy increases with social closeness. Using evolutionary game theory, I examine how empathy evolves in various social-economic ecologies. I situate the evolutionary game on a network with distinct social and economic layers (where strategy transmission and game interactions occur, respectively), and I define empathy as cooperating conditional on social proximity. The numerical and analytical results reveal that when interactions occur among both familiars and strangers, discriminatory empathy outperforms unconditional cooperation, but when interactions occur only among strangers, empathy produces no cooperation. Using pair approximation, I show that empathy is selected for when the gains from cooperation (b/c) exceed the degree of the social network (s). This pattern parallels Hamilton's rule, with 1/s corresponding to the relatedness coefficient. Broadly, the results illustrate that if behaviors spread among familiars, familiarity serves as a proxy for behavioral relatedness; in this way, conditioning cooperation on familiarity concentrates the benefits of cooperation among cooperators, allowing this behavior to persist and spread. These findings highlight the behavioral consequences of modern human ecologies characterized by anonymity and provide insights for designing institutions and structuring human ecologies that better foster cooperation.

Keywords: Empathy; Cooperation; Cultural evolution; Evolutionary game theory; Social networks

1. Introduction

We are a social species. Our aptitude for cooperation and social learning has enabled us to adapt to a wide variety of ecosystems and organize into diverse sociocultural configurations. This array of human ecologies gives rise to cooperation problems that range in scale and scope from mammoth hunts to climate change. Simultaneously, human ecology shapes the evolution and expression of cooperative behaviors such as empathy. In this paper, ecology refers to the pattern of relationships between individuals and their environments (Gonzalez & Rice, 2024), cooperation refers to any action that is individually costly but collectively beneficial (Rand & Nowak, 2013), and empathy refers to an internal state that motivates cooperative action contingent on social proximity (Davis, 2015). The social context of economic interactions constitutes an important dimension of human ecology that varies across different societies. Whereas hunter-gatherers know most of those whom they depend upon for subsistence, the constituents of market societies depend primarily on strangers. In this paper, I explore the effects of social context on the evolution and behavioral expression of empathy.

Empirical evidence indicates that cooperation increases with the sociality of the interaction context. Game theoretic experiments have demonstrated this effect by manipulating social context in various ways. Reducing anonymity by identifying subjects by name or photograph or by conducting the game in a face-to-face setting tends to promote cooperation (Ariely et al., 2009; Bohnet & Frey, 1999; Burnham, 2003; Charness & Gneezy, 2008; Choi & Storr, 2020; de Oliveira et al., 2014; Frey & Bohnet, 1997; Krysowski & Tremewan, 2021; Leider et al., 2008; Ligon & Schechter, 2012; List et al., 2004; Rankin, 2006; Rege & Telle, 2004; Schram & Charness, 2015), and so does permitting discussion among subjects before the game (Andreoni & Rao, 2011; Batson & Ahmad, 2001; Batson & Moran, 1999; Bochet et al., 2006; Frey & Bohnet, 1997; Ostrom & Walker, 1997; Sally, 1995). Revealing the group identities of subjects tends to prompt in-group cooperation (Bicchieri et al., 2022; Glaeser et al., 2000; Habyarimana et al., 2007; Howe et al., 2023). Most notably, in non-anonymous games, social distance between interaction partners in their real-world social networks correlates negatively with cooperation (Apicella et al., 2012; Brañas-Garza et al., 2010; Chandrasekhar et al., 2018; Goeree et al., 2010; Leider et al., 2008). Similarly, questionnaires of hypothetical choices suggest that cooperation increases with subjective relational closeness (Jones, 2022; Jones & Rachlin, 2009; Jones & Rachlin, 2006; Kardos et al., 2023.; Locey et al., 2011, 2013; Osiński, 2009; Rachlin & Jones, 2008; Vekaria et al., 2017).

Psychologists also expound the importance of relational context in understanding cooperative behavior (Clark et al., 2015; Kardos et al., 2023), especially in the literature on empathy. Although empathy has been defined in a number of ways (Batson, 2009; Davis, 2015), the concept generally refers to a cognitive, affective, and/or behavioral response of one individual to the state of another (Davis, 2015; de Waal & Preston, 2017). Empathic responses are mediated by psychological and social context (Cialdini et al., 1997; Davis, 2015; Hoffman, 2008; Preston & Waal, 2002; Singer & Lamm, 2009; Singer & Tusche, 2014). Namely, empathy increases with familiarity, similarity, past experience, and salience (de Waal, 2009; de Waal & Preston, 2017; Hoffman, 2008; Loewenstein & Small, 2007; Preston & Waal, 2002; Singer & Tusche, 2014). These effects of "psychological distance" (Davis, 2015) or "relationship closeness" (Cialdini et al., 1997) encapsulate the general observation that the strongest empathic responses occur between close friends or family and the weakest between out-group strangers (Davis, 2015). The internal cognitive and affective aspects are often associated with cooperative behavior (Batson et al., 1981; Batson, 2009, 2010; Batson & Ahmad, 2001; Batson & Moran, 1999; Cialdini et al., 1997; Davis, 2015; de Waal, 2009; de Waal, 2008; de Waal & Preston, 2017; Hoffman, 2008; Kamas & Preston, 2021; Preston & Waal, 2002; Rumble et al., 2010; Singer & Lamm, 2009). In game theoretic experiments, encouraging perspective-taking in the game instructions increases cooperation (Andreoni & Rao, 2011; Batson & Ahmad, 2001; Batson & Moran, 1999, p. 199; E. Hoffman et al., 2000; Rumble et al., 2010). Measures of dispositional empathy sometimes correlate with cooperation (Edele et al., 2013; Kamas & Preston, 2021; Takagishi et al., 2010), and sometimes do not (Artinger et al., 2014; Büchner et al., 2007; de Oliveira et al., 2014; Edele et al., 2013; Pelligra, 2011; Sautter et al., 2007), but notably, all those experiments that did not demonstrate an effect were conducted in anonymous settings.

These experimental findings suggest we are more prone to empathize and cooperate in situations that are more saliently social and with individuals who are socially closer to us. How might this strategy of cooperating conditional on social proximity have evolved? And how does the social context of interactions influence its evolution?

Whereas situational effects of sociality on cooperation are readily observable in experimental settings, evolutionary effects are not. Evolutionary game theory provides a framework for analyzing evolutionary dynamics by modelling the persistence and spread of behavioral strategies in a population as a function of the relative payoffs the strategies yield in intrapopulation game interactions. The human behavioral repertoire (including cooperation) is a product of interacting genetic and cultural evolution (Boyd & Richerson, 1988, 2005; Henrich & Muthukrishna, 2021). Genes persist and spread via survival and reproduction of the carrier. Cultural traits spread via social learning. Given the long history of misuse and misinterpretation of evolutionary concepts applied to humans (Gonzalez & Rice, 2024), it is important to distinguish between explanation and justification: evolution is not synonymous with progress nor is fitness synonymous with desirability.

Cooperation poses an evolutionary puzzle given the separation of the actor from the benefits of their action. If individuals interact randomly with each other and employ unconditional strategies (always cooperate or never cooperate), those who never cooperate inevitably accumulate a higher payoff than those who always cooperate. Cooperation can only evolve if the benefits of cooperation are sufficiently concentrated among cooperators, such that the average cooperation benefits received by cooperators exceed the cooperation costs they pay. This logic applies to both cultural and genetic evolution and can thus produce both learned and innate cooperative traits. Hamilton's rule expresses this logic in mathematical form: $rb > c$, where b and c represent, respectively, the benefits and costs of cooperation and r, the relatedness coefficient, represents the proportion of cooperation benefits received by cooperators (Bowles & Gintis, 2011; Hamilton, 1964; Henrich, 2018; Smaldino, 2023). This concentration of benefits among cooperators can occur through assortment due to the coincidence of behavioral transmission and interaction ecologies as in the cases of kin selection (Hamilton, 1964), group selection (Bowles & Gintis, 2011), and static networks (Rand et al., 2014), or it can arise through conditional behavior as in the cases of direct and indirect reciprocity (Nowak & Sigmund, 2005; Schmid et al., 2021; Trivers, 1971), parochial altruism (Bowles & Gintis, 2011), altruistic punishment (Boyd & Richerson, 2005), and choice of interaction partners (which results in assortment) (Rand et al., 2011). These evolutionary mechanisms may manifest in a variety of proximate behavioral motives including social emotions (Batson, 2010; Bowles & Gintis, 2011; Davis, 2015; de Waal & Preston, 2017; Hoffman, 2008) and internalized or externally enforced social and moral norms (Akçay & Van Cleve, 2021; Bowles & Gintis, 2011; Boyd & Richerson, 2005; Ohtsuki & Iwasa, 2006; Ostrom, 2000).

Models of evolutionary games situated on networks illuminate the effects of population structure on the evolution of cooperation (e.g. Allen et al., 2017; Débarre et al., 2014; Lieberman et al., 2005; Nowak et al., 2010; Nowak & May, 1992; Ohtsuki et al., 2006; Sheng et al., 2024; Su et al., 2019, 2022, 2023; Su & Stewart, 2025). A seminal paper by Ohtsuki, Hauert, Lieberman, and Nowak (2006) demonstrates that, in the case of a single-layer network where both game interactions and strategy transmission occur, unconditional cooperation can persist and spread if the gains from cooperation are larger than the number of interaction and transmission neighbors. The network structure results in strategy assortment, thereby promoting the evolution of cooperation. A number of subsequent studies have shown that separating the loci of strategy transmission and game interactions can undermine the evolution of unconditional cooperation by lessening the assortment of strategies (Allen & Nowak, 2014; Dong et al., 2023; Ohtsuki et al., 2007; Su et al., 2019). Since genetic transmission occurs between kin and cultural transmission (social learning) occurs among both kith and kin, the transmission network might best be thought of as a network of kith and kin—i.e. a social network. This, in turn, suggests that unconditional cooperation can evolve only when the interaction context is highly social. But what if cooperation is conditioned on social proximity?

To explore the effect of social context on the evolution of empathy, I develop a model consisting of an evolutionary game situated on a network with distinct social and economic layers (Fig. 1, 2). The game interactions occur on the economic layer, and the strategy transmission occurs on the social layer. In this way, I can vary the social context of the evolutionary game by varying the social-economic overlap of the network. I define empathy as the strategy of cooperating with those within a certain range in a social network and defecting against everyone further away. While several others have operationalized empathy in a utility function or as a game theoretic strategy (Binmore, 2005; Grohn et al., 2014), the social distance conditionality is novel. Each of the situational and relational mediating factors associated with increased empathy in the psychology literature (Preston & Waal, 2002) can be related to proximity in a social network: familiarity by the definition of a social relation, similarity due to network homophily (McPherson et al., 2001), past experience from durable social relationships, and salience in the context of face-to-face interactions. To analyze the evolutionary viability and stability of empathy, I estimate fixation probabilities numerically by running simulations of the evolutionary game model, and I employ a novel method involving pair approximation (Baalen, 2000; Ohtsuki et al., 2006, 2007) to obtain an analytical rule for the evolution of empathy.

2. Model

The model I develop consists of a prisoner's dilemma game, the competing strategies for playing the game, the fitness function dictating how strategies spread, and the multilayer network on which the evolutionary game dynamics occur. I run evolutionary game simulations and apply a technique called pair approximation to derive numerical and analytical results describing the impact of social context on the evolution of empathy.

The two-layer network structures the interactions among the population of N nodes. The game occurs on the economic layer (G), and the strategy updating occurs on the social layer (S). The overlap network (O) is obtained from the intersection of the edge sets of G and S. The economic, social, and overlap networks are all regular degree: Every node has g economic neighbors and s social neighbors, o of which are both social and economic neighbors. (The term "regular" refers to the constant number of neighbors or "degree".) The sociality of the interaction context is represented by the overlap degree divided by the economic degree o/g, which indicates the proportion of interactions which occur among familiars. To generate a random regular multilayer network from the set of all possible regular multilayer networks given the parameters N, g, s, and o, I construct a circulant network of degree $g + s - o$, partition the edges into $(G' - O')$, $(S' - O')$, and O'; stochastically rewire the edges of each partitioned edge set, $G - O = R(G' - O')$, $S - O = R(S' - O')$, and $O = R(O')$; and reconstruct $G = (G - O) + O$ and $S = (S - O) + O$. Figure 1 illustrates this network generation algorithm, which I developed by modifying the Watts-Strogatz model (Watts & Strogatz, 1998). A detailed description can be found in Appendix A.

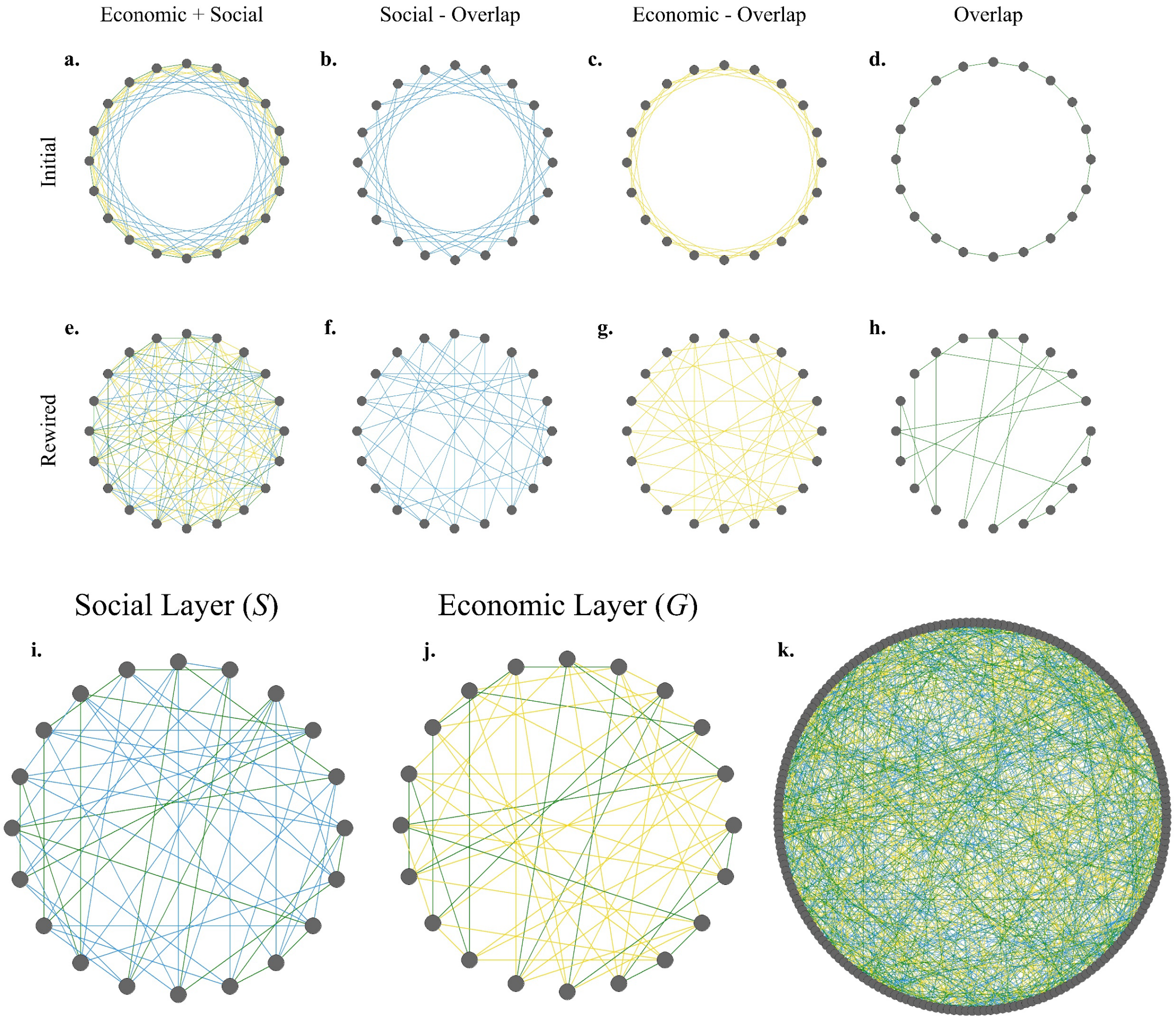
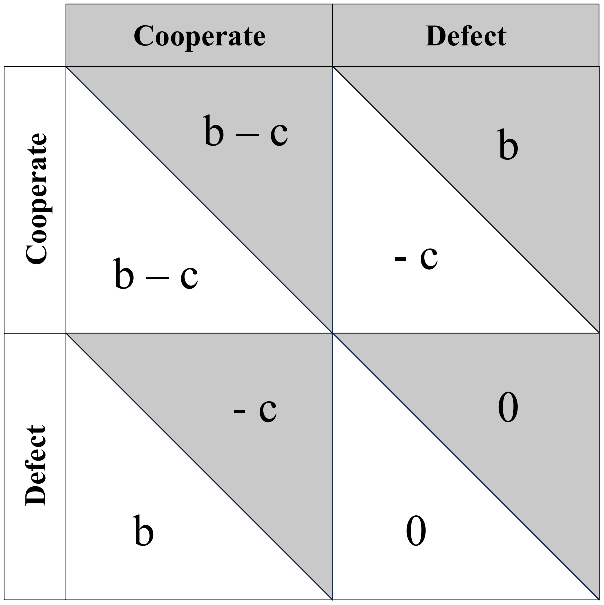


Fig. 1. Generating a multilayer network. The evolutionary game takes place on a random regular N-node multilayer network with distinct social and economic layers. Both layers are random regular networks, and the intersection of the layers is also a random regular network, so every node has the same number of economic neighbors (g), the same number of social neighbors (s), and the same number of social-economic neighbors (o). The random regular multilayer networks used in the simulations are constructed by partitioning and rewiring circulant networks. **a.–h.** depict the rewiring of the circulant union network, $S \cup G = R(S' \cup G')$, and each of its partitioned subsets: $S - O = R(S' - O')$, $G - O = R(G' - O')$, and $O = R(O')$ ($N = 20$, $g = s = 4$, $o = 2$). **i.** depicts the social layer obtained from $S = (S - O) + O$ (i.e. from the union of **f.** and **h.**). **j.** depicts the economic layer obtained from $G = (G - O) + O$ (i.e. from the union of **g.** and **h.**). **k.** depicts an example of a multilayer network used in the actual simulations ($N = 200$, $g = s = 8$, $o = 4$) with the edges of $S - O$, $G - O$, and O colored blue, yellow, and green, respectively.

The nodes of this network interact and update their strategies in a series of rounds. During each round, every node plays a one-shot game with each of their economic neighbors. In these games, each player chooses to either cooperate or defect (Fig. 2). The act of cooperation consists of paying a cost c to convey a benefit b to the other player, where $c > 0$ and $b > c$. The act of defection is costless and conveys no benefit to the other. The game defined by these actions and payoffs is an example of a prisoner's dilemma.



	Cooperate	**Defect**
Cooperate	b – c b – c	b - c
Defect	- c b	0 0

Fig. 2. Game payoffs. Each round, every node plays a one-shot prisoner's dilemma with each of its economic neighbors, accumulating payoffs according to the payoff matrix shown. *b* and *c* denote the benefits and costs of cooperation, respectively.

Each node plays the game according to their inherited or learned behavioral strategy. This strategy is defined by an empathic range variable, $x_i \in \mathbb{N}$, which indicates that node *i* will cooperate with every economic neighbor who is x_i or fewer social edges away and defect against all others. In other words, i will cooperate with j if and only if $d_{i,j} \leq x_i$, where $d_{i,j}$ is the distance between i and j in the social layer of the multilayer network. If $x_i = 0$, node *i* defects unconditionally. If $x_i \geq \text{diameter}(S)$, node *i* cooperates unconditionally. If $x_i \in (0, \text{diameter}(S))$, node *i* cooperates conditional on social proximity. Throughout this paper, I refer to these three types as defectors, unconditional cooperators, and empathizers, respectively. Note that empathizers who cooperate only with social neighbors $(x_i = 1)$ behave identically to unconditional cooperators in the context of full sociality $(o/g = 1)$ but behave identically to defectors in the absence of sociality $(o/g = 0)$. The prevalence of cooperation in a population of empathizers thus depends directly on the sociality of the interaction context. I demonstrate in Appendix B how this behavioral strategy defined by x_i can be obtained from a utility function with an other-regarding component that decays with social distance.

Each round, each node, i, accumulates a total payoff from the games played with their economic neighbors:

$$\Pi_i = \sum_{m \in N_i(G)} \pi_{i,m} \tag{1}$$

$N_i(G)$ denotes the set of economic neighbors of i. The fitness of node i (i.e. i's propensity to spread its behavioral strategy) is then calculated from this total payoff according to the following equation (Ohtsuki et al., 2006, 2007):

$$f_i = 1 - w + w \cdot \Pi_i \tag{2}$$

where $w \in [0,1]$ is the strength of selection. This parameter specifies the importance of payoffs in the updating dynamic. When $w = 0$, strategies spread according to neutral drift (i.e. they spread stochastically and independently from payoffs); when $w = 1$, strategies spread according to the relative size of the payoffs reaped by the strategy-bearing nodes.

Each round, after all games have been played, one node is selected randomly from the population to update their strategy by emulating one of their social neighbors. This updating mechanism is called the death-birth rule (Ohtsuki et al., 2007). The updating node, i, adopts the strategy of any one of their social neighbors, j, with a probability proportionate to j's relative fitness:

$$\mathrm{P}(i \text{ emulates } j) = \frac{f_j}{\sum_{m \in N_i(S)} f_m} \tag{3}$$

where $N_i(S)$ is the set of social neighbors of i.

I analyze the evolutionary properties of cooperation in this model by employing two different approaches, one numerical and the other analytical. For the first, I run simulations to estimate fixation probabilities (the probability of a single invading strategies spreading to the entire population). For the second, I apply the technique of pair approximation to derive an analytical rule describing the game and network conditions under which cooperation can spread.

3. Results

3.1. Numerical results

Simulations begin with a population containing multiple strategies and run until one of the strategies has fixated (i.e. eliminated its competitor(s)). The fixation probability of a strategy is estimated based on the proportion of 8000 simulations which end with that strategy fixating. I focus primarily on the case of invasion, where one "mutant" is introduced into a homogeneous population (Fig. 3a, 3b, 4a-d). I am interested not only in the conditions under which cooperators might invade a population of defectors (viability) but also the conditions under which a population of cooperators might fend of invading defectors (stability). Both cooperative and non-cooperative equilibria can be found in any society, so understanding how cooperation is maintained is just as important as understanding how it is initially fostered. And as the results show (Fig. 3, 4), viability and stability are not symmetric since cooperators rely on assortment for their fitness advantage.

To run one of these simulations, I must specify network parameters, N, g, s, and o, game parameters, b and c, and the selection strength, w. I limit the N to 200 and g and s to 8 in my main numerical analysis (Fig. 4), and my analytical results suggest little generality is lost in this specification. I vary the overlap degree, o, from 0 to 8 by steps of 2 to analyze the evolutionary dynamics under various social contexts. (The step size is constrained by the network generation algorithm.) I vary the relative gains from cooperation, b/c, around a median value of 8 to capture the interactive effects of network structure and game payoffs on the evolutionary dynamics. Earlier findings by Ohtsuki et al. (2006, 2007) as well as my own numerical and analytical results suggest that the boundary of the evolutionary feasibility of cooperation lies in this region of the game-network parameter space (where b/c is close to g and s). I vary b/c by varying b and holding c constant at $c = 1$. Multiplying both b and c by a scalar alters the payoff of every outcome by the same factor. Scaling the payoffs up or down in this manner is equivalent to altering the strength of selection (w). I conduct my primary analyses under conditions of weak selection ($w = 0.1$) in accordance with much of the theoretical literature on the evolution of cooperation on networks (Allen et al., 2017; Allen & Nowak, 2014; Dong et al., 2023; Ohtsuki et al., 2006, 2007; Tarnita et al., 2009). In Appendix C, I discuss selection strength more extensively and analyze its impact on the evolution of cooperation (see Figures C.1 and C.2).

The fixation probability estimates of the evolutionary game simulations indicate that whereas unconditional cooperation relies on a highly social interaction context to evolve, empathy (cooperation conditional on social proximity) can persist and spread when this sociality is low as long as the relative gains from cooperation are sufficiently large (Fig. 4). In other words, the numerical results suggest that conditioning cooperation on social proximity is adaptive when interactions occur among both familiars and strangers.

A strategy is selected for if its fixation probability is larger than that of a neutral mutant $(1/N)$. In Figures 3 and 4, fixation probability estimates for invading cooperators larger than (smaller than) $1/N$ are colored blue (red). Similarly, fixation probability estimates for resident cooperators larger than (smaller than) $1-(1/N)$ are colored blue (red). Thus, blue (red) cells indicate game and network conditions under which cooperation is selected for (against). Note that in the absence of sociality ($o = 0$), empathizers are no different from defectors, so their fixation probability estimates hover around those of neutral mutants, and in the context of full sociality ($o = g$), empathizers are no different than unconditional cooperators, so their fixation probability estimates align.

Figures 3a and 3b report the fixation probability estimates for empathizers with different empathic ranges ($x = 1, 2, 3, 4, 5$) competing against defectors in the context of partial sociality $(o/g = 0.5)$ when the gains from cooperation are slightly larger than the social and economic degrees $(b/c = 6 > g = s = 4)$. These conditions select for all empathy variants with $x \leq 3$. More generally, fitness decreases with empathic range. Figure 3c reports the fixation probability estimates for simulations run under the same game and network conditions but starting from a uniform distribution of all seven strategies from $x = 0$ (defectors) to $x = 6$ (unconditional cooperators. In this environment of many competing strategies, as in the case of a single competing strategy, discriminatory empathizers ($x = 1, 2$) are selected for.

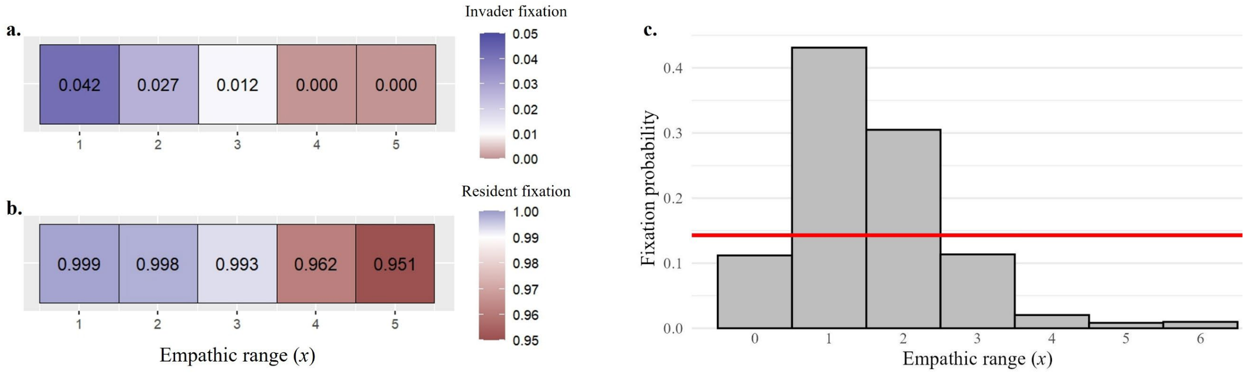


Fig. 3. The evolutionary advantage of conditioning cooperation on social proximity. *N* = 100, *g* = *s* = 4, *o* = 2, *w* = 0.1, *b* = 6, *c* = 1, 8000 simulations per estimate. In the case of partial sociality (*o*/*g* = 0.5), evolutionary game simulations suggest that conditioning cooperation on social proximity is adaptive. Fitness appears to decrease with empathic range. **a,** The fixation probability estimates for empathizers with different ranges introduced into a population of defectors (viability). **b,** The fixation probability estimates for resident empathizers with different ranges fending off invading defectors (stability). Empathizers with a range of 1 cooperate only with social neighbors whereas empathizers with a range of 5 cooperate with almost everyone. If a cell is blue (red), the empathizers are selected for (against). **c,** The fixation probability estimates for empathizers with ranges 0 (defectors) through 6 (unconditional cooperators) starting from a uniform distribution of all seven strategies, $x_i \sim \mathcal{U}(0,6)$. Strategies with fixation probabilities above (below) the red line—marking the fixation probability of a neutral mutant—are selected for (against).

Figure 4 reports the fixation probability estimates for unconditional cooperators ($x \geq \mathrm{diameter}(S)$) and empathizers ($x = 1$) competing against defectors under various conditions of sociality and gains from cooperation. The fitness of unconditional cooperators increases with gains from cooperation as well as sociality. Empathizer fitness increases with gains from cooperation, but its relation to social context is more complex. In the context of partial sociality ($o/g > 0$) and when the gains from cooperation exceed the degrees of the social and economic layers ($b/c > 8 = s = g$), empathizers are selected for (viability fixation probability estimates exceed $1/N$, and stability fixation probability estimates exceed $1 - (1/N)$). Below this b/c threshold, empathizers are selected against. Viability fixation probability estimates increase with sociality when $b/c > 8$, but stability fixation probability estimates decrease with sociality when $b/c < 8$. The explanation is this: as sociality increases, empathizers cooperate more, and when the gains from cooperation are sufficiently high, this increases their average fitness relative to competing defectors, but when the gains from cooperation are low, this decreases their average fitness relative to competing defectors.

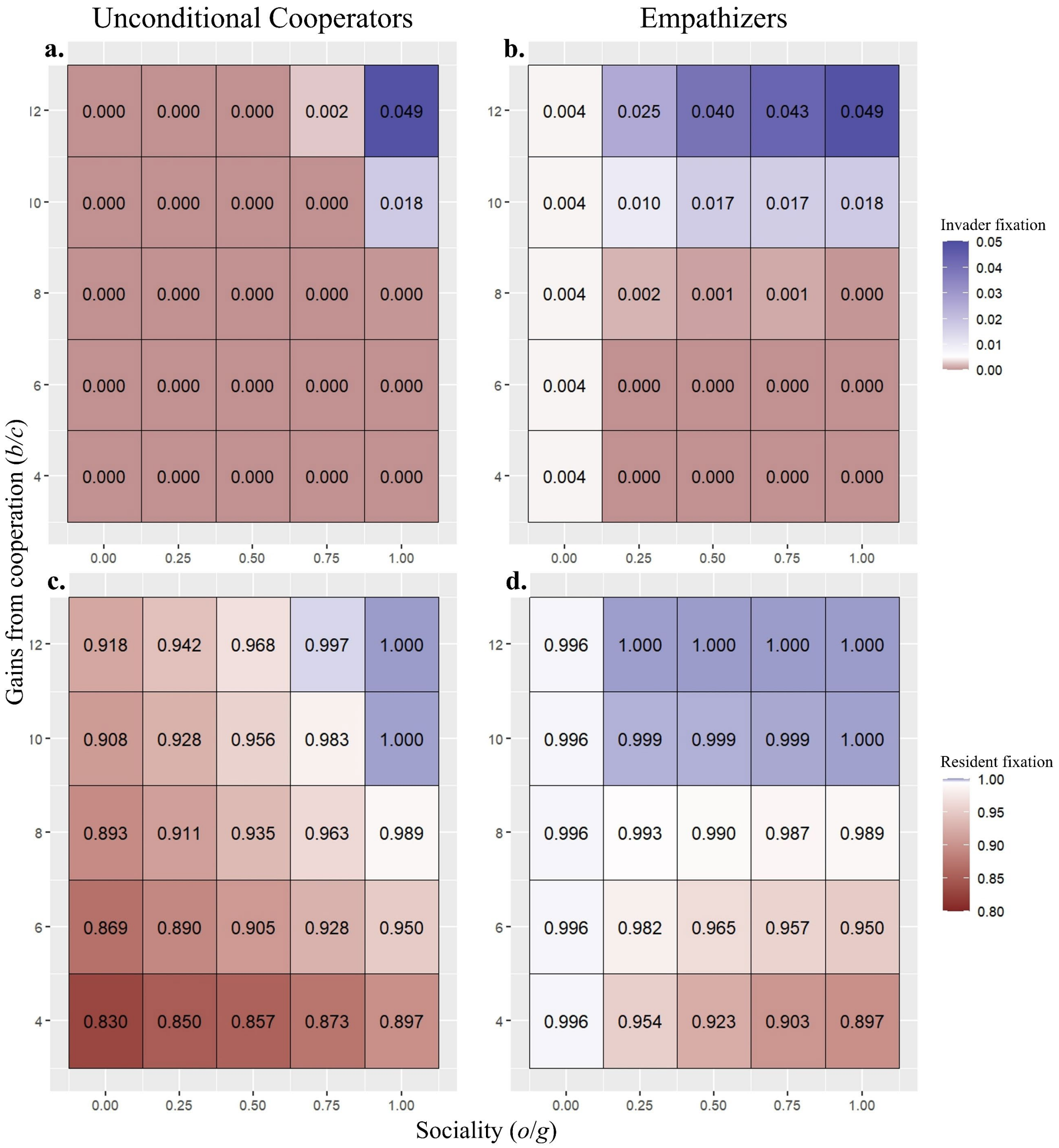


Fig. 4. The impact of social context on the evolution of cooperation. $N = 200$, $g = s = 8$, $w = 0.1$, $c = 1$, 8000 simulations per estimate. Unconditional cooperation relies on high sociality and gains from cooperation to evolve, but discriminatory empathy can persist and spread even when sociality is low as long as the gains from cooperation are sufficiently large. **a, b,** The fixation probability estimates for cooperators invading a population of defectors (viability). **c, d,** The fixation probability estimates for resident cooperators fending off invading defectors (stability). **a, c,** The estimates for unconditional cooperators. **b, d,** The estimates for empathizers who cooperate only with social neighbors. If a cell is blue (red), the cooperators are selected for (against).

3.2. Analytical results

In their 2007 paper, Ohtsuki, Nowak, and Pacheco apply the technique of pair approximation to an evolutionary game model on a multilayer network and find that unconditional cooperation is favored by selection when $\frac{b}{c} > \frac{g \cdot s}{o}$. I apply this technique to replicate this rule for the evolution of unconditional cooperation and to derive a similar rule for the evolution of empathy: $b/c > s$. This analytical result aligns with the numerical results. Both suggest that discriminating on the basis of social distance mitigates the damage low sociality inflicts on the evolutionary feasibility of cooperation.

Pair approximation comprises a method for approximating evolutionary dynamics and equilibrium conditions on networks by focusing on the dynamics as they occur among node pairs (Baalen, 2000). This abstraction away from the complex geometry of the full network makes possible the derivation of differential equations that describe the essential dynamics of the model. In the context of the model in this paper, pair approximation shows that when selection strength is weak ($w \ll 1$), the local density of strategies on the network equilibrates faster than the global density of strategies. The strategy with the higher fitness in this local strategy density equilibrium is likely to prevail in the simulation.

Let p_A be the prevalence of strategy A in the entire population, and let $q_{A|B}^{(*)}$ be the probability that any neighbor on layer $*$ of a B-strategy node is an A-strategy node (i.e. $q_{A|B}^{(*)}$ represents the local density of A around B in the $*$ layer). Then, in a population of competing A and B strategies, the following identities always hold:

$$q_{A|A}^{(G-O)} = q_{A|B}^{(G-O)} = p_A \tag{4}$$

$$q_{*|*}^{(O)} = q_{*|*}^{(S)} \tag{5}$$

On the network of edges belonging only to the economic layer (*G* – *O*), the probability of a neighbor of either strategy being an A player is equal to the global prevalence of that strategy. This is because these non-social ties are unrelated to strategy transmission, so no assortment arises on *G* – *O*. The local strategy densities on the social (*S*) and overlap (*O*) layers are identical since the overlap layer is a subset of the social layer.

In the case of weak selection, $q_{A|A}^{(S)}$ equilibrates before p_A does. This local strategy density equilibrium approximates the state of the population in the simulations while $p_A \in \left(\frac{1}{N}, 1-\frac{1}{N}\right)$. In this equilibrium, the following relationship holds (Ohtsuki et al., 2006, Supp. Eq. 17):

$$q_{A|A}^{(S)} - q_{A|B}^{(S)} = \frac{1}{s-1} \tag{6}$$

In words, A nodes are more likely than B nodes to be social neighbors with other A nodes. This strategy homophily (i.e. clustering of strategies) arises on the social layer because social neighbors may, at some point, have acquired their current strategies from one another. The mathematical derivation can be found in Appendix D.

At this point, rather than estimating and comparing fixation probabilities as Ohtsuki and his coauthors do (2006, 2007), I employ a shortcut I developed: I compare the fitness of the strategies competing for emulation in the local strategy density equilibrium. In Appendix D, I present a proof of the validity of this approach.

To compare the fitness of two strategies competing for emulation in this equilibrium, I must compare their payoffs, and to compare their payoffs, I must compare their expected economic neighborhoods. But since $q_{A|A}^{(G-O)} = q_{A|B}^{(G-O)} = p_A$, the difference in economic neighborhoods is the same as the difference in overlap neighborhoods. Let A_A be the number of A nodes in the overlap neighborhood of an A node and A_B be the number of A nodes in the overlap neighborhood of a B node. From the local strategy density equilibrium relationship (Appendix D), I find that

$$E(A_A - A_B) = \frac{o}{s} \tag{7}$$

Thus, the expected difference in total round payoffs between an unconditional cooperator (*C*) and a defector (*D*) competing for emulation is given by

$$E(\Pi_C - \Pi_D) = (b \cdot C_C - c \cdot g) - (b \cdot C_D) = \frac{b \cdot o}{s} - c \cdot g \tag{8}$$

which is positive when

$$\frac{b}{c} > \frac{g \cdot s}{o} \tag{9}$$

This inequality—originally derived by Ohtsuki et al. (2007)—specifies the conditions of sociality and gains from cooperation under which unconditional cooperation is selected for.

In the case of empathizers who cooperate only with social neighbors ($x = 1$), the expected difference in total round payoffs between an empathizer (*M*) and a defector (*D*) competing for emulation is given by

$$E(\Pi_M - \Pi_D) = (b \cdot M_M - c \cdot o) - (b \cdot M_D) = \frac{b \cdot o}{s} - c \cdot o \tag{10}$$

which is positive when

$$\frac{b}{c} > s \tag{11}$$

This inequality specifies that empathy is selected for as long as the gains from cooperation exceed the degree of the social layer.

The numerical estimates from the simulations corroborate the analytical rules for both unconditional cooperators and empathizers. As shown in Figure 4, in the simulations where $g = s = 8$, unconditional cooperators are only selected for in the upper right corner where $b/c > 64/o$ whereas empathizers are selected for whenever $b/c > 8$.

The evolutionary success of any strategy in the simulations depends firstly on its capacity to duplicate before being eliminated (Fig. C.2) and secondly on its relative fitness in the local strategy density equilibrium. Under weak selection, cooperators invading a population of defectors may survive to the local density equilibrium; then, if the network and game conditions satisfy the relevant analytical rule, the invading cooperators reap a higher average payoff and are therefore likely to spread. Similarly, under weak selection, defectors invading a population of cooperators may survive to the local density equilibrium; then, if the network and game conditions satisfy the relevant analytical rule, the resident cooperators reap a higher average payoff and are therefore likely to fend off the defectors.

4. Discussion

In this paper, I developed an evolutionary game model on a multilayer network to explore the impact of social context on the evolution of cooperation. The main results are intuitive: when we interact with both familiars and strangers, it is adaptive to cooperate with the former but not the latter. Low sociality impedes the evolution of unconditional cooperation but not the evolution of socially-contingent cooperation—i.e. empathy. However, if empathy spreads to the whole population (i.e. fixates), the prevalence of cooperative behavior depends directly on the degree of sociality.

Empathy, as I have defined it, combines the evolutionary tactics of discrimination and assortment to concentrate the benefits of cooperation among cooperators: empathizers cooperate with those who, due to the strategy transmission process, are more likely than chance to also be empathizers. The limiting factor then becomes the degree of the social layer which dictates the “relatedness” of overlap neighbors—i.e. the likelihood they share the same strategy. This interpretation is embodied in the analytical rule for the evolution of empathy, which can be rewritten as $(1/s)b > c$, such that $1/s$ corresponds with the relatedness coefficient of Hamilton’s rule (Bowles & Gintis, 2011; Hamilton, 1964; Smaldino, 2023).

These results illustrate a basic logic underlying the evolution of empathy—a logic that does not depend on the model’s assumptions of static, homogeneous networks, weak selection, or death-birth updating. If behaviors spread among familiars (via biological reproduction or social learning), then familiarity functions as a proxy for behavioral relatedness. Thus, conditioning cooperation on familiarity concentrates the benefits of cooperation among fellow cooperators, allowing this behavior to persist and spread.

Studies of empathy and socially-contingent cooperation have most often explained the evolution of this form of conditional cooperation with reference to Hamilton’s theory of kin selection (Hamilton, 1964) and Trivers’ theory of reciprocal altruism (Trivers, 1971), since social proximity may coincide with genetic relatedness and interaction duration (Batson, 2010; Binmore, 2005; de Waal, 2009; de Waal, 2008; de Waal & Preston, 2017; Osiński, 2009; Rachlin & Jones, 2008). If we interpret the dynamics and results of the model in the context of genetic evolution, then the social layer, where the strategies spread, is a kin network, and social distance corresponds to genetic relatedness. The strategy of empathy in this case is identical to discriminatory kin-based altruism: cooperate with close kin; defect with others. And this strategy is an evolutionary improvement over unconditional cooperation when individuals interact with both kin and non-kin. If, instead, we interpret the dynamics and results of the model in terms of cultural evolution, then social distance corresponds to cultural rather than genetic relatedness, and the evolutionary mechanism illustrated is more closely related to parochial altruism (Bowles & Gintis, 2011) than kin selection. Both genetic and cultural interpretations of the model align with the evidence that relational closeness correlates with cooperation between kith and between kin (Rachlin & Jones, 2008). These dual interpretations also align with the evidence that individuals' cooperative proclivities are products of both heredity and socialization (Eisenberg et al., 2013; Kosse et al., 2020).

The model suggests that modern human ecologies may impact cooperative behavior through two distinct channels: first, through the context-dependent expression of our evolutionary inheritance, and second, through cultural evolution itself. Since genetic evolution occurs at a much slower rate than cultural evolution (due to the different means of transmission), our genetic inheritance may reflect ancient human ecologies while our cultural inheritance may reflect modern human ecologies (Boyd & Richerson, 2005). Consequently, our innate behaviors may be maladaptive to large-scale, anonymous market societies. Whereas empathy may effectively foster cooperation in big game hunts and childcare, it is poorly suited to address climate change and international conflict. The effects of social context on cultural evolution may also hinder the resolution of modern cooperation problems. The results of the model suggest that asocial human ecologies could erode the evolutionary feasibility of unconditional cooperation and select for non-cooperative behavior (either in the form of unconditional defection or vestigial empathy).

The insights the model provides into the behavioral ecology of cooperation suggest cooperation problems might be solved either by altering individual incentives to align self-interest with collective welfare or by changing the social context to better select for and provoke cooperative behavior. The most effective solution depends on the nature of the specific cooperation problem. If the context is unavoidably asocial, the incentive approach might be justified by the reality of self-interested actors. This case is exemplified by a carbon tax implemented to mitigate climate change. Here, the tax brings the cost individuals impose on the global population back into their own self-interested calculation, thus relieving the need for any internally motivated cooperation.

In other contexts, it may be possible to increase the sociality of interactions, thus provoking more cooperation. This case is exemplified by the management of local resource commons (Dietz et al., 2003; Ostrom, 2009; Pretty, 2003). Here, the delegation of commons governance to local communities that depend on the resource may function well on the basis of empathy without any top-down policy solution tailored to self-interested actors (Bowles & Gintis, 2007; Ostrom, 1990).

More broadly, the model developed here exemplifies how the mathematical tools of evolutionary game theory and multilayer networks lend themselves to modelling the evolution of context-dependent behaviors. Such theoretical models provide important insights into the nuance and dynamism of human behavioral ecology (Gonzalez & Rice, 2024), which could prove key to navigating the complex social and environmental problems we face today (Schill et al., 2019).

Code availability

The code for running the evolutionary game simulations and replicating the figures can be found in Supplementary Materials.

Declarations of Competing Interest

None.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Multilayer network generation

The Watts-Strogatz model (Watts & Strogatz, 1998) specifies an algorithm capable of generating networks with small-world properties (i.e. small distances and high transitivity). By modifying the rewiring method to preserve degree distribution, this same algorithm can produce random regular networks as follows:

1. Construct a circulant network with the desired even degree.
2. Arrange the edges in a list and for each edge, designate one adjacent node as the "tail" and the other as the "head."
3. Rewire each edge in the list by selecting an edge at random and swapping their "tail" nodes (adding as a constraint the condition that no self-loops or duplicate edges are created). If a random subset of the edges are rewired such that some of the initial circulant structure remains, then the resulting network will have the small-world properties of short distances and high transitivity. The more edges rewired, the lower the transitivity of the resulting network.

To extend this algorithm to generate a multilayer network with the given parameters of *N, s, g*, and o, (where *s, g*, and *o* are even and $o \leq \min\,(g, s)$), I begin with a circulant network of degree $s + g - o$. After rewiring, this network ($S' \cup G'$) will become the union of the social and economic layers ($S \cup G$). I partition the edges of this union circulant network into three disjoint subsets: the social sans overlap layer ($S' - O'$), the economic sans overlap layer ($G' - O'$), and the overlap layer (O'). I then rewire the networks to obtain $S - O = R(S' - O')$, $G - O = R(G' - O')$, and $O = R(O')$, swapping only edges within the same subset while also avoiding the creation duplicate edges in the superset union network. The original edge is included in the set of potential rewiring pairs, so with the same probability as rewiring with any other given edge, the edge is not rewired. Also, if no pair rewiring is possible that does not create duplicate edges, then the original edge is kept. This process is illustrated in Figure 1 for *N* = 20, *s* = *g* = 4, and *o* = 2. When all edges have been rewired, the social layer is obtained by adding the rewired social sans overlap edge set to the rewired overlap edge set ($S = (S - O) + O$), and the economic layer is obtained by adding the rewired economic sans overlap edge set to the rewired overlap edge set ($G = (G - O) + O$).

This algorithm can be modified to generate a three-layer multilayer network with regular, controlled overlap between each layer. In this case, the edge set of the union circulant network must be partitioned into seven disjoint subsets, one for layer 1 alone, one for layer 2 alone, one for layer 3 alone, one for layer 1 and 2 without layer 3, one for layer 1 and 3 without layer 2, one for layer 2 and 3 without layer 1, and one for the overlap of all three layers, each of which must be rewired separately and without creating duplicate edges in the edge superset. Accordingly, with more than three layers, the process becomes even more unwieldy.

Appendix B. Utility function specification

Let i and j be nodes interacting in the game described in Figure 2. If i is selfish, she will defect since $b > b - c$ and $0 > -c$. If i cares sufficiently about j, she will cooperate since $b > c$. These motivations can be expressed in terms of a utility function incorporating the payoffs of i and j: $u_i(\pi_{i,j}, \pi_{j,i})$. To operationalize empathy, I specify the other-regarding component of i's utility function as a hyperbolic function of the social distance between i and j in accordance with the empirical findings of the social discounting literature (Jones, 2022; Jones & Rachlin, 2006):

$$u_i = \pi_{i,j} + \frac{a_i \pi_{j,i}}{1 + \delta_i\left(d_{i,j} - 1\right)} \qquad (B.1)$$

where $\pi_{i,j}$ represents the game payoff of player i interacting with player j, $d_{i,j}$ is the social distance between i and j, a_i represents how much i values the payoffs of a social neighbor relative to her own, and δ_i is the social discount factor which specifies how quickly empathy decays with social distance.

Individual i will thus cooperate with j as long as

$$\frac{a_i b}{1 + \delta_i\left(d_{i,j} - 1\right)} > c \qquad (B.2)$$

Individuals for whom $a_i < c/b$ will defect unconditionally. Individuals for whom $a_i > c/b$ and $\delta_i = 0$ will cooperate unconditionally. And individuals for whom $a_i > \frac{c}{b}$ and $\delta_i > 0$ will cooperate conditional on sufficient social proximity. Specifically, these empathizers cooperate with all j for whom $d_{i,j} < \frac{a_i b - c}{c\delta_i} + 1$. Thus, this utility function produces a behavioral strategy that can be alternatively defined by the single discrete empathic range variable, $x_i \in \mathbb{N}$, indicating that i will cooperate with all j for whom $d_{i,j} \leq x_i$. (I define x_i based on the weak inequality for intuitiveness of interpretation.) Figure B.1 illustrates graphically this derivation of x_i from b, c, a_i, and δ_i. The strategy defined by x_i could also be obtained from alternative utility function specifications. For example, the specification $u_i = \pi_i + \frac{a_i \pi_j}{d_{i,j} - 1}$ (which reflects Goeree et al.'s "1/d Law of Giving" (2010)) yields the same discrete empathic range trait x_i. The evolutionary dynamics in this paper, however, depend only on the empathic range trait, so my results are agnostic with regard to the choice of utility function specification.

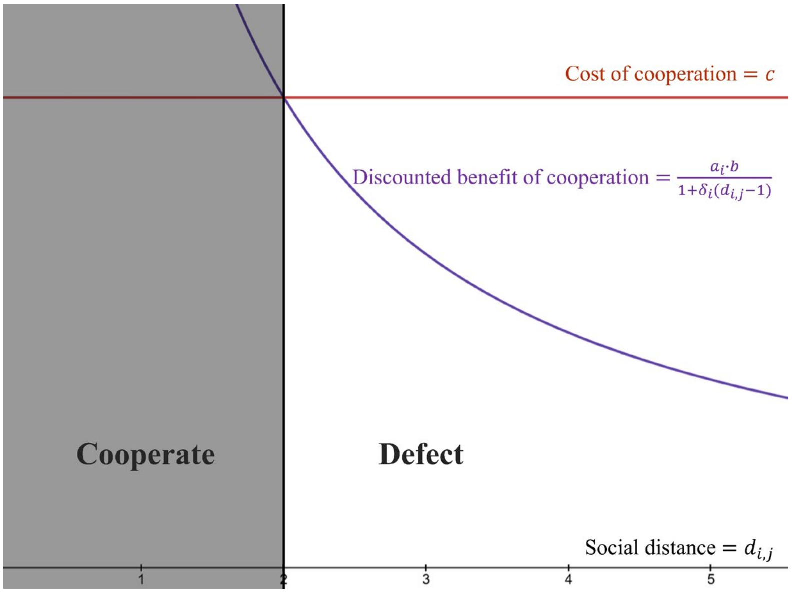


Fig. B.1. Social discounting utility function specification. If the socially discounted benefit of cooperation (which declines with social distance) exceeds the cost of cooperation, the empathizer will cooperate. This means that given a_i and δ_i, i will cooperate with every economic neighbor within a certain distance in the social layer and defect with all others. In the example shown in this figure, i will cooperate only with economic neighbors who are 1 or 2 edges away in the social layer. This behavioral strategy can be represented as a discrete empathic range: $x_i = 2$.

Appendix C. Selection strength

C.1. Conceptual discussion

In evolutionary models, weak selection allows a single cooperator to invade a population of defectors. Isolated cooperators necessarily have a lower payoff than the defectors they are competing against, so they must initially spread stochastically against the gradient of selection to form a cluster of cooperators, which can then compete with defectors in terms of fitness. How realistic, then, is this assumption?

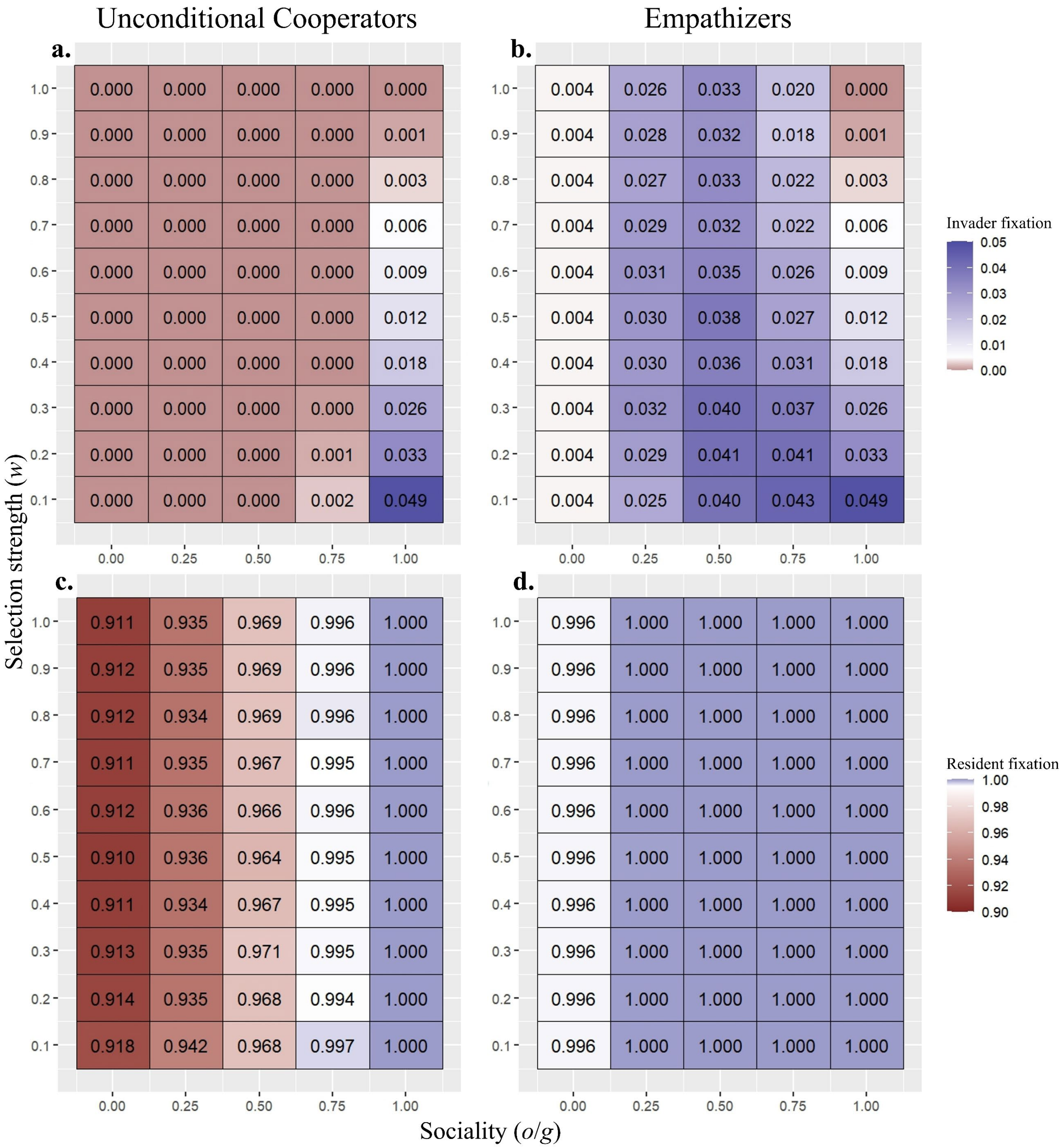
Selection strength likely differs across the realms of cultural and genetic evolution. In the case of genetic evolution, fitness refers to an organism's propensity to survive and reproduce and thus pass on their genetic traits. Here, fitness is a product of a multitude of interactions of which the trait-relevant game is only one, so weak selection is justified.

In the case of cultural evolution, since enculturation is not limited to vertical transmission following biological reproduction, fitness applies to the traits carried by the organism rather than the organism itself. The propensity of a socially learned behavior to persist and spread need not be directly correlated to the overall material success of the trait-bearer. The latter could be the case if the spread of the cultural trait is achieved by means of power and wealth (e.g. control of the media, intentional indoctrination, conquest and replacement). But cultural traits could also be selected for by voluntary preferential adoption, in which case, the fitness of the trait in the individual is likely a function of their payoffs from trait-relevant interactions.

C.2. Numerical results

Figure C.1 reports the fixation probability estimates for unconditional cooperators and empathizers $(x = 1)$ competing against defectors under various conditions of sociality (o/g) and selection strength (w). I hold the gains from cooperation at a constant level sufficient to promote some evolutionary success of cooperation $(b/c = 12)$. The viability fixation probability estimates decrease with selection strength for unconditional cooperators and for empathizers when sociality is high (in which case they act like unconditional cooperators). On the other hand, no clear relationship exists between selection strength and the stability fixation probabilities estimates for resident cooperators (either unconditional cooperators or empathizers).

The detrimental effect of selection strength on the evolutionary viability of invading cooperators can be explained with reference to the probability of those invaders forming a cluster. As selection strength increases, invading cooperators are less likely to duplicate before they are eliminated, and for empathizers, this effect intensifies with increasing sociality (Fig. C.2a). The effect of selection strength on the duplication probability of unconditional cooperators is not mediated by sociality since their behavior is not conditioned on it. In the case of the evolutionary stability of resident cooperators, however, cluster formation is not a challenge since the initial conditions are characterized by one massive cooperator cluster. Higher selection strength still increases the chance of the invading defector duplicating, but the effect is miniscule compared to the case of cooperator invasion (Fig. C.2b). In both cases, when payoffs cease to matter (due to a selection strength of zero or when empathizers are indistinguishable from defectors), the probability that an invading strategy duplicates before it can be eliminated becomes 0.5.



a.

w \ o/g	0.00	0.25	0.50	0.75	1.00
1.0	0.000	0.000	0.000	0.000	0.000
0.9	0.000	0.000	0.000	0.000	0.001
0.8	0.000	0.000	0.000	0.000	0.003
0.7	0.000	0.000	0.000	0.000	0.006
0.6	0.000	0.000	0.000	0.000	0.009
0.5	0.000	0.000	0.000	0.000	0.012
0.4	0.000	0.000	0.000	0.000	0.018
0.3	0.000	0.000	0.000	0.000	0.026
0.2	0.000	0.000	0.000	0.001	0.033
0.1	0.000	0.000	0.000	0.002	0.049

b.

w \ o/g	0.00	0.25	0.50	0.75	1.00
1.0	0.004	0.026	0.033	0.020	0.000
0.9	0.004	0.028	0.032	0.018	0.001
0.8	0.004	0.027	0.033	0.022	0.003
0.7	0.004	0.029	0.032	0.022	0.006
0.6	0.004	0.031	0.035	0.026	0.009
0.5	0.004	0.030	0.038	0.027	0.012
0.4	0.004	0.030	0.036	0.031	0.018
0.3	0.004	0.032	0.040	0.037	0.026
0.2	0.004	0.029	0.041	0.041	0.033
0.1	0.004	0.025	0.040	0.043	0.049

c.

w \ o/g	0.00	0.25	0.50	0.75	1.00
1.0	0.911	0.935	0.969	0.996	1.000
0.9	0.912	0.935	0.969	0.996	1.000
0.8	0.912	0.934	0.969	0.996	1.000
0.7	0.911	0.935	0.967	0.995	1.000
0.6	0.912	0.936	0.966	0.996	1.000
0.5	0.910	0.936	0.964	0.995	1.000
0.4	0.911	0.934	0.967	0.995	1.000
0.3	0.913	0.935	0.971	0.995	1.000
0.2	0.914	0.935	0.968	0.994	1.000
0.1	0.918	0.942	0.968	0.997	1.000

d.

w \ o/g	0.00	0.25	0.50	0.75	1.00
1.0	0.996	1.000	1.000	1.000	1.000
0.9	0.996	1.000	1.000	1.000	1.000
0.8	0.996	1.000	1.000	1.000	1.000
0.7	0.996	1.000	1.000	1.000	1.000
0.6	0.996	1.000	1.000	1.000	1.000
0.5	0.996	1.000	1.000	1.000	1.000
0.4	0.996	1.000	1.000	1.000	1.000
0.3	0.996	1.000	1.000	1.000	1.000
0.2	0.996	1.000	1.000	1.000	1.000
0.1	0.996	1.000	1.000	1.000	1.000

Fig. C.1. The impact of selection strength on the evolution of cooperation. $N = 200$, $g = s = 8$, $b = 12$, $c = 1$, 8000 simulations per estimate. In the case of cooperator invasion (**a.** and **b.**), the evolutionary viability of unconditional cooperation appears to decrease with increasing selection strength. This effect also holds for empathizers when sociality is high. In the case of defector invasion (**c.** and **d.**), the evolutionary stability of cooperation appears to be unrelated to selection strength. If a cell is blue (red), the cooperators are selected for (against).

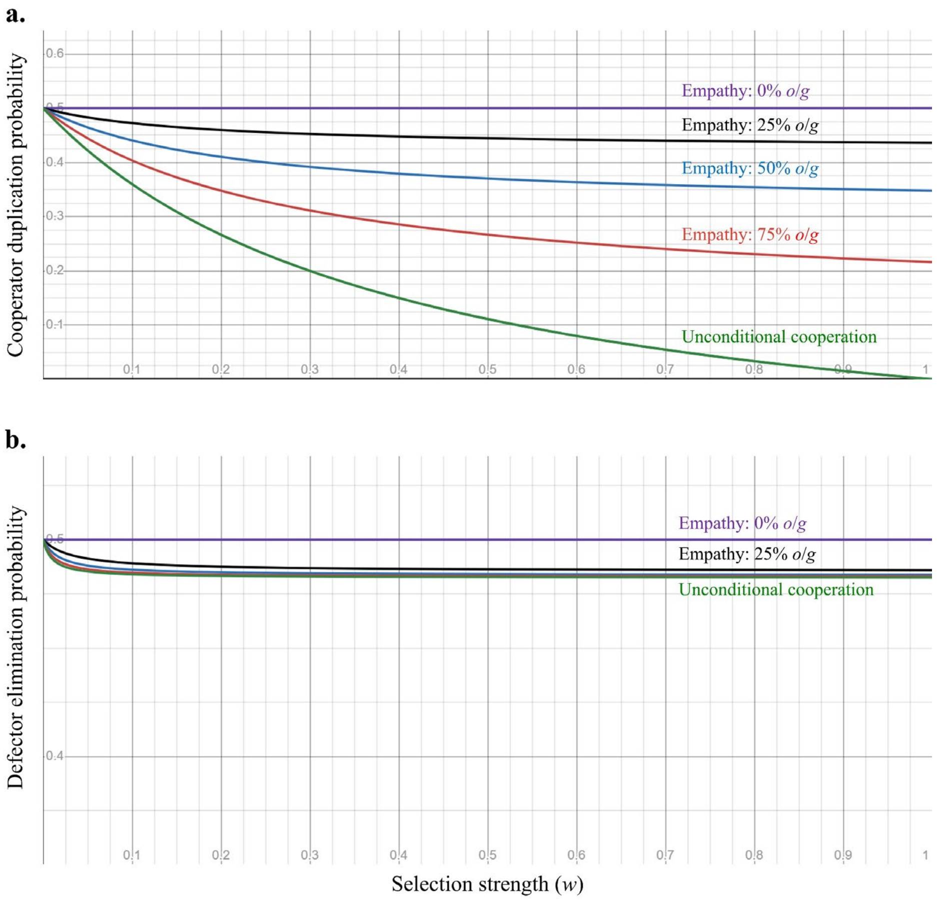


Fig. C.2. The impact of selection strength on initial trait proliferation. $N = \infty$, $h = g = 8$, $b = 12$, $c = 1$. **a.** As selection strength increases, invading cooperators are less likely to duplicate before they are eliminated, and for empathizers, this effect intensifies with increasing sociality. **b.** In the case of resident cooperators, the probability that the invading defector is eliminated before it can duplicate also decreases with selection strength, but the effect is miniscule. In both cases, when payoffs cease to matter (due to a selection strength of zero or when empathizers are indistinguishable from defectors), the probability that an invading strategy duplicates before it can be eliminated becomes 0.5.

C.3. Calculating the probability of invader duplication

The probability that event X occurs before event Y, given that X occurs each period with probability x and Y with the probability y, is $\frac{x}{x+y}$. Using the fitness equation and adding c to all payoffs in Figure 2 to avoid negative probabilities (as is done in the simulations), we can calculate these duplication probabilities. For simplicity, I assume *N* to be arbitrarily large (if I keep a finite N in the calculations, the results don't change much, but the functions become even more cumbersome). For unconditional cooperators, the probability in each period that the invader duplicates is

$$\left(\frac{s}{N}\right)\frac{f_C}{f_C + (s-1)f_D} = \left(\frac{s}{N}\right)\frac{1 - w}{s(1-w) + (s-1)wgc} \qquad (C.1)$$

And the probability in each period that the invader is eliminated is $\frac{1}{N}$. The probability then that the unconditional cooperator duplicates before it can be eliminated is

$$\frac{sf_C}{(s+1)f_C + (s-1)f_D} = \frac{s(1-w)}{2s(1-w) + (s-1)wgc} \qquad (C.2)$$

For empathizers, the probability in each period that the invader duplicates is

$$\left(\frac{s}{N}\right)\frac{f_E}{f_E + (s-1)f_D} = \left(\frac{s}{N}\right)\frac{1 - w + w(g-o)c}{s(1-w+wgc) - woc} \qquad (C.3)$$

And the probability in each period that the invader is eliminated is $\frac{1}{N}$. The probability then that the empathizer duplicates before it can be eliminated is

$$\frac{sf_E}{(s+1)f_E + (s-1)f_D} = \frac{s(1-w+w(g-o)c)}{2s(1 - w + wgc) - woc(s+1)} \qquad (C.4)$$

These duplication probabilities are plotted in Figure C.2a. Note that if there is no difference in fitness (if the payoffs are equal or if w = 0), the duplication probabilities simplify to

$$\frac{sf}{(s+1)f + (s-1)f} = \frac{1}{2} \qquad (C.5)$$

Note also that

$$\frac{d}{do}\left(\frac{sf_E}{(s+1)f_E + (s-1)f_D}\right) = \frac{-swc(s-1)f_D}{[(s+1)f_E + (s-1)f_D]^2} < 0 \qquad (C.6)$$

i.e. duplication probability declines with overlap.

Using the same tools and assumptions, we can calculate probability that invading defectors are eliminated before they can duplicate. In the case of unconditional cooperators, the probability in each period that the defector is eliminated is $1/N$, and the probability in each period that the defector duplicates is

$$\left(\frac{s}{N}\right)\frac{1 - w + wg(b+c)}{s(1 - w) + wg(b+c) + (s-1)wbg} \qquad (C.7)$$

The probability then that the defector is eliminated before it can duplicate is

$$1-\frac{s\big(1-w+wg(b+c)\big)}{2s(1-w)+(s+1)wg(b+c)+(s-1)wbg} \tag{C.8}$$

In the case of empathizers, the probability in each period that the defector is eliminated is $\frac{1}{N}$, and the probability in each period that the invader duplicates is

$$\left(\frac{s}{N}\right)\frac{1-w+w(ob+gc)}{s(1\,-\,w+wob)+wc(sg-so+o)} \tag{C.9}$$

The probability then that the defector is eliminated before it can duplicate is

$$1-\frac{s\big(1-w+w(ob+gc)\big)}{2s(1\,-\,w+wob)+wc(2sg-so+o)} \tag{C.10}$$

These duplication probabilities are plotted in Figure C.2b.

Appendix D. Derivation of the analytical rule

In this section, I derive the local density equilibrium conditions and formulate a proof demonstrating the logic of the shortcut I employ to obtain the analytical rule for the evolution of empathy.

As before, let p_A be the prevalence of strategy A in the entire population, and let $q_{A|B}^{(*)}$ be the probability that any $*$ neighbor of a B-strategy node is an A-strategy node. Also, let $p_{AA}^{(*)}$ be the prevalence of AA strategy pairs among all $Nk/2$ edges in any layer $*$ of k degree. Then, in a population of competing A and B strategies, the following identities always hold:

$$p_A+p_B=1 \tag{D.1}$$

$$p_{AA}^{(*)}+p_{AB}^{(*)}+p_{BB}^{(*)}=1 \tag{D.2}$$

$$p_{AB}^{(*)}=p_{BA}^{(*)} \tag{D.3}$$

$$q_{A|B}^{(*)}=\frac{p_{AB}^{(*)}}{p_B} \tag{D.4}$$

$$q_{A|A}^{(*)} = 1 - q_{B|A}^{(*)} \qquad (D.5)$$

$$q_{A|A}^{(G-O)} = q_{A|B}^{(G-O)} = p_A \qquad (D.6)$$

$$q_{*|*}^{(O)} = q_{*|*}^{(S)} \qquad (D.7)$$

The intuition behind equation (D.4) is that the prevalence of AB pairs in the vicinity of a B multiplied by the global prevalence of B is equal to the global prevalence of AB pairs—multiplying $q_{A|B}^{(*)}$ by p_B undoes the conditionality of the prevalence. As discussed in the main text, the last two identities are the consequence of the separation of the interaction (economic) and updating (social) layers.

The probability that the number of A nodes increases in any given round is equal to the probability that a B node is chosen to update and becomes an A:

$$P\left(\Delta p_A = \frac{1}{N}\right) = p_B \sum_{s_A=0}^{s} \frac{s!}{s_A!\, s_B!} {q_{A|B}^{(S)}}^{s_A} {q_{B|B}^{(S)}}^{s_B} \frac{s_A f_A}{s_A f_A + s_B f_B} \qquad (D.8)$$

where s_A denotes the number of social neighbors of the updater with strategy A (so $s_A + s_B = s$), and f_A denotes the expected fitness of an A. The intuition behind this expression is this: p_B is the probability a B node is chosen to update; $\frac{s!}{s_A! s_B!} {q_{A|B}^{(S)}}^{s_A} {q_{B|B}^{(S)}}^{s_B}$ is the probability of any given B node having s_A social neighbors with strategy A; and $\frac{s_A f_A}{s_A f_A + s_B f_B}$ is the probability that one of the s_A social neighbors is chosen to be emulated. The sum is taken across these probabilities for all possible neighborhood configurations and their prevalence. Consequently, the probability that the number of social AA pairs increases by s_A is given by

$$P\left(\Delta p_{AA}^{(S)} = \frac{2s_A}{sN}\right) = p_B \frac{s!}{s_A!\, s_B!} {q_{A|B}^{(S)}}^{s_A} {q_{B|B}^{(S)}}^{s_B} \frac{s_A f_A}{s_A f_A + s_B f_B} \qquad (D.9)$$

since updating a B node with s_A social neighbors with strategy A results in s_A new AA pairs out of $sN/2$ total social edges.

Now, let each round of each simulation be considered as one unit of time. Then,

$$\frac{d}{dt}(p_A) = \frac{1}{N} \cdot P\left(\Delta p_A = \frac{1}{N}\right) - \frac{1}{N} \cdot P\left(\Delta p_A = -\frac{1}{N}\right) \qquad (D.10)$$

and

$$\frac{d}{dt}\left(p_{AA}^{(S)}\right) = \sum_{s_A=0}^{s} \frac{2s_A}{sN}\left[P\left(\Delta p_{AA}^{(S)} = \frac{2s_A}{sN}\right) - P\left(\Delta p_{AA}^{(S)} = -\frac{2s_A}{sN}\right)\right] \qquad (D.11)$$

In the case of weak selection ($w \ll 1$), $p_{AA}^{(S)}$ changes more quickly than p_A since

$$\lim_{w \to 0}\left(\frac{d}{dt}(p_A)\right) = 0 \qquad (D.12)$$

but

$$\lim_{w \to 0}\left(\frac{d}{dt}\left(p_{AA}^{(S)}\right)\right) = \frac{2}{sN} p_{AB}^{(S)}\left[1 + (s-1)\left(q_{A|B}^{(S)} - q_{A|A}^{(S)}\right)\right] > 0 \qquad (D.13)$$

This is to say that when there is no selection occurring and therefore no differences in fitness, A nodes are just as likely to become B nodes as B nodes are to become A nodes, so the expected rate of change of the global prevalence of strategies is 0. But strategy updating even without selection still leads to a clustering of strategies in the social layer of the network since in any pair of social neighbors, one may have adopted its current strategy by emulating the other. A more detailed derivation of these equations can be found in the supplementary section of Ohtsuki et al. (2006).

The intuition for equation (D.11) is this: The change in the prevalence of AA social pairs depends on the prevalence of AB social pairs, since both the creation and destruction of AA pairs relies on the existence of a B near an A (to create an AA, an AB pair must become an AA pair, and to destroy an AA pair, a B node must occur in the AA pair's extended neighborhood to be emulated). Thus, the number of AA pairs changes only along the strategy boundaries marked by AB pairs. Now, consider a given AB pair. The neighborhood of the A includes on average $(s-1)q_{A|A}^{(S)}$ A nodes, and the neighborhood of the B includes on average $1 + (s-1)q_{A|B}^{(S)}$ A nodes. If the B updates and emulates an A, then, on average, $1 + (s-1)q_{A|B}^{(S)}$ new AA pairs will be created, and if the A updates and emulates a B, then, on average, $(s-1)q_{A|A}^{(S)}$ AA pairs will be lost. Thus, the rate of change of the number of AA pairs is proportionate to $1 + (s - 1)\left(q_{A|B}^{(S)} - q_{A|A}^{(S)}\right)$.

Since $q_{A|A}^{(S)} = p_{AA}^{(S)}/p_A$, we can also derive the time derivative for $q_{A|A}^{(S)}$ as strength of selection approaches zero:

$$\lim_{w\to 0}\left(\frac{d}{dt}\left(q_{A|A}^{(S)}\right)\right) = \frac{2}{sN}\frac{p_{AB}^{(S)}}{p_A}\left[1 + (s-1)\left(q_{A|B}^{(S)} - q_{A|A}^{(S)}\right)\right] \qquad (D.14)$$

It is due to this difference in the rate of change of p_A and $q_{A|A}^{(S)}$ that, in the case of weak selection, $q_{A|A}^{(S)}$ equilibrates before p_A does. The $q_{A|A}^{(S)}$ equilibrium, defined by $\frac{d}{dt}\left(q_{A|A}^{(S)}\right) = 0$, yields the following relationships (Ohtsuki et al., 2006):

$$q_{A|A}^{(S)} = p_A + \frac{1}{s-1}(1 - p_A) \qquad (D.15)$$

$$q_{B|A}^{(S)} = 1 - q_{A|A}^{(S)} = 1 - p_A - \frac{1}{s-1}(1 - p_A) = p_B - \frac{1}{s-1}p_B \qquad (D.16)$$

$$q_{A|A}^{(S)} - q_{A|B}^{(S)} = p_A + \frac{1}{s-1}(1 - p_A) - p_A + \frac{1}{s-1}p_A = \frac{1}{s-1} \qquad (D.17)$$

Thus far, I have followed the method of Ohtsuki et al. (2006, 2007). At this point, Ohtsuki and his coauthors estimate fixation probabilities and derive their analytical rules by comparing these formulas. In contrast, I derive the rules for the evolution of unconditional cooperation and empathy using a shortcut whereby I compare the expected payoff of the competing strategies at the local strategy density equilibrium.

But why does this work? Consider the probability that an updater emulates an A or a B node:

$$E[\Pr(U \to A)] = p_A\left[\frac{s \cdot q_{A|A}^{(S)} \cdot E[f_A]}{\sum_{N_s(U)} f}\right] + p_B\left[\frac{s \cdot q_{A|B}^{(S)} \cdot E[f_A]}{\sum_{N_s(U)} f}\right] \qquad (D.18)$$

$$E[\Pr(U \to B)] = p_A\left[\frac{s \cdot q_{B|A}^{(S)} \cdot E[f_B]}{\sum_{N_s(U)} f}\right] + p_B\left[\frac{s \cdot q_{B|B}^{(S)} \cdot E[f_B]}{\sum_{N_s(U)} f}\right] \qquad (D.19)$$

so

$$
\begin{aligned}
E[\Pr(U \to A)] > E[\Pr(U \to A)]& \\
\Leftrightarrow p_A \left[\frac{s \cdot q_{A|A}^{(S)} \cdot E[f_A]}{\sum_{N_S(U)} f}\right] + p_B \left[\frac{s \cdot q_{A|B}^{(S)} \cdot E[f_A]}{\sum_{N_S(U)} f}\right]& \\
> p_A \left[\frac{s \cdot q_{B|A}^{(S)} \cdot E[f_B]}{\sum_{N_S(U)} f}\right] + p_B \left[\frac{s \cdot q_{B|B}^{(S)} \cdot E[f_B]}{\sum_{N_S(U)} f}\right]& \\
\Leftrightarrow p_A\left(q_{A|A}^{(S)} \cdot E[f_A]\right) + p_B\left(q_{A|B}^{(S)} \cdot E[f_A]\right)& \\
> p_A\left(q_{B|A}^{(S)} \cdot E[f_B]\right) + p_B\left(q_{B|B}^{(S)} \cdot E[f_B]\right)& \qquad (D.20)
\end{aligned}
$$

Using Equations D.15 and D.16, this inequality simplifies to

$$p_A \cdot E[f_A] > p_B \cdot E[f_B] \qquad (D.21)$$

Now, recall that a trait is said to be selected for if it fares better than a neutral mutant. If A is a neutral mutant of B, then $E[f_A] = E[f_B]$, in which case Inequality (D.21) simplifies to

$$p_A > p_B \qquad (D.22)$$

Thus, strategy A fares better than a neutral mutant if

$$p_A \cdot E[f_A] - p_B \cdot E[f_B] > p_A - p_B \qquad (D.23)$$

which is true when

$$
\begin{aligned}
E[f_A] > E[f_B]& \\
\Leftrightarrow E[\Pi_A] > E[\Pi_B]& \qquad (D.24)
\end{aligned}
$$

Interestingly, this rule is also arrived at by comparing $E[\Pr(U \to A)]$ with $E[\Pr(U \to B)]$ when the population is equally split between A and B ($p_A = p_B$)—or by switching to a deterministic updating rule based on the maximum payoff once the local strategy density equilibrium is reached.

Now, to compare the expected payoffs of the two competing strategies, I compare their expected economic neighborhoods in the local strategy density equilibrium. In the non-social economic network ($G - O$), there is no difference in the expected neighborhoods (since assortment only arises on the social network). Therefore, the expected difference in economic neighborhood is equal to the expected difference in the overlap neighborhood. Let A_A be the number of A nodes in the overlap neighborhood of an A node and A_B be the number of A nodes in the overlap neighborhood of a B node. To find $E(A_A - A_B)$ in the local strategy density equilibrium, we must consider eight scenarios arising from the strategy of the updater, *U*, and its position relative to the A and B nodes competing for emulation:

1. *U* is an A player
    1. *U* is in the overlap neighborhood of both A and B (occurs with probability $\left(\frac{o}{s}\right)^2$)

       $$E(A_A - A_B) = (o-1)\left(q_{A|A}^{(O)} - q_{A|B}^{(O)}\right) \tag{D.25}$$

    2. *U* is in the overlap neighborhood of A but not B ($Pr = \frac{o}{s}\left(1 - \frac{o}{s}\right)$)

       $$E(A_A - A_B) = 1 + (o-1)q_{A|A}^{(O)} - oq_{A|B}^{(O)} \tag{D.26}$$

    3. *U* is in the overlap neighborhood of B but not A ($Pr = \frac{o}{s}\left(1 - \frac{o}{s}\right)$)

       $$E(A_A - A_B) = oq_{A|A}^{(O)} - (o-1)q_{A|B}^{(O)} - 1 \tag{D.27}$$

    4. *U* is in the overlap neighborhood of both A and B ($Pr = \left(1 - \frac{o}{s}\right)^2$)

       $$E(A_A - A_B) = o\left(q_{A|A}^{(O)} - q_{A|B}^{(O)}\right) \tag{D.28}$$

2. *U* is a B player
    1. *U* is in the overlap neighborhood of both A and B ($Pr = \left(\frac{o}{s}\right)^2$)

       $$E(A_A - A_B) = (o-1)\left(q_{A|A}^{(O)} - q_{A|B}^{(O)}\right) \tag{D.29}$$

    2. *U* is in the overlap neighborhood of A but not B ($Pr = \frac{o}{s}\left(1 - \frac{o}{s}\right)$)

       $$E(A_A - A_B) = (o-1)q_{A|A}^{(O)} - oq_{A|B}^{(O)} \tag{D.30}$$

    3. *U* is in the overlap neighborhood of B but not A ($Pr = \frac{o}{s}\left(1 - \frac{o}{s}\right)$)

       $$E(A_A - A_B) = oq_{A|A}^{(O)} - (o-1)q_{A|B}^{(O)} \tag{D.31}$$

    4. *U* is in the overlap neighborhood of both A and B ($Pr = \left(1 - \frac{o}{s}\right)^2$)

       $$E(A_A - A_B) = o\left(q_{A|A}^{(O)} - q_{A|B}^{(O)}\right) \tag{D.32}$$

Whether U is an A or a B player, the expected difference in the number of A nodes in the overlap neighborhoods is then

$$
\begin{aligned}
E(A_A - A_B) = {} & \left(\frac{o}{s}\right)^2 (o-1)\left(q_{A|A}^{(O)} - q_{A|B}^{(O)}\right) \\
& + \frac{o}{s}\left(1 - \frac{o}{s}\right)\left((o-1)q_{A|A}^{(O)} - oq_{A|B}^{(O)} + oq_{A|A}^{(O)} - (o-1)q_{A|B}^{(O)}\right) \\
& + \left(1 - \frac{o}{s}\right)^2 o\left(q_{A|A}^{(O)} - q_{A|B}^{(O)}\right) \qquad (D.33)
\end{aligned}
$$

Substituting in $q_{A|A}^{(O)} - q_{A|B}^{(O)} = \frac{1}{s-1}$, this becomes

$$
\begin{aligned}
E(A_A - A_B) = {} & \left(\frac{o}{s}\right)^2 \frac{o-1}{s-1} \\
& + \frac{o}{s}\left(1 - \frac{o}{s}\right)\left(\frac{2o}{s-1} - \left(p_A - \frac{1}{s-1}(1-p_A)\right) + p_A - \frac{1}{s-1}p_A\right) \\
& + \left(1 - \frac{o}{s}\right)^2 \frac{o}{s-1} \\
= {} & \frac{o}{s} \qquad (D.34)
\end{aligned}
$$

Lastly, note that if we separate the social layer (*S*) from a "cultural" trait transmission layer (*L*), the rule for the evolution of empathy becomes $\frac{b}{c} > \frac{o_{gs} \cdot l}{o_{gl}}$, where o_{gs} is the social-economic overlap degree, *l* is the cultural degree, and o_{gl} is the cultural-economic overlap degree. As mentioned above, the multilayer network generation algorithm could be modified to generate a three-layer multilayer network on which such evolutionary game simulations could be run.

References

Akçay, E., & Van Cleve, J. (2021). Internalizing cooperative norms in group-structured populations. *Cooperation and Conflict: The Interaction of Opposites in Shaping Social Behavior, W. Wilczynski, S. Brosnan, Eds.(Cambridge University Press, 2021)*, 26–44.

Allen, B., Lippner, G., Chen, Y.-T., Fotouhi, B., Momeni, N., Yau, S.-T., & Nowak, M. A. (2017). Evolutionary dynamics on any population structure. *Nature*, *544*(7649), Article 7649. https://doi.org/10.1038/nature21723

Allen, B., & Nowak, M. A. (2014). Games on graphs. *EMS Surveys in Mathematical Sciences*, *1*(1), 113–151. https://doi.org/10.4171/emss/3

Andreoni, J., & Rao, J. M. (2011). The power of asking: How communication affects selfishness, empathy, and altruism. *Journal of Public Economics*, *95*(7), 513–520. https://doi.org/10.1016/j.jpubeco.2010.12.008

Apicella, C. L., Marlowe, F. W., Fowler, J. H., & Christakis, N. A. (2012). Social networks and cooperation in hunter-gatherers. *Nature*, *481*(7382), Article 7382. https://doi.org/10.1038/nature10736

Ariely, D., Bracha, A., & Meier, S. (2009). Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. *American Economic Review*, *99*(1), 544–555. https://doi.org/10.1257/aer.99.1.544

Artinger, F., Exadaktylos, F., Koppel, H., & Sääksvuori, L. (2014). In others' shoes: Do individual differences in empathy and theory of mind shape social preferences? *PLOS ONE*, *9*(4), e92844. https://doi.org/10.1371/journal.pone.0092844

Baalen, M. van. (2000). Pair approximations for different spatial geometries. In J. A. J. Metz, R. Law, & U. Dieckmann (Eds.), *The Geometry of Ecological Interactions: Simplifying Spatial Complexity* (pp. 359–387). Cambridge University Press. https://doi.org/10.1017/CBO9780511525537.023

Batson, C. D. (2009). These things called empathy: Eight related but distinct phenomena. In J. Decety & W. Ickes (Eds.), *The Social Neuroscience of Empathy* (p. 0). The MIT Press. https://doi.org/10.7551/mitpress/9780262012973.003.0002

Batson, C. D. (2010). *Empathy-induced altruistic motivation.* https://psycnet.apa.org/record/2009-08941-001

Batson, C. D., & Ahmad, N. (2001). Empathy-induced altruism in a prisoner's dilemma II: What if the target of empathy has defected? *European Journal of Social Psychology*, *31*(1), 25–36. https://doi.org/10.1002/ejsp.26

Batson, C. D., Duncan, B. D., Ackerman, P., Buckley, T., & Birch, K. (1981). Is empathic emotion a source of altruistic motivation? *Journal of Personality and Social Psychology*, *40*(2), 290–302. https://doi.org/10.1037/0022-3514.40.2.290

Batson, C. D., & Moran, T. (1999). Empathy-induced altruism in a prisoner's dilemma. *European Journal of Social Psychology*, *29*(7), 909–924. https://doi.org/10.1002/(SICI)1099-0992(199911)29:7%253C909::AID-EJSP965%253E3.0.CO;2-L

Bicchieri, C., Dimant, E., Gächter, S., & Nosenzo, D. (2022). Social proximity and the erosion of norm compliance. *Games and Economic Behavior*, *132*, 59–72. https://doi.org/10.1016/j.geb.2021.11.012

Binmore, K. (2005). *Natural justice*. Oxford University Press.

Bochet, O., Page, T., & Putterman, L. (2006). Communication and punishment in voluntary contribution experiments. *Journal of Economic Behavior & Organization*, *60*(1), 11–26. https://doi.org/10.1016/j.jebo.2003.06.006

Bohnet, I., & Frey, B. S. (1999). Social distance and other-regarding behavior in dictator games: Comment. *American Economic Review*, *89*(1), 335–339. https://doi.org/10.1257/aer.89.1.335

Bowles, S., & Gintis, H. (2007). Social capital, moral sentiments, and community governance. In H. Gintis, S. Bowles, R. Boyd, & E. Fehr (Eds.), *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life* (Vol. 109). https://anthrosource.onlinelibrary.wiley.com/doi/10.1525/aa.2007.109.2.380

Bowles, S., & Gintis, H. (2011). *A cooperative species: Human reciprocity and its evolution*. Princeton University Press. https://doi.org/10.1515/9781400838837

Boyd, R., & Richerson, P. J. (1988). *Culture and the evolutionary process*. University of Chicago Press.

Boyd, R., & Richerson, P. J. (2005). *The origin and evolution of cultures*. Oxford University Press.

Brañas-Garza, P., Cobo-Reyes, R., Espinosa, M. P., Jiménez, N., Kovářík, J., & Ponti, G. (2010). Altruism and social integration. *Games and Economic Behavior*, *69*(2), 249–257. https://doi.org/10.1016/j.geb.2009.10.014

Büchner, S., Coricelli, G., & Greiner, B. (2007). Self-centered and other-regarding behavior in the solidarity game. *Journal of Economic Behavior & Organization*, *62*(2), 293–303. https://doi.org/10.1016/j.jebo.2004.12.006

Burnham, T. C. (2003). Engineering altruism: A theoretical and experimental investigation of anonymity and gift giving. *Journal of Economic Behavior & Organization*, *50*(1), 133–144. https://doi.org/10.1016/S0167-2681(02)00044-6

Chandrasekhar, A. G., Kinnan, C., & Larreguy, H. (2018). Social networks as contract enforcement: Evidence from a lab experiment in the field. *American Economic Journal: Applied Economics*, *10*(4), 43–78. https://doi.org/10.1257/app.20150057

Charness, G., & Gneezy, U. (2008). What's in a name? Anonymity and social distance in dictator and ultimatum games. *Journal of Economic Behavior & Organization*, *68*(1), 29–35. https://doi.org/10.1016/j.jebo.2008.03.001

Choi, G. S., & Storr, V. H. (2020). Market interactions, trust and reciprocity. *PLOS ONE*, *15*(5), e0232704. https://doi.org/10.1371/journal.pone.0232704

Cialdini, R. B., Brown, S. L., Lewis, B. P., Luce, C., & Neuberg, S. L. (1997). Reinterpreting the empathy–altruism relationship: When one into one equals oneness. *Journal of Personality and Social Psychology*, *73*(3), 481–494. https://doi.org/10.1037/0022-3514.73.3.481

Clark, M. S., Boothby, E., Clark-Polner, E., & Reis, H. T. (2015). Understanding prosocial behavior requires understanding relational context. In D. A. Schroeder & W. G. Graziano (Eds.), *The Oxford Handbook of Prosocial Behavior* (p. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195399813.013.37

Davis, M. H. (2015). Empathy and prosocial behavior. In D. A. Schroeder & W. G. Graziano (Eds.), *The Oxford Handbook of Prosocial Behavior* (p. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195399813.013.026

de Oliveira, A. C. M., Eckel, C. C., & Croson, R. T. A. (2014). Solidarity among the poor. *Economics Letters*, *123*(2), 144–148. https://doi.org/10.1016/j.econlet.2014.01.025

de Waal, F. (2009). *The age of empathy: Nature's lessons for a kinder society*. Crown.

de Waal, F. B. M. (2008). Putting the altruism back into altruism: The evolution of empathy. *Annual Review of Psychology*, *59*(1), 279–300. https://doi.org/10.1146/annurev.psych.59.103006.093625

de Waal, F. B. M., & Preston, S. D. (2017). Mammalian empathy: Behavioural manifestations and neural basis. *Nature Reviews Neuroscience*, *18*(8), Article 8. https://doi.org/10.1038/nrn.2017.72

Débarre, F., Hauert, C., & Doebeli, M. (2014). Social evolution in structured populations. *Nature Communications*, *5*(1), Article 1. https://doi.org/10.1038/ncomms4409

Dietz, T., Ostrom, E., & Stern, P. C. (2003). The struggle to govern the commons. *Science*, *302*(5652), 1907–1912. https://doi.org/10.1126/science.1091015

Dong, G., Li, S., Du, R., & Su, Q. (2023). Asymmetric effects of interaction and dispersal graphs on evolutionary dynamics. *Chaos, Solitons & Fractals*, *173*, 113758. https://doi.org/10.1016/j.chaos.2023.113758

Edele, A., Dziobek, I., & Keller, M. (2013). Explaining altruistic sharing in the dictator game: The role of affective empathy, cognitive empathy, and justice sensitivity. *Learning and Individual Differences*, *24*, 96–102. https://doi.org/10.1016/j.lindif.2012.12.020

Eisenberg, N., Spinrad, T. L., & Morris, A. S. (2013). Prosocial development. In P. D. Zelazo (Ed.), *The Oxford Handbook of Developmental Psychology, Vol. 2: Self and Other* (p. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199958474.013.0013

Frey, B. S., & Bohnet, I. (1997). Identification in democratic society. *The Journal of Socio-Economics*, *26*(1), 25–38. https://doi.org/10.1016/S1053-5357(97)90050-0

Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., & Soutter, C. L. (2000). Measuring trust*. *The Quarterly Journal of Economics*, *115*(3), 811–846. https://doi.org/10.1162/003355300554926

Goeree, J. K., McConnell, M. A., Mitchell, T., Tromp, T., & Yariv, L. (2010). The 1/d law of giving. *American Economic Journal: Microeconomics*, *2*(1), 183–203. https://doi.org/10.1257/mic.2.1.183

Gonzalez, M. Z., & Rice, M. A. (2024). Behavioural sciences need behavioural ecology. *Nature Human Behaviour*, *8*(7), 1240–1242. https://doi.org/10.1038/s41562-024-01906-8

Grohn, J., Huck, S., & Valasek, J. M. (2014). A note on empathy in games. *Journal of Economic Behavior & Organization*, *108*, 383–388. https://doi.org/10.1016/j.jebo.2014.01.008

Habyarimana, J., Humphreys, M., Posner, D. N., & Weinstein, J. M. (2007). Why does ethnic diversity undermine public goods provision? *The American Political Science Review*, *101*(4), 709–725.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. *Journal of Theoretical Biology*, *7*(1), 1–52. https://doi.org/10.1016/0022-5193(64)90038-4

Henrich, J. (2018). Human cooperation: The hunter-gatherer puzzle. *Current Biology*, *28*(19), R1143–R1145. https://doi.org/10.1016/j.cub.2018.08.005

Henrich, J., & Muthukrishna, M. (2021). The origins and psychology of human cooperation. *Annual Review of Psychology*, *72*(1), 207–240. https://doi.org/10.1146/annurev-psych-081920-042106

Hoffman, E., McCabe, K., & Smith, V. (2000). The Impact of Exchange Context on the Activation of Equity in Ultimatum Games. *Experimental Economics*, *3*(1), 5–9. https://doi.org/10.1023/A:1009925123187

Hoffman, M. L. (2008). Empathy and prosocial behavior. In *Handbook of emotions* (Vol. 3, pp. 440–455). https://books.google.com/books?hl=en&lr=&id=uIQQskejGwUC&oi=fnd&pg=PA440&dq=hoffman+2000+empathy&ots=3RTnVRicD6&sig=NqSk9N1_ju_RD5EfvSW7xDPbqyc

Howe, E. L., Murphy, J. J., Gerkey, D., Stoddard, O. B., & West, C. T. (2023). Sharing, social norms, and social distance: Experimental evidence from Russia and Western Alaska. *Journal of Economic Behavior & Organization*, *213*, 345–358. https://doi.org/10.1016/j.jebo.2023.07.026

Jones, B. A. (2022). A review of social discounting: The impact of social distance on altruism. *The Psychological Record*, *72*(3), 511–515. https://doi.org/10.1007/s40732-021-00488-5

Jones, B. A., & Rachlin, H. (2009). Delay, probability, and social discounting in a public goods game. *Journal of the Experimental Analysis of Behavior*, *91*(1), 61–73. https://doi.org/10.1901/jeab.2009.91-61

Jones, B., & Rachlin, H. (2006). Social discounting. *Psychological Science*, *17*(4), 283–286. https://doi.org/10.1111/j.1467-9280.2006.01699.x

Kamas, L., & Preston, A. (2021). Empathy, gender, and prosocial behavior. *Journal of Behavioral and Experimental Economics*, *92*, 101654. https://doi.org/10.1016/j.socec.2020.101654

Kardos, P., Leidner, B., & Nawalkha, S. (2024). What can we learn about human nature from interacting with strangers? Relationship type determines behavior in the dictator game. *The Journal of Psychology*, *0*(0), 1–23. https://doi.org/10.1080/00223980.2024.2437380

Kosse, F., Deckers, T., Pinger, P., Schildberg-Hörisch, H., & Falk, A. (2020). The formation of prosociality: Causal evidence on the role of social environment. *Journal of Political Economy*, *128*(2), 434–467. https://doi.org/10.1086/704386

Krysowski, E., & Tremewan, J. (2021). Why does anonymity make us misbehave: Different norms or less compliance? *Economic Inquiry*, *59*(2), 776–789. https://doi.org/10.1111/ecin.12955

Leider, S., Mobius, M. M., Rosenblat, T. S., & DO, Q.-A. (2008). Directed altruism and enforced reciprocity in social networks. *Quarterly Journal of Economics*. https://ink.library.smu.edu.sg/soe_research/1147

Lieberman, E., Hauert, C., & Nowak, M. A. (2005). Evolutionary dynamics on graphs. *Nature*, *433*(7023), Article 7023. https://doi.org/10.1038/nature03204

Ligon, E., & Schechter, L. (2012). Motives for sharing in social networks. *Journal of Development Economics*, *99*(1), 13–26. https://doi.org/10.1016/j.jdeveco.2011.12.002

List, J. A., Berrens, R. P., Bohara, A. K., & Kerkvliet, J. (2004). Examining the role of social isolation on stated preferences. *The American Economic Review*, *94*(3), 741–752.

Locey, M. L., Jones, B. A., & Rachlin, H. (2011). Real and hypothetical rewards in self-control and social discounting. *Judgment and Decision Making*, *6*(6), 552–564. https://doi.org/10.1017/S1930297500002515

Locey, M. L., Safin, V., & Rachlin, H. (2013). Social discounting and the prisoner's dilemma game. *Journal of the Experimental Analysis of Behavior*, *99*(1), 85–97. https://doi.org/10.1002/jeab.3

Loewenstein, G., & Small, D. A. (2007). The scarecrow and the tin man: The vicissitudes of human sympathy and caring. *Review of General Psychology*, *11*(2), 112–126. https://doi.org/10.1037/1089-2680.11.2.112

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*(1), 415–444. https://doi.org/10.1146/annurev.soc.27.1.415

Nowak, M. A., & May, R. M. (1992). Evolutionary games and spatial chaos. *Nature*, *359*(6398), Article 6398. https://doi.org/10.1038/359826a0

Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, *437*(7063), Article 7063. https://doi.org/10.1038/nature04131

Nowak, M. A., Tarnita, C. E., & Antal, T. (2010). Evolutionary dynamics in structured populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1537), 19–30. https://doi.org/10.1098/rstb.2009.0215

Ohtsuki, H., Hauert, C., Lieberman, E., & Nowak, M. A. (2006). A simple rule for the evolution of cooperation on graphs and social networks. *Nature*, *441*(7092), Article 7092. https://doi.org/10.1038/nature04605

Ohtsuki, H., & Iwasa, Y. (2006). The leading eight: Social norms that can maintain cooperation by indirect reciprocity. *Journal of Theoretical Biology*, *239*(4), 435–444. https://doi.org/10.1016/j.jtbi.2005.08.008

Ohtsuki, H., Pacheco, J. M., & Nowak, M. A. (2007). Evolutionary graph theory: Breaking the symmetry between interaction and replacement. *Journal of Theoretical Biology*, *246*(4), 681–694. https://doi.org/10.1016/j.jtbi.2007.01.024

Osiński, J. (2009). Kin altruism, reciprocal altruism and social discounting. *Personality and Individual Differences*, *47*(4), 374–378. https://doi.org/10.1016/j.paid.2009.04.011

Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press.

Ostrom, E. (2000). Collective action and the evolution of social norms. *Journal of Economic Perspectives*, *14*(3), 137–158. https://doi.org/10.1257/jep.14.3.137

Ostrom, E. (2009). A General Framework for Analyzing Sustainability of Social-Ecological Systems. *Science*, *325*(5939), 419–422. https://doi.org/10.1126/science.1172133

Ostrom, E., & Walker, J. (1997). Neither markets nor states: Linking transformation processes in collective action arenas. *Perspectives on Public Choice: A Handbook*, 35–72.

Pelligra, V. (2011). Empathy, guilt-aversion, and patterns of reciprocity. *Journal of Neuroscience, Psychology, and Economics*, *4*(3), 161–173. https://doi.org/10.1037/a0024688

Preston, S. D., & Waal, F. B. M. de. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, *25*(1), 1–20. https://doi.org/10.1017/S0140525X02000018

Pretty, J. (2003). Social capital and the collective management of resources. *Science*, *302*(5652), 1912–1914. https://doi.org/10.1126/science.1090847

Rachlin, H., & Jones, B. A. (2008). Altruism among relatives and non-relatives. *Behavioural Processes*, *79*(2), 120–123. https://doi.org/10.1016/j.beproc.2008.06.002

Rand, D. G., Arbesman, S., & Christakis, N. A. (2011). Dynamic social networks promote cooperation in experiments with humans. *Proceedings of the National Academy of Sciences*, *108*(48), 19193–19198. https://doi.org/10.1073/pnas.1108243108

Rand, D. G., & Nowak, M. A. (2013). Human cooperation. *Trends in Cognitive Sciences*, *17*(8), 413–425. https://doi.org/10.1016/j.tics.2013.06.003

Rand, D. G., Nowak, M. A., Fowler, J. H., & Christakis, N. A. (2014). Static network structure can stabilize human cooperation. *Proceedings of the National Academy of Sciences*, *111*(48), 17093–17098. https://doi.org/10.1073/pnas.1400406111

Rankin, F. W. (2006). Requests and social distance in dictator games. *Journal of Economic Behavior & Organization*, *60*(1), 27–36. https://doi.org/10.1016/j.jebo.2004.08.004

Rege, M., & Telle, K. (2004). The impact of social approval and framing on cooperation in public good situations. *Journal of Public Economics*, *88*(7), 1625–1644. https://doi.org/10.1016/S0047-2727(03)00021-5

Rumble, A. C., Van Lange, P. A. M., & Parks, C. D. (2010). The benefits of empathy: When empathy may sustain cooperation in social dilemmas. *European Journal of Social Psychology*, *40*(5), 856–866. https://doi.org/10.1002/ejsp.659

Sally, D. (1995). Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992. *Rationality and Society*, *7*(1), 58–92. https://doi.org/10.1177/1043463195007001004

Sautter, J. A., Littvay, L., & Bearnes, B. (2007). A dual-edged sword: Empathy and collective action in the prisoner's dilemma. *Annals of the American Academy of Political and Social Science*, *614*(1), 154–171. https://doi.org/10.1177/0002716207306360

Schill, C., Anderies, J. M., Lindahl, T., Folke, C., Polasky, S., Cárdenas, J. C., Crépin, A.-S., Janssen, M. A., Norberg, J., & Schlüter, M. (2019). A more dynamic understanding of human behaviour for the Anthropocene. *Nature Sustainability*, *2*(12), Article 12. https://doi.org/10.1038/s41893-019-0419-7

Schmid, L., Chatterjee, K., Hilbe, C., & Nowak, M. A. (2021). A unified framework of direct and indirect reciprocity. *Nature Human Behaviour*, *5*(10), 1292–1302. https://doi.org/10.1038/s41562-021-01114-8

Schram, A., & Charness, G. (2015). Inducing social norms in laboratory allocation choices. *Management Science*, *61*(7), 1531–1546.

Sheng, A., Su, Q., Wang, L., & Plotkin, J. B. (2024). Strategy evolution on higher-order networks. *Nature Computational Science*, *4*(4), 274–284. https://doi.org/10.1038/s43588-024-00621-8

Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Sciences*, *1156*(1), 81–96. https://doi.org/10.1111/j.1749-6632.2009.04418.x

Singer, T., & Tusche, A. (2014). Understanding others: Brain mechanisms of theory of mind and empathy. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics (Second Edition)* (pp. 513–532). Academic Press. https://doi.org/10.1016/B978-0-12-416008-8.00027-9

Smaldino, P. E. (2023). *Modeling social behavior: Mathematical and agent-based models of social dynamics and cultural evolution*. Princeton University Press.

Su, Q., Li, A., Wang, L., & Eugene Stanley, H. (2019). Spatial reciprocity in the evolution of cooperation. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1900), 20190041. https://doi.org/10.1098/rspb.2019.0041

Su, Q., McAvoy, A., Mori, Y., & Plotkin, J. B. (2022). Evolution of prosocial behaviours in multilayer populations. *Nature Human Behaviour*, *6*(3), Article 3. https://doi.org/10.1038/s41562-021-01241-2

Su, Q., McAvoy, A., & Plotkin, J. B. (2023). Strategy evolution on dynamic networks. *Nature Computational Science*, *3*(9), 763–776. https://doi.org/10.1038/s43588-023-00509-z

Su, Q., & Stewart, A. J. (2025). Evolutionary dynamics of behavioral motivations for cooperation. *Nature Communications*, *16*(1), 4023. https://doi.org/10.1038/s41467-025-59366-1

Takagishi, H., Kameshima, S., Schug, J., Koizumi, M., & Yamagishi, T. (2010). Theory of mind enhances preference for fairness. *Journal of Experimental Child Psychology*, *105*(1), 130–137. https://doi.org/10.1016/j.jecp.2009.09.005

Tarnita, C. E., Ohtsuki, H., Antal, T., Fu, F., & Nowak, M. A. (2009). Strategy selection in structured populations. *Journal of Theoretical Biology*, *259*(3), 570–581. https://doi.org/10.1016/j.jtbi.2009.03.035

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, *46*(1), 35–57. https://doi.org/10.1086/406755

Vekaria, K. M., Brethel-Haurwitz, K. M., Cardinale, E. M., Stoycos, S. A., & Marsh, A. A. (2017). Social discounting and distance perceptions in costly altruism. *Nature Human Behaviour*, *1*(5), 1–7. https://doi.org/10.1038/s41562-017-0100

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature*, *393*(6684), 440–442. https://doi.org/10.1038/30918