

MAGKS



**Joint Discussion Paper
Series in Economics**

by the Universities of
Aachen · Gießen · Göttingen
Kassel · Marburg · Siegen

ISSN 1867-3678

No. 39-2017

Hannes Rusch

**The Evolution of Collaboration in Symmetric
2x2-Games with Imperfect Recognition of Types**

This paper can be downloaded from
<http://www.uni-marburg.de/fb02/makro/forschung/magkspapers>

Coordination: Bernd Hayo • Philipps-University Marburg
School of Business and Economics • Universitätsstraße 24, D-35032 Marburg
Tel: +49-6421-2823091, Fax: +49-6421-2823088, e-mail: hayo@wiwi.uni-marburg.de

"Forthcoming as Rusch, H. The evolution of collaboration in symmetric 2x2-games with imperfect recognition of types. Games and Economic Behavior (2019), <https://doi.org/10.1016/j.geb.2018.12.005>"

MACIE PAPER SERIES

Marburg Centre for
Institutional Economics



Nr. 2017/09

The Evolution of Collaboration in Symmetric 2x2-Games with Imperfect Recognition of Types

Hannes Rusch

MACIE, Philipps-Universität Marburg

Marburg Centre for Institutional Economics • Coordination: Prof. Dr. Elisabeth Schulte
c/o Research Group Institutional Economics • Barfuessertor 2 • D-35037 Marburg

Phone: +49 (0) 6421-28-23196 • Fax: +49 (0) 6421-28-24858 •
www.uni-marburg.de/fb02/MACIE • macie@wiwi.uni-marburg.de



The Evolution of Collaboration in Symmetric 2×2 -Games with Imperfect Recognition of Types

(Formerly: “Shared Intentions: Collaboration
Evolving”)

Hannes Rusch
University of Marburg & TU München
(hannes.rusch@tum.de)

This version: December 18, 2018

Abstract

A recent series of papers has introduced a fresh perspective on the problem of the evolution of human cooperation by suggesting an amendment to the concept of cooperation itself: instead of thinking of cooperation as playing a particular strategy in a given game, usually C in the prisoner’s dilemma, we could also think of cooperation as collaboration, i.e. as coalitional strategy choice, such as jointly switching from (D, D) to (C, C) . The present paper complements previous work on collaboration by expanding on its genericity while relaxing the assumption that collaborators are able to perfectly identify their own kind. Conditions for the evolutionary viability of such collaboration under fairly undemanding assumptions about population and interaction structure are derived. Doing so, this paper shows that collaboration is an adaptive principle of strategy choice in a broad range of *niches*, i.e., stochastic mixtures of games.

JEL: C73; **Keywords:** cooperation; collaboration; strategy choice; evolution; **Note:** This is the accepted manuscript version. It is not the copy of record. The citable final article is available through [doi:10.1016/j.geb.2018.12.005](https://doi.org/10.1016/j.geb.2018.12.005).

1 Introduction

Ample game theoretic research on the conditions allowing for specific types of cooperative behavior to be fostered by natural and/or cultural selection exists (see, e.g.: Bowles and Gintis 2011; Nowak 2012; Rand and Nowak 2013; Newton 2018). Nonetheless, a recent series of papers has introduced a fresh perspective on the subject within the game theoretic framework by suggesting an amendment to the concept of cooperation itself (Newton 2012; Sawa 2014; Angus and Newton 2015; Newton and Angus 2015; Newton 2017a). These authors argue that, instead of thinking of cooperation as playing a particular strategy in a specific game, usually C in the prisoner's dilemma [Pd], we could also think of cooperation as coalitional strategy choice, such as jointly switching from (D, D) to (C, C) in the Pd . To disambiguate play of a cooperative strategy from coalitional strategy choice, Angus and Newton (2015) suggest to refer to the latter as *collaboration*.

One particular strength of this concept of collaboration is its genericity, i.e. it provides a unified formal approach to describing cooperative behavior in more than one game. Correspondingly, Angus and Newton (2015) and Newton (2017a) have already shown that collaboration can be positively selected for by evolutionary processes when social interaction between individuals is modeled as one of a range of specific games.

The present paper complements previous work on collaboration by expanding on its genericity: conditions for the evolutionary viability of collaboration under fairly undemanding assumptions about population and interaction structure are derived. Doing so, this paper shows that collaboration is an adaptive principle of strategy choice in a broad range of *niches*, i.e., stochastic mixtures of games—a concept to be concretized later. Naturally, and importantly, analyses also characterize niches in which collaboration does not readily evolve. Readers interested more generally in the strengths and limitations of the concept of collaboration are referred to the papers referenced above, Newton (2017a) in particular.

This paper is organized as follows: Section 2 provides the motivation for the formal model presented in Section 3. This model is analyzed in Section 4. Section 5 discusses results and concludes.

2 Motivation

Canonically, game theoretic studies of the evolution of cooperativeness start with a given game, usually some variety of the *Pd* (e.g. Axelrod and Hamilton 1981; Nowak et al. 2004; van Veelen et al. 2012; for a literature review see, e.g., Nowak 2012). Then they add assumptions about population structure, interaction patterns, and information available to players. Next, they analyze under which conditions these ingredients facilitate the proliferation of strategies that entail some form of cooperative behavior. The fruitfulness of this approach is evident from the vast literature it has produced (for reviews see, e.g., Nowak 2006b; van Veelen 2009; Rand and Nowak 2013).

Beginning analyses by specifying a particular ‘base-game’ is inevitable as long as cooperativeness, i.e. the very phenomenon in focus, needs to be defined in terms of players playing a specific strategy of that game—be it *C* in the one-shot *Pd*, *TFT* in iterated *Pds*, ‘*Stag*’ in Rousseau’s stag hunt [*Sh*], or positive contributions in public good games. However, the concept of collaboration renders an alternative approach possible. As collaboration represents a principle of strategy choice, shorthand: a *maxim*, it can be defined generically, i.e. independently of any concrete game (for comprehensive discussions of the relation of collaborative maxims with other principles of strategy choice, best-responding in particular, see: Gold and Colman 2018; Karpus and Radzvilas 2018; Newton 2012, 2017a, 2018).

2.1 What is studied here?

Following Newton (2017a), we will operationalize collaboration here as an ability of players to determine a status quo strategy profile for any given game and to jointly optimize their payoffs subsequently, i.e. to search for possible Pareto-improvements from the status quo and to coordinate on them if available. Studying (the evolution of) such collaborative maxims has already proven fruitful; Newton (2018) provides a comprehensive overview.

Notably, previous work has also already identified several scenarios in which the evolution of collaboration is hampered. This is remarkable since successfully collaborating players can only attain mutual benefits relative to the status quo. However, as Newton (2017a) shows, collaboration can suffer from free-riding when collaborating entails costs while its benefits are

shared among more players than just the collaborators. Moreover, when positive assortativity of player types and negative externalities of collaboration are assumed, non-collaborating players can fare better as they are less likely to suffer the harm that collaborators impose on their vicinity (*ibid.*). In network-structured populations, furthermore, collaboration can have retarding effects on the spreading of advantageous strategy choice, which can lead to collaborative groups being outcompeted by less collaborative ones (Angus and Newton 2015; Newton and Angus 2015).

Instead of studying the performance of collaboration in similarly complex cases, the present paper looks at unstructured populations of unsophisticated agents inhabiting simple but very variable environments and asks if collaboration can evolve in such primordial settings. Thus, roughly speaking, this paper traces some elementary ecological conditions that need to be met in order for collaboration to be able to play out its potential advantages.

We stress that collaboration is not construed here as being a *rational* principle of strategy choice from any individual’s perspective—note, e.g., that collaboration can lead to the choice of strictly dominated strategies (like C in the one-shot Pd). We also emphasize that collaboration as construed here does not require a separate evolution of individuals’ preferences (*sensu* Bergstrom 1995 and Alger and Weibull 2013, 2016, 2017; also see Newton 2017a, 2017b, 2018). Instead, collaborators in our model are coalitional fitness optimizers, i.e. they exploit all potentials for mutual benefits, including those that require asymmetric strategy choice. Possibly, hence, collaborators leave some opportunities for individual fitness maximization unused. Thus, while collaborators avoid the inefficiencies that all-out individual fitness maximization can result in, they remain vulnerable to invasions by more specialized individual fitness maximizers in case such specialization yields sufficiently large fitness benefits.

This paper, however, is concerned with the question of whether such collaboration represents an evolutionarily viable solution to the more generic problem sitting at the core of the literature concerned with the origins of (human) cooperativeness, namely reaping “those many benefits sought by living things that are disproportionately available to cooperating groups” (adapted from Axelrod and Hamilton 1981, p. 1391). Complementing that

literature, however, we relax the constraints on the strategic nature of the game played, i.e. we include more games than just the *Pd*, and analyze the evolutionary performance of maxims—as opposed to strategies.

2.2 How is it studied?

In short, thus, this paper tries to shed some light on possible beginnings of collaboration as a maxim governing behavior in social interaction. Therefore, the core version of the model presented in Section 3 comprises only three types of players: (i) self-sufficient ‘loners’, who do not interact socially at all, thereby foregoing the potential benefits of, but simultaneously avoiding all risks entailed by, interacting socially; (ii) ‘self-protectors’, who interact socially but minimize their risk from doing so; and (iii) collaborators, who can potentially reap all the mutual benefits of social interaction when interacting with their own kind. However, note that we deviate from earlier work on collaboration by dropping the assumption that collaborators are capable of recognizing each other flawlessly. Instead, we assume that attempts to collaborate may result in detrimental failures for collaborators who interact with non-collaborators, making collaboration the most vulnerable maxim in the competition.

We assume that interactions are one-shot and that the games played are voluntary, symmetric, simultaneous 2×2 -games with random payoffs. We confine ourselves to symmetric 2×2 -games to maintain comparability with the bulk of the previous literature on the evolution of cooperativeness. We use voluntary games, because we are interested in the question of how collaboration can get started in an asocial world (see Axelrod and Hamilton 1981, p. 1391; Hauert et al. 2002, 2007; Silva et al. 2010; also note that compulsory games are included as a special case in the model).

Furthermore, for our main analyses in Section 3 we assume that players are unsophisticated in the following sense: (i) they cannot choose mixed strategies, and (ii) they do not reason strategically. The motivation for fixing these two restrictions is mostly biological, as, arguably, humans developed their potentials for strategic finesse only after they had evolved into a cooperative species (Tomasello 1999, 2012). However, results obtained for a version of the model relaxing these two restrictions show that collabora-

tion can also evolve when faced with slightly more sophisticated competitors (see Appendix A.2).

For now, though, we let collaboration compete with two other maxims: ‘self-sufficiency’ and ‘self-protection’. All three maxims determine behavior in voluntary 2×2 -games that have the following structure. In stage one, players independently decide whether to engage in social interaction or not. If at least one player opts out, no interaction takes place and both receive a fixed baseline payoff. The maxim of self-sufficiency always opts out at this stage, the other two maxims opt in.

Stage two is a move by nature: the four payoffs required to define a symmetric 2×2 -game are drawn from a distribution centered around the baseline payoff and arranged such that the probability that a 2×2 -game of type γ is realized is given by $p_\gamma \in [0, 1]$, where γ denotes the Pd , Sh , etc. (see Section 3).

In stage three, players select a strategy using their maxim given these payoffs. To this end, players using the maxim of self-protection apply the *maximin* rule (leading, e.g., to choice of D in the Pd). Players using the maxim of collaboration also determine a status quo strategy profile by applying *maximin*. However, if (and only if) that status quo profile is inefficient, they jointly switch to playing an efficient profile. A possible process resulting in such coordination is the following. When two collaborators meet, nature randomly selects one of them to be quicker. That quicker one sends a signal to its partner conveying a message saying “Hey, we’re about to do (X, X) . Let’s do (Y, Z) together instead!”, where (Y, Z) is a Pareto-improvement relative to (X, X) and efficient. Then, the quicker collaborator plays Y and the slower one acquiesces and plays Z . For the case of multiple possible Pareto-improvements from the status quo, we assume that the quicker collaborator chooses what is best for herself. When a collaborator meets a ‘self-protector’ and the status quo is inefficient, the collaborator follows the procedure just described; the ‘self-protector’, however, does not follow and sticks with the status quo strategy determined by *maximin*.

Of course, this setup is stylized in many respects. What we want to model here are only certain essential features of primordial ecologies. One, by allowing for voluntary entry into social interaction as a first stage, we model asocial self-sufficiency as the reference case—as opposed to ‘defection’

or ‘cheating’ or ‘free-riding’. This can be considered as being biologically more realistic (Hauert et al. 2002, 2007), as social behavior obviously needs to evolve from asociality as a first step. Two, by letting payoffs realize only after players have opted for interacting socially, i.e. disallowing players to opt out of dilemmatic 2×2 -games like the *Pd*, we make it harder for collaboration to evolve, as it is vulnerable to detrimental failures of coordination in these types of interactions. Three, self-protection is a rather natural choice of opponent maxim, because of (i) its low marginal advancement relative to self-sufficiency—only information about own payoffs is required for applying the maximin rule; (ii) its behavioral equivalence to defective strategies like ‘*AllD*’ in the *Pd*; and, maybe most importantly, (iii) its genericity—unlike best-responding, e.g., the maximin rule always selects a unique pure strategy in the games studied here, i.e. it is generically applicable. Also note that ‘defection’ or ‘uncooperative behavior’ may be well defined for variants of the *Pd*, i.e. on the strategy level. However, a formal concept of ‘defection’ on the maxim level is not available, yet.

The performance of collaboration against other opponent maxims is certainly worth being studied, too. (For results on how it fares against mixed strategy best-responding and security strategy play see the Appendix, A.2). However, definitions of more exploitative opponent maxims than self-protection are also more demanding with respect to the cognitive abilities of players. The maxims compared here, instead, are cognitively rather facile and thus, arguably, more likely to represent first steps in a series of evolutionary refinements of maxims guiding social interaction (Tomasello et al. 2005; Tomasello et al. 2012; Rusch and Luetge 2016). Furthermore, the results presented in Section 4 already show that collaboration, even when merely faced with self-protection as its opponent maxim, does not evolve as readily as one may be tempted to expect given its intuitively quite obvious advantages.

3 Model description

We analyze evolutionary dynamics in an unstructured population consisting of N animals. Reproductive success is fitness proportional. The baseline fitness of all animals is 1. Animals have one of three types: L (‘loners’), M

(‘maximiners’), or S (‘intention sharers’). L -types do not engage in social interaction with other animals, whereas M - and S -types do. When two animals engage in social interaction, they face a simultaneous, one-shot, symmetric 2×2 -game as given in matrix A .

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad (1)$$

Herein, payoffs a, b, c, d are determined by drawing four i.i.d. random variables from a symmetric distribution, $F(X)$, with mean 1 and support Z , and arranging them into A such that the probability that A represents a game of type γ is given by p_γ . Thus, each time two animals interact socially, they play one of the twelve strategically distinct symmetric 2×2 -games with payoffs $X_{(i)}, i \in \{1, 2, 3, 4\}$, where $X_{(i)}$ denotes the i th order statistic of sampling four values from Z according to $F(X)$. Shorthand, we write $\mathbb{1} = \mathbb{E}(X_{(1)}), \mathbb{2} = \mathbb{E}(X_{(2)}),$ etc. (In the following we let $F(X) = X/2$ and $Z = [0, 2]$, i.e. payoffs are uniformly distributed over $[0, 2]$, yielding $4 = 1.6, 3 = 1.2, 2 = 0.8, 1 = 0.4$. However, *mutatis mutandis*, qualitative results hold for any symmetric CDF.)

Using the notation suggested by Bruns (2015), the set of game types potentially played by social types is

$$\Gamma = \{As, Ba, Ch, Cm, Co, Dl, Ha, Hr, Nc, Pc, Pd, Sh\}.$$

More precisely, thus, when two social types are matched for their one-shot interaction, the probability that they will be playing a game of type $\gamma \in \Gamma$ is given by p_γ with $\sum_{\gamma \in \Gamma} p_\gamma = 1$.

As explained in Section 2, the two social types differ in their maxims. M -types apply the maximin rule: they choose their strategy such that they never receive the lowest possible payoff ($= \mathbb{1}$), irrespective of their opponent’s choice. S -types, on the other hand, use the maximin rule to determine a status quo strategy profile but then check for mutually beneficial, i.e. Pareto-better, deviations from that status quo profile. If one such Pareto-better strategy profile exists, they jointly deviate accordingly. If two such Pareto-better profiles exist, S -types coordinate on each of them with equal probability. If none exists, they stick to the status quo profile. Table 1

<i>Ch</i>	L	R	<i>Cm</i>	L	R	<i>Co</i>	L	R
U	<u> 3, 3 </u>	2, 4	U	<u> 3, 3 </u>	4, 2	U	<u> 4, 4 </u>	2, 1
D	4, <u>2</u>	1, 1	D	2, 4	1, 1	D	1, 2	3, 3
<i>Dl</i>	L	R	<i>Ha</i>	L	R	<i>Nc</i>	L	R
U	<u> 3, 3 </u>	4, 1	U	<u> 4, 4 </u>	3, 2	U	<u> 4, 4 </u>	2, 3
D	1, 4	2, 2	D	2, 3	1, 1	D	3, 2	1, 1
<i>Pc</i>	L	R	<i>As</i>	L	R	<i>Ba</i>	L	R
U	<u> 4, 4 </u>	3, 1	U	<u> 4, 4 </u>	1, 2	U	2, 2	<u> 3, 4 </u>
D	1, 3	2, 2	D	2, 1	<u>3, 3</u>	D	<u> 4, 3 </u>	1, 1
<i>Hr</i>	L	R	<i>Pd</i>	L	R	<i>Sh</i>	L	R
U	2, 2	<u> 4, 3 </u>	U	<u> 3, 3 </u>	1, 4	U	<u> 4, 4 </u>	1, 3
D	<u> 3, 4 </u>	1, 1	D	4, 1	<u>2, 2</u>	D	3, 1	<u>2, 2</u>

Table 1: Overview of the 12 strict symmetric ordinal 2×2 -games; underlined profiles are reached by M -types, profiles in norm dashes (|| \bullet , \bullet ||) are reached by S -types

shows the resulting strategy choices by S - and M -types for all games in Γ . When an S -type plays with an M -type, the S -type behaves as if matched with another S -type and is thus vulnerable to failures of coordination on Pareto-better profiles. When a social type plays with a loner, finally, no interaction takes place, and both receive the baseline payoff of 1 ($\neq 1$).

As can be seen from Table 1, M - and S -types choose the same strategies in seven games (Ch, Cm, Co, Dl, Ha, Nc , and Pc). In the remaining five games, however, their choices differ: $\theta = \{As, Ba, Hr, Pd, Sh\}$. Obviously, these five games are the ones decisive for the dynamics of the population. Therefore, let their individual occurrence probabilities be denoted by: $0 \leq p_{As}, p_{Ba}, p_{Hr}, p_{Pd}, p_{Sh} \leq 1$, with $\sum_{i \in \theta} p_i \leq 1$. Shorthand, we say that $\eta = (p_i)_{i \in \theta}$ characterizes the *niche* that the population is inhabiting.

Finally, we assume that the expected fitness of the types in the population is approximated sufficiently well by the expected payoffs given in matrix G , with $\pi_{X,Y}(\eta)$ denoting the expected fitness of type X when play-

ing against type Y in niche η .

$$G = \begin{pmatrix} \pi_{S,S}(\eta) & \pi_{S,M}(\eta) & \pi_{S,L}(\eta) \\ \pi_{M,S}(\eta) & \pi_{M,M}(\eta) & \pi_{M,L}(\eta) \\ \pi_{L,S}(\eta) & \pi_{L,M}(\eta) & \pi_{L,L}(\eta) \end{pmatrix} \quad (2)$$

Herein,

$$\pi_{S,S}(\eta) = [50 + 6(p_{As} + p_{Sh}) - p_{Ba} - p_{Hr} - 8p_{Pd}] / 35,$$

$$\pi_{S,M}(\eta) = [50 - 36(p_{As} + p_{Pd} + p_{Sh}) + 6p_{Ba} - 22p_{Hr}] / 35,$$

$$\pi_{M,S} = [50 - 22(p_{As} + p_{Hr}) - 8(p_{Ba} + p_{Sh}) + 6p_{Pd}] / 35,$$

$$\pi_{M,M}(\eta) = [50 - 8p_{As} - 22(p_{Ba} + p_{Hr} + p_{Pd} + p_{Sh})] / 35,$$

and $\pi_{L,\bullet}(\eta) = \pi_{\bullet,L}(\eta) = 1$ always. (The non-trivial expected payoffs are obtained by summing over the respectively probability-weighted payoffs obtained by the types S and M in the games in Γ using $4 = 1.6$, $3 = 1.2$, $2 = 0.8$, $1 = 0.4$. The general form of G is derived in the Appendix, A.1.)

4 Results

Given a population of size N inhabiting a niche η , is it possible for S -types to invade? And if so, will they prevail?

4.1 Very large populations

We focus on the case of very large, well-mixed populations first, i.e. $N = \infty$. In these, population dynamics can be described using the replicator equation (eq. 3), wherein $\phi(t) = (s, m, l)^T$ denotes the shares of the respective types in the population at time t , implying $s + m + l = 1$ always.

$$\dot{\phi}_i(t) = \phi_i [(G\phi)_i - \phi^T G \phi], i \in \{1, 2, 3\} \quad (3)$$

First, we check for equilibria on the edges of the (s, m, l) -simplex. Short-hand, slightly abusing notation, let $S = (1, 0, 0)^T$, $M = (0, 1, 0)^T$, and $L = (0, 0, 1)^T$ denote the three monomorphic equilibria, i.e. the corners

of the simplex. We find an equilibrium on the S/M -edge, i.e. in the $(s, 1 - s, 0)^T$ -hyperplane, at

$$s^* = \frac{4p_{As} - 4p_{Ba} + 2p_{Pd} + 2p_{Sh}}{8p_{As} - 3p_{Ba} + 3p_{Hr} + 4p_{Sh}}. \quad (4)$$

The S/L - and M/L -edges, in contrast, are degenerate in the following sense. As $\pi_{L,\bullet}(\eta) = \pi_{\bullet,L}(\eta) = 1$ always, solving for payoff equality between M - and L -types yields that this edge either contains only monomorphic equilibria or is entirely equilibrial. The latter is the case if $\pi_{M,M}(\eta) = 1$, i.e. if

$$p_{Pd} = [15 - 8p_{As} - 22(p_{Ba} + p_{Hr} + p_{Sh})] / 22 =: p_{Pd}^{M/L}. \quad (5)$$

As $\pi_{S,S}(\eta) > 1$ always holds for $\sum_{i \in \theta} p_i \leq 1$, the S/L -edge only contains the monomorphic equilibria at S and L .

Comparative statics applied on eq. 5 reveal that particularly increasing probabilities of Ba , Hr , Pd and Sh work against M -types' success relative to L -types. This, of course, is due to the fact that M -types playing against their own type receive a payoff of $2 < 1$ in these four games. Moreover, comparative statics on eq. 4 show that S -types are more successful against M -types in niches with relatively higher probabilities of Ba and Hr , i.e. in niches featuring those two games more frequently in which mutual collaboration by S -types cures an inefficiency resulting from maximin strategy choice but is not harmed by the coordination failures occurring when S -types interact with M -types.

Next, we check for asymptotic stability of S and M . (Note that as $\pi_{S,S}(\eta) > 1$ always holds, L can never be stable.) We find that S is asymptotically stable as long as $\pi_{S,S}(\eta) > \pi_{M,S}(\eta)$, i.e.,

$$p_{Pd} < 2p_{As} + \frac{1}{2}p_{Ba} + \frac{3}{2}p_{Hr} + p_{Sh} =: p_{Pd}^S, \quad (6)$$

with the signs of the terms in eq. 6 reflecting the fact that the Pd is the only game in Γ in which M -types do better against S -types than these do against their own kind. Similarly, we find that M is asymptotically stable as long as $p_{Pd} < p_{Pd}^{M/L}$ and $\pi_{M,M}(\eta) > \pi_{S,M}(\eta)$, the latter condition being

$$p_{Pd} > 2p_{Ba} - 2p_{As} - p_{Sh} =: p_{Pd}^M, \quad (7)$$

or, equivalently, $p_{Ba} < p_{As} + \frac{1}{2}(p_{Pd} + p_{Sh})$, and reflecting the fact that Ba is the only game in Γ in which S -types do better against M -types than these do against their own kind.

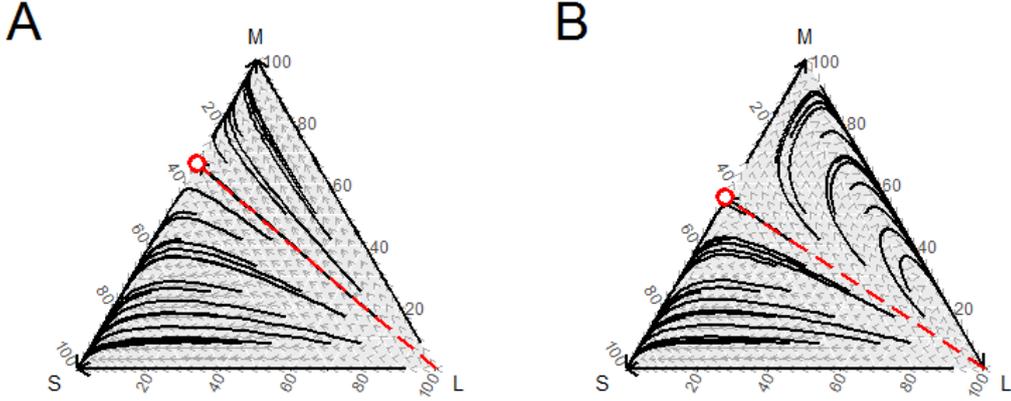


Figure 1: *Two illustrative dynamics.*

Panel A: $p_{As} = p_{Ba} = p_{Hr} = p_{Pd} = p_{Sh} = \frac{1}{12}$;
 Panel B: $p_{As} = p_{Hr} = p_{Sh} = \frac{1}{10}, p_{Ba} = \frac{2}{10}, p_{Pd} = \frac{3}{10}$

Figure 1 illustrates evolutionary dynamics for two niches. Panel A of Fig. 1 shows the dynamics for $p_i = \frac{1}{12}, \forall i \in \theta$, implying $p_{Pd}^M < p_{Pd} < p_{Pd}^S, p_{Pd}^{M/L}$, i.e. both S and M are stable (and $s^* = \frac{1}{3}$). Parameters in panel B are $p_{As} = p_{Hr} = p_{Sh} = \frac{1}{10}, p_{Ba} = \frac{2}{10}$, and $p_{Pd} = \frac{3}{10}$, implying $p_{Pd}^M, p_{Pd}^{M/L} < p_{Pd} < p_{Pd}^S$, i.e. S is stable, M is unstable (and $s^* = \frac{4}{9}$).

4.2 Finite populations

We have just derived the conditions that niches must fulfill in order for S -types to be able to invade into and grow to dominate very large populations consisting of S -, M - and L -types. Furthermore, when $s^* \leq 0$ in eq. 4 and $p_{Pd} < p_{Pd}^S$ hold simultaneously, S -types even prevail when exclusively competing against resident M -types. The condition $s^* \leq 0$ can be relaxed further in finite populations, i.e. when $N < \infty$. As shown by Nowak et al. (2004), a straightforward 1/3-rule applies for large finite populations in the limit of weak selection, i.e. when G is assumed to represent only a minor

component of animals' overall fitness (i.e. when w in eq. 9 is sufficiently small; also see, e.g., Taylor et al. 2004 and Sample and Allen 2017). For this case, we obtain that selection favors invading S -types replacing resident M -types (L -types being absent) for sufficiently large N and sufficiently weak selection if $p_{Pd}^M < p_{Pd} < p_{Pd}^S$ and $s^* < 1/3$. The latter condition holds if

$$p_{Pd} < [9p_{Ba} - 4p_{As} + 3p_{Hr} - 2p_{Sh}] / 6 =: p_{Pd}^{1/3}. \quad (8)$$

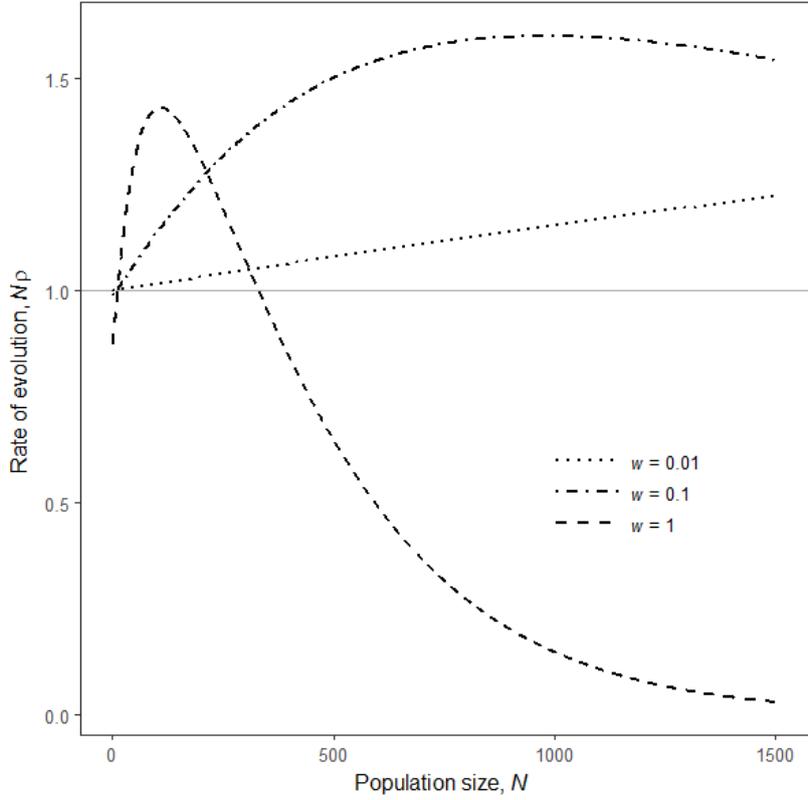


Figure 2: *Rates of evolution in finite populations of size N .*
Parameters: $p_{As} = p_{Pd} = \frac{2}{10}$, $p_{Ba} = \frac{3}{10}$, $p_{Hr} = 0$, and $p_{Sh} = \frac{1}{10}$

More generally, for any strength of selection $w \in [0, 1]$ and population size N , we can use the methods of Taylor et al. (2004) to calculate the fixation probability, ρ_S , for a single S -type in a finite population with $N - 1$ resident M -types using

$$\rho_S = 1 / \left(1 + \sum_{k=1}^{N-1} \prod_{i=1}^k \frac{g_i}{f_i} \right), \quad (9)$$

wherein $f_i = 1 - w + w [\pi_{S,S}(\eta)(i - 1) + \pi_{S,M}(\eta)(N - i)] / [N - 1]$ and $g_i = 1 - w + w [\pi_{M,S}(\eta)i + \pi_{M,M}(\eta)(N - i - 1)] / [N - 1]$ (see, e.g., Taylor et al. 2004 and Nowak 2006a for methodological details). Whenever $\rho_S > 1/N$, i.e. whenever the fixation probability of a single S -type is larger than its fixation probability in the case of no selection ($1/N$), we have positive selection for S -types. As eq. 9 contains N th-order polynomials, though, no convenient form of this condition can be obtained. Figure 2 shows rates of evolution ($N\rho_S$) for different population sizes and selection strengths for $p_{As} = p_{Pd} = \frac{2}{10}$, $p_{Ba} = \frac{3}{10}$, $p_{Hr} = 0$, and $p_{Sh} = \frac{1}{10}$. As can be seen from Fig. 2, numerical evaluations of eq. 9 indicate that selection favors S -types replacing M -types, i.e. $N\rho_S > 1$, for $w \in \{1, 0.1, 0.01\}$ as long as $13 \leq N \leq 330$ in this particular niche.

5 Discussion and conclusion

The model devised and analyzed here demonstrates that collaboration as a principle of strategy choice, i.e. as a maxim, can be evolutionarily viable and successful in both finite and infinite populations. Collaboration can prevail against both self-sufficiency and self-protection as opponent maxims provided that the niches inhabited by the respective populations fulfill certain conditions.

Notably, collaboration’s potential for evolutionary success in this model is not based on repeated encounter, population structure, or information about past behavior nor any of the other previously studied factors favoring the evolution of cooperativeness (see, e.g.: Nowak 2006b). In fact, we have seen that collaboration can potentially prevail in entirely unstructured populations, even when all interaction is assumed to be one-shot. Rather, collaboration’s evolutionarily fate in this model depends on whether social interaction offers sufficiently many opportunities for attaining mutual benefits, i.e. on whether a population’s niche favors collaboration or not. In light of these results, several observations are worth being addressed.

One, previous work on the evolution of cooperativeness has mostly focused on the Pd in its many varieties, as it represents “the most stringent cooperative dilemma” (Nowak 2012, p. 1). The model presented here reconfirms this focus. In niches that are ‘too dilemmatic’, i.e. whenever combina-

tions of p_{As} , p_{Pd} , and p_{Sh} exceed certain thresholds, collaboration does not evolve. However, the model also shows that there are ‘quite dilemmatic’ niches in which it still does. (For appropriate N and w , e.g., collaboration can evolve in finite populations inhabiting the niche characterized by $p_{As} = 0$, $p_{Ba} = 0.2$, $p_{Hr} = 0.3$, $p_{Sh} = 0.1$, and $p_{Pd} = 0.4$, i.e. a niche in which forty percent of all social interactions are Pds .)

Two, it may be deemed a weakness of collaboration that it cannot evolve in niches that are too dilemmatic. However, when applied to the question of why humans are highly collaborative while other species are not, or not as much, this weakness may have some explanatory value. Think of the rudimentary collaborative maxim studied here as modeling an early step in the evolution of human cooperative behavior. Then, the main implication of the present model is that we should try to find out what types of niches our ancestors were inhabiting and how these differed from those occupied by other animals. This way of phrasing and formally modeling the problem of ‘the evolution of human cooperation’ seamlessly connects with less formal biological theorizing, particularly in evolutionary anthropology (Tomasello 2009; Tomasello et al. 2012), and follows the principles of behavioral ecology (Davies, Krebs, and West 2012).

Three, apart from its potential value for the study of the evolutionary origins of human cooperative behavior, studying collaboration as a maxim may also prove helpful in explaining choice behavior of contemporary humans. A recent strand of experimental literature in economics and psychology has begun to study the question of whether participants in laboratory experiments use distinct strategies for different games they play or whether they follow more generic heuristics that do not distinguish too sharply between different strategic contexts (e.g. Bednar et al. 2012; Peysakhovich, Nowak, and Rand 2014; Rand et al. 2014; Peysakhovich and Rand 2016; Rusch and Luetge 2016; Angelovski et al. 2018). The evidence collected in these studies points more in the direction of the latter conjecture, rendering maxims a promising formal tool for modeling decision behavior of this kind.

Finally, the model presented here has several limitations, including the following. One, only symmetric 2×2 -games were studied. Two, maxims were assumed to be inherited without mutations. Three, the results presented are limited to stylized primordial ecologies and rather unsophisti-

cated agents typifying a limited set of maxims. Removing these limitations, particularly by allowing for advancement of the strategic finesse of maxims, is a promising task for future research.

Appendix

A.1: Derivation of payoff matrix G

Given a niche $\eta = (p_{As}, p_{Ba}, p_{Hr}, p_{Pd}, p_{Sh})$ we can derive the entries of G as follows. First, note that with probability $p_R = 1 - p_{As} - p_{Ba} - p_{Hr} - p_{Pd} - p_{Sh}$ two animals play one of the seven games in which S - and M -types obtain the same payoff; these are: Ch, Cm, Co, Dl, Ha, Nc , and Pc . For simplicity, we assume that each of these realizes with the same probability, resulting in an expected payoff of $p_R \cdot (4 \cdot 4 + 3 \cdot 3) / 7$ for S - and M -types in these cases. Payoffs in the remaining cases differ for S - and M -types; these are $\theta = \{As, Ba, Hr, Pd, Sh\}$.

Take the example of the Pd . It realizes with probability p_{Pd} . S -types are able to obtain 3 when playing against each other, because of their ability to jointly switch from the inefficient status quo, (D,R) in Table 1, to the Pareto-better and efficient cooperative strategy profile, (U,L). S -types playing against M -types, however, obtain 1: assuming, w.l.o.g., that the S -type is the row player, the S -type will try to initiate joint switching to (U,L), but the M -type will not follow, resulting in (U,R) being played. M -types, vice versa, obtain 4 when matched with an S -type. M -types matched with other M -types, finally, stick with the status quo and thus obtain 2.

Take Hr as another example and look, w.l.o.g., at the row player. When two S -types meet, each of them is equally likely to be the first to signal where to jointly deviate from the status quo, which is (U,L) in Table 1. Thus, each obtains an expected payoff of $(4 + 3)/2$. When an S -type meets an M -type, both receive 2, because the possible Pareto-improvement that the S -type ‘suggests’, (U,R), would require the column player to switch to R, which she does not. When two M -types meet, finally, both stick with the status quo and obtain 2.

Payoffs for the other games in θ are calculated analogously to these examples, resulting in

$$\pi_{S,S}(\eta) = 4 \cdot (p_{As} + p_{Sh}) + \frac{4 + 3}{2} \cdot (p_{Hr} + p_{Ba}) + 3 \cdot p_{Pd} + \frac{4 \cdot 4 + 3 \cdot 3}{7} \cdot p_R,$$

$$\pi_{S,M}(\eta) = 4 \cdot p_{Ba} + 2 \cdot p_{Hr} + 1 \cdot (p_{As} + p_{Pd} + p_{Sh}) + \frac{4 \cdot 4 + 3 \cdot 3}{7} \cdot p_R,$$

$$\pi_{M,S}(\eta) = 4 \cdot p_{Pd} + 3 \cdot (p_{Ba} + p_{Sh}) + 2 \cdot (p_{As} + p_{Hr}) + \frac{4 \cdot 4 + 3 \cdot 3}{7} \cdot p_R,$$

$$\pi_{M,M}(\eta) = 3 \cdot p_{As} + 2 \cdot (p_{Ba} + p_{Hr} + p_{Pd} + p_{Sh}) + \frac{4 \cdot 4 + 3 \cdot 3}{7} \cdot p_R.$$

A.2: Additional dynamics

In this supplementary section, we derive the conditions under which S -types can invade into and grow to dominate very large populations consisting of two other potential opponent maxims: R -types and C -types. Both these types use mixed strategies. R -types play the mixed strategy that makes their opponent exactly indifferent between her two strategies. I.e., R -types play their part of the mixed strategy Nash equilibrium of any symmetric 2×2 -game. Note that such strategy choice is quite elaborate and thus requires rather advanced cognitive capacities. Also note that pure strategy best responses lead to equilibrium selection problems in six of the twelve strict symmetric ordinal 2×2 -games—which is why pure strategy best-responding is ignored here (see, e.g., Newton 2017a for additional comparisons of collaborative types with best-responders). C -types are a bit less sophisticated. They play the mixed strategy that maximizes their minimum payoff under the assumption that their opponent chooses her pure strategy maximally malevolently, i.e. they play their *security strategy* (see Binmore 2007, Section 7.4). Table A.1 shows strategy choice by all three types.

Game	S -types	C -types	R -types
As	U	1/4	1/2
Ba	U or D	3/4	1/2
Ch	U	1	1/2
Co	U	1/2	1/4
Hr	U or D	1/2	3/4
Pd	U	0	0
Sh	U	0	1/2
Cm, Dl, Ha, Nc, Pc	U	1	1

Table A.1: strategy choice conditional on type of game by S -, C -, and R -types; probabilities stated are probabilities of playing U in the notation of Table 1; S -types choose pure strategies as before, C - and R -types choose mixed strategies.

All three types select the same strategy, U in the notation of Table 1, in five of the twelve 2×2 -games: Cm , Dl , Ha , Nc , and Pc . For simplicity, thus, let these games' respective probabilities all be equal to $p_R = (1 - p_{As} - p_{Ba} - p_{Ch} - p_{Co} - p_{Hr} - p_{Pd} - p_{Sh})/5$. Further, let $\eta = (p_i)_{i \in \{As, Ba, Ch, Co, Hr, Pd, Sh\}}$ in the following. Fixing $4 = 1.6$, $3 = 1.2$, $2 = 0.8$, and $1 = 0.4$, this yields

$$H = \begin{pmatrix} \pi_{S,S}(\eta) & \pi_{S,C}(\eta) & \pi_{S,R}(\eta) \\ \pi_{C,S}(\eta) & \pi_{C,C}(\eta) & \pi_{C,R}(\eta) \\ \pi_{R,S}(\eta) & \pi_{R,C}(\eta) & \pi_{R,R}(\eta) \end{pmatrix} \quad (\text{A.1})$$

with

$$\begin{aligned}
\pi_{S,S}(\eta) &= [36 + 4(p_{Sh} + p_{Co} + p_{As}) - 6(p_{Pd} + p_{Ch}) - p_{Hr} - p_{Ba}] / 25, \\
\pi_{S,C}(\eta) &= [72 - 52(p_{Sh} + p_{Pd}) - 12(p_{Hr} + p_{Co} + p_{Ch}) - 7p_{Ba} - 37p_{As}] / 50, \\
\pi_{S,R}(\eta) &= [36 - 11(p_{Sh} + p_{Hr} + p_{Co} + p_{Ch} + p_{Ba} + p_{As}) - 26p_{Pd}] / 25, \\
\pi_{C,S}(\eta) &= [36 - 6(p_{Sh} + p_{Ch}) + 4p_{Pd} - 11(p_{Hr} + p_{Co} + p_{Ba} + p_{As})] / 25, \\
\pi_{C,C}(\eta) &= [36 - 16(p_{Sh} + p_{Pd}) - 11(p_{Hr} + p_{Co} + p_{Ba} + p_{As}) - 6p_{Ch}] / 25, \\
\pi_{C,R}(\eta) &= [36 - 11(p_{Sh} + p_{Hr} + p_{Co} + p_{Ch} + p_{Ba} + p_{As}) - 16p_{Pd}] / 25, \\
\pi_{R,S}(\eta) &= [72 - 2(p_{Sh} + p_{Ch}) + 8p_{Pd} - 27p_{Hr} - 37p_{Co} - 32p_{Ba} - 12p_{As}] / 50, \\
\pi_{R,C}(\eta) &= [72 - 42p_{Sh} - 32p_{Pd} - 17(p_{Ba} + p_{Hr}) - 27(p_{Co} + p_{As}) - 2p_{Ch}] / 50, \\
\pi_{R,R}(\eta) &= [36 - 16p_{Pd} - 11(p_{Sh} + p_{Hr} + p_{Co} + p_{Ch} + p_{Ba} + p_{As})] / 25.
\end{aligned}$$

Applying standard replicator dynamics (as in eq. 3), we find an equilibrium on the S/R -edge at:

$$s^{S/R} = \frac{4p_{Pd}}{2p_{Sh} + 5p_{Hr} + 9p_{Co} - 2p_{Ch} + 6p_{Ba} + 4p_{As}}. \quad (\text{A.2})$$

Further, we find an equilibrium on the S/C -edge at:

$$s^{S/C} = \frac{4p_{Sh} + 4p_{Pd} - 2p_{Hr} - 2p_{Co} - 3p_{Ba} + 3p_{As}}{8p_{Sh} + 2p_{Hr} + 4p_{Co} + p_{Ba} + 9p_{As}}. \quad (\text{A.3})$$

The C/R -edge, finally, is degenerate. As $\pi_{C,R}(\eta) = \pi_{R,R}(\eta), \forall \eta$ (see eq. A.1), the dynamics on this edge solely depend on whether $\pi_{C,C}(\eta) \lesseqgtr \pi_{R,C}(\eta)$, implying that when

$$\pi_{C,C}(\eta) - \pi_{R,C}(\eta) = 2p_{Sh} - p_{Hr} + p_{Co} - 2p_{Ch} - p_{Ba} + p_{As} > 0$$

holds, C -types invade and fully replace R -types, S -types being absent.

Apart from a theoretically interesting competition between mixed-strategy maximin and best-response maxims (see Pruzhansky 2011, 2013), a noteworthy observation that this variant of our model yields is the following. As eq. A.2 shows, the success of S -types in competition with R -types crucially depends on the probability of Pds in a niche. When Pds are absent, R -types are always invaded and fully replaced by S -types in direct competitions between these two types; see Figure A.1 for an example. While this observation can again serve to justify the central role that has been assigned to the Pd in the literature on the evolution of cooperativeness, it also shows that best-responding, at least as we have operationalized it here, quickly succumbs to collaboration when we broaden the scope of adaptive problems faced by the agents under selection.

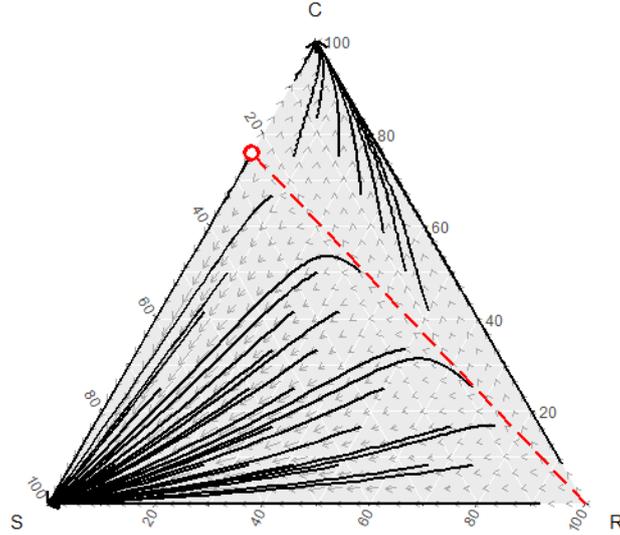


Figure A.1: *Illustrative dynamics in a population consisting of S-, C- and R-types.* Parameters: $p_{As} = p_{Sh} = \frac{3}{12}$, $p_{Ba} = p_{Ch} = p_{Co} = p_{Hr} = \frac{1}{12}$, $p_{Pd} = 0$

Acknowledgements

Helpful comments were provided by participants of the MACIE research seminar in Marburg, Matthijs van Veelen, and the Editors and Reviewers at *Games and Economic Behavior*. Special thanks are due to Jonathan Newton for inspiring discussions and very helpful feedback on earlier versions of this paper.

References

- Alger, I., and J. Weibull. 2016. “Evolution and Kantian morality”. *Games and Economic Behavior* 98:56–67. doi:10.1016/j.geb.2016.05.006.
- . 2013. “Homo Moralis–Preference Evolution Under Incomplete Information and Assortative Matching”. *Econometrica* 81 (6): 2269–2302. doi:10.3982/ECTA10637.
- . 2017. “Strategic Behavior of Moralists and Altruists”. *Games* 8 (3): 38. doi:10.3390/g8030038.
- Angelovski, A., D. Di Cagno, W. Güth, F. Marazzi, and L. Panaccione. 2018. “Behavioral spillovers in local public good provision: An experimental study”. *Journal of Economic Psychology* 67:116–134. doi:10.1016/j.joep.2018.05.003.
- Angus, S. D., and J. Newton. 2015. “Emergence of Shared Intentionality Is Coupled to the Advance of Cumulative Culture”. *PLOS Computational Biology* 11 (10): e1004587. doi:10.1371/journal.pcbi.1004587.
- Axelrod, R. M., and W. D. Hamilton. 1981. “The evolution of cooperation”. *Science* 211 (27 March): 1390–1396. doi:10.1126/science.7466396.
- Bednar, J., Y. Chen, T. X. Liu, and S. Page. 2012. “Behavioral spillovers and cognitive load in multiple games: An experimental study”. *Games and Economic Behavior* 74 (1): 12–31. doi:10.1016/j.geb.2011.06.009.
- Bergstrom, T. C. 1995. “On the Evolution of Altruistic Ethical Rules for Siblings”. *American Economic Review* 85 (1): 58–81. doi:10.2307/2117996.
- Binmore, K. G. 2007. *Playing for real: A text on game theory*. Oxford, New York: Oxford University Press.
- Bowles, S., and H. Gintis. 2011. *A cooperative species: Human reciprocity and its evolution*. Princeton: Princeton University Press.
- Bruns, B. 2015. “Names for Games: Locating 2×2 Games”. *Games* 6 (4): 495–520. doi:10.3390/g6040495.
- Davies, N. B., J. R. Krebs, and S. A. West. 2012. *An introduction to behavioural ecology*. 4. ed. Hoboken: Wiley-Blackwell.
- Gold, N., and A. M. Colman. 2018. “Team Reasoning and the Rational Choice of Payoff-Dominant Outcomes in Games”. *Topoi* 118 (4): 431. doi:10.1007/s11245-018-9575-z.
- Hauert, C., A. Traulsen, H. Brandt, M. A. Nowak, and K. Sigmund. 2007. “Via Freedom to Coercion: The Emergence of Costly Punishment”. *Science* 316 (5833): 1905–1907. doi:10.1126/science.1141588.
- Hauert, C., S. de Monte, J. Hofbauer, and K. Sigmund. 2002. “Volunteering as Red Queen mechanism for cooperation in public goods games”. *Science* 296 (5570): 1129–1132. doi:10.1126/science.1070582.
- Karpus, J., and M. Radzvilas. 2018. “Team Reasoning and a Measure of Mutual Advantage in Games”. *Economics and Philosophy* 34 (01): 1–30. doi:10.1017/S0266267117000153.
- Newton, J. 2012. “Coalitional stochastic stability”. *Games and Economic Behavior* 75 (2): 842–854. doi:10.1016/j.geb.2012.02.014.
- . 2018. “Evolutionary Game Theory: A Renaissance”. *Games* 9 (2): 31. doi:10.3390/g9020031.
- . 2017a. “Shared intentions: The evolution of collaboration”. *Games and Economic Behavior* 104:517–534. doi:10.1016/j.geb.2017.06.001.

- . 2017b. “The preferences of Homo Moralis are unstable under evolving assortativity”. *International Journal of Game Theory* 46 (2): 583–589. doi:10.1007/s00182-016-0548-4.
- Newton, J., and S. D. Angus. 2015. “Coalitions, tipping points and the speed of evolution”. *Journal of Economic Theory* 157:172–187. doi:10.1016/j.jet.2015.01.003.
- Nowak, M. A. 2006a. *Evolutionary dynamics: Exploring the equations of life*. Cambridge, Mass: Belknap Press of Harvard University Press.
- . 2012. “Evolving cooperation”. *Journal of Theoretical Biology* 299:1–8. doi:10.1016/j.jtbi.2012.01.014.
- . 2006b. “Five rules for the evolution of cooperation”. *Science* 314 (8 December): 1560–1563. doi:10.1126/science.1133755.
- Nowak, M. A., A. Sasaki, C. Taylor, and D. Fudenberg. 2004. “Emergence of cooperation and evolutionary stability in finite populations”. *Nature* 428 (6983): 646–650. doi:10.1038/nature02414.
- Peysakhovich, A., M. A. Nowak, and D. G. Rand. 2014. “Humans display a ‘cooperative phenotype’ that is domain general and temporally stable”. *Nature Communications* 5:4939. doi:10.1038/ncomms5939.
- Peysakhovich, A., and D. G. Rand. 2016. “Habits of Virtue: Creating Norms of Cooperation and Defection in the Laboratory”. *Management Science* 62 (3): 631–647. doi:10.1287/mnsc.2015.2168.
- Pruzhansky, V. 2013. “Maximin play in completely mixed strategic games”. *Theory and Decision* 75 (4): 543–561. doi:10.1007/s11238-013-9376-1.
- . 2011. “Some interesting properties of maximin strategies”. *International Journal of Game Theory* 40 (2): 351–365. doi:10.1007/s00182-010-0249-3.
- Rand, D. G., and M. A. Nowak. 2013. “Human cooperation”. *Trends in Cognitive Sciences* 17 (8): 413–425. doi:10.1016/j.tics.2013.06.003.
- Rand, D. G., A. Peysakhovich, G. T. Kraft-Todd, G. E. Newman, O. Wurzbacher, M. A. Nowak, and J. D. Greene. 2014. “Social heuristics shape intuitive cooperation”. *Nature Communications* 5. doi:10.1038/ncomms4677.
- Rusch, H., and C. Luetge. 2016. “Spillovers From Coordination to Cooperation: Evidence for the Interdependence Hypothesis?” *Evolutionary Behavioral Sciences*. doi:10.1037/ebs0000066.
- Sample, C., and B. Allen. 2017. “The limits of weak selection and large population size in evolutionary game theory”. *Journal of Mathematical Biology* 75 (5): 1285–1317. doi:10.1007/s00285-017-1119-4.
- Sawa, R. 2014. “Coalitional stochastic stability in games, networks and markets”. *Games and Economic Behavior* 88:90–111. doi:10.1016/j.geb.2014.07.005.
- Silva, H. de, C. Hauert, A. Traulsen, and K. Sigmund. 2010. “Freedom, enforcement, and the social dilemma of strong altruism”. *Journal of Evolutionary Economics* 20 (2): 203–217. doi:10.1007/s00191-009-0162-8.
- Taylor, C., D. Fudenberg, A. Sasaki, and M. A. Nowak. 2004. “Evolutionary game dynamics in finite populations”. *Bulletin of Mathematical Biology* 66 (6): 1621–1644. doi:10.1016/j.bulm.2004.03.004.
- Tomasello, M. 1999. *The Cultural Origins of Human Cognition*. Cambridge: Harvard University Press.

- . 2009. *Why we cooperate*. Cambridge: MIT Press.
- Tomasello, M., A. P. Melis, C. Tennie, E. Wyman, and E. Herrmann. 2012. “Two Key Steps in the Evolution of Human Cooperation”. *Current Anthropology* 53 (6): 673–692. doi:10.1086/668207.
- Tomasello, M., M. Carpenter, J. Call, T. Behne, and H. Moll. 2005. “Understanding and sharing intentions: The origins of cultural cognition”. *Behavioral and Brain Sciences* 28 (5): 675–691. doi:10.1017/S0140525X05000129.
- van Veelen, M. 2009. “Does it pay to be good? Competing evolutionary explanations of pro-social behaviour”. In *The Moral Brain*, ed. by J. Verplaetse, J. Schrijver, S. Vanneste, and J. Braeckman, 185–200. Dordrecht: Springer Netherlands. doi:10.1007/978-1-4020-6287-2\textunderscore}8.
- van Veelen, M., J. García, D. G. Rand, and M. A. Nowak. 2012. “Direct reciprocity in structured populations”. *Proceedings of the National Academy of Sciences of the United States of America* 109 (25): 9929–9934. doi:10.1073/pnas.1206694109.