Joint Discussion Paper Series in Economics

by the Universities of

Aachen · Gießen · Göttingen
Kassel · Marburg · Siegen

ISSN 1867-3678

No. 39-2017

Hannes Rusch

Shared Intentions: Collaboration Evolving

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

MACIE PAPER SERIES

Marburg Centre for Institutional Economics

MACIE

Nr. 2017/09

Shared Intentions: Collaboration Evolving

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

Philipps Universität Marburg

Shared Intentions: Collaboration Evolving

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

September 28, 2017

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: conditions for the evolutionary viability and stability 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.

1 Introduction

Humans are highly cooperative animals (Tomasello 2009). We possess a rich toolbox of cooperative behaviors and are equipped with the cognitive abilities required to carry them out. Arguably, cooperation with conspecifics is so essential to our way of life that we cannot survive more than a couple of days on our own (Tomasello 1999).

It may appear unambitious to claim that a species whose members are able to reap mutual benefits by working together in the pursuit of joint goals must evolutionarily fare better than species whose members are not. Still, comparative research in evolutionary anthropology suggests that the extent to which humans are able to detect opportunities for mutual benefit and to coordinate their actions in order to exploit them is unparalleled in the animal world (Tomasello et al. 2005; Bowles and Gintis 2011; Tomasello et al. 2012). This immediately prompts the question of why this is so, i.e. why similarly generic cooperative capabilities do not seem to have evolved in other species (Pennisi 2005).

Ample game theoretic research on the conditions allowing for cooperative behavioral traits to be fostered by natural and/or cultural selection exists (see, e.g.: Nowak 2006; West, Griffin, and Gardner 2007; West, El Mouden, and Gardner 2011; Nowak 2012; Rand and Nowak 2013). However, 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 2017). These authors argue that, instead of thinking of cooperation as playing a particular strategy in a given 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 the 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 New-

ton (2017) 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 and stability 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, 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.)

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; for a literature review see: Nowak 2012). Subsequently, they add assumptions about population structure, interaction patterns, and/or 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 2006; West, Griffin, and Gardner 2007; 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 its iterated variant, or positive levels of contribution in a public good game. However, the concept of collaboration renders an al-

terative approach possible. As collaboration represents a principle of strategy choice, shorthand: a *maxim*, it can be defined independently of any concrete game (for comprehensive discussions of the relation of collaborative maxims with more traditional principles of strategy choice, e.g. best-responding, see, e.g.: Karpus and Radzvilas, forthcoming; Newton 2012).

For the purpose of this paper, following Newton (2017), we will just assume that collaborative players are able 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. The only assumptions we will make about the games played are that these are voluntary, symmetric, simultaneous, one-shot 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 allow for voluntary games, as these may be more apt to capture the essence of the problem of the evolution of cooperativeness (Hauert et al. 2002, 2007; Silva et al. 2010; also note that compulsory games are included as a special case in the model presented below). We add to the existing literature by relaxing constraints on the strategic nature of the game played and analyzing the evolutionary performance of maxims—as opposed to strategies—in such a variable environment.

We will compare the performance of collaboration to that of two other maxims: self-sufficiency and self-protection. All three maxims determine behavior in volutary, one-shot 2×2-games, which are assumed have the following timing. 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-suffiency always opts out at this stage, while the other two opt in. In stage two, the payoffs of the 2×2-game realize. In stage three, players simultaneously play the strategy determined by their maxim given these payoffs. To this end, the maxim of self-protection applies the *maximin* rule, thereby securing that it cannot be taken advantage off by interaction partners (leading, e.g., to play of D in the Pd). The maxim of collaboration, in contrast, determines a status quo strategy profile by applying maximin, but then proceeds as described

above (leading, e.g., to play of C in the Pd).

The motivation for this setup is threefold. One, allowing for voluntary entry into social interaction as a first stage yields self-suffiency as the reference case—as opposed to defectiveness. This can be considered as being biologically more realistic (Hauert et al. 2002, 2007). Two, letting payoffs realize only after players have opted in, i.e. disallowing players to opt out of dillemmatic 2×2-games like the Pd, makes it harder for collaboration to evolve, as it is vulnarable to being exploited in these types of interactions. Three, self-protection is chosen here as an opponent maxim to collaboration because of (i) its simplicity—only information about own payoffs is required for applying the maximin rule—, (ii) its behavioral equivalency to defective strategies like *'AllD'* in the Pd, and (iii) its genericity. (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 'defectiveness' on the maxim level is not available, yet.)

Other opponent maxims are certainly worth being studied, too. However, definitions of more exploitative opponent maxims than self-protection are more demanding with respect to the cognitive abilities of players. The maxims compared here, instead, are relatively abstemious and thus, arguably, more likely to represent first steps in a series of evolutionary refinements of maxims guiding social interaction (Tomasello et al. 2012; Rusch and Luetge 2016). Furthermore, the model presented in the following Section 3 already shows that collaboration, even when 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 play a simultaneous, one-shot, symmetric 2×2-game represented by matrix A.

$$A = \begin{pmatrix} a & c \\ b & d \end{pmatrix} \tag{1}$$

Herein, payoffs a, b, c, d are i.i.d. random variables following a symmetric distribution, $F(X)$, with mean 1 and support Z. In expected terms, 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 ith order statistic of sampling four values from Z according to $F(X)$. Shorthand, we write $\mathbb{1} = X_{(1)}, \mathbb{2} = X_{(2)}$, etc. (In the following we let $F(X) = X/2$ and $Z = [0, 2]$, i.e. payoffs are uniformly distributed over $[0, 2]$, yielding $\mathbb{4} = 1.6, \mathbb{3} = 1.2, \mathbb{2} = 0.8, \mathbb{1} = 0.6$. However, *mutatis mutandis*, qualitative results hold for any symmetric CDF.)

Using the notation suggested by Bruns (2015), the set of games played by social types is $\Gamma = \{As, Ba, Ch, Cm, Co, Dl, Ha, Hr, Nc, Pc, Pd, Sh\}$. If no further assumptions are made, each of these games realizes as a game played by social types with equal probability (= 1/12, but see below).

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 S1 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 ($\neq \mathbb{1}$).

As can be seen from Table S1, M- and S-types choose the same strate-

gies in seven games (Ch, Cm, Co, Dl, Ha, Nc, and Pc). In the remaining five games, however, their choices differ: $\gamma = \{As, Ba, Hr, Pd, Sh\}$. Obviously, these five games are the ones decisive for the dynamics of the population. Therefore, we introduce an individual occurrence probability for each of them: p_{As}, p_{Ba}, p_{Hr}, p_{Pd}, and p_{Sh}, respectively, with $\sum_{i\in\gamma} p_i \leq 1$. Shorthand, we say that $\eta = (p_i)_{i\in\gamma}$ characterizes the *niche* that the population is inhabiting.

Finally, we assume that fitness is evaluated after animals have lived long enough, such that it is approximated sufficiently well by the expected payoffs given in matrix G.

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

Herein,

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

,

$$\pi_{S,M}(\eta) = \frac{50 - 29(p_{As} + p_{Pd} + p_{Sh}) - 8p_{Ba} - 15p_{Hr}}{35},$$

$$\pi_{M,S} = \frac{50 - 22p_{As} - 15p_{Ba} - 8(p_{Hr} + p_{Sh}) + 6p_{Pd}}{35},$$

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

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 obained by the types S and M in the games in Γ using $\mathbb{4} = 1.6, \mathbb{3} = 1.2, \mathbb{2} = 0.8, \mathbb{1} = 0.6$. The general form of G is derived in the supplements, S2.)

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 (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.

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

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

$$s^* = \frac{3p_{As} - 2p_{Ba} - p_{Hr} + p_{Pd} + p_{Sh}}{7p_{As} - p_{Pd} + 3p_{Sh}}. \tag{4}$$

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 trivial equilibria or is entirely equilibrial. The latter is the case if $\pi_{M,M}(\eta) = 1$, i.e. if

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

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

Second, we check for asymptotic stability of S and M. (Note that as $\pi_{S,S}(\eta) > 1$ always holds, L can never be stable.) Deriving conditions for the negativity of all eigenvalues of $J_\phi(S)$, we find that S is asymptotically stable as long as

$$p_{Pd} < 2p_{As} + p_{Ba} + \frac{1}{2}p_{Hr} + p_{Sh} =: p_{Pd}^{S}. \tag{6}$$

(Note that eq. 6 is equivalent to $\pi_{S,S}(\eta) > \pi_{M,S}(\eta)$.) Similarly, we find that

M is asymptotically stable as long as $p_{Pd} < p_{Pd}^{M/L}$ and

$$p_{Pd} > 2p_{Ba} - 3p_{As} + p_{Hr} - p_{Sh} =: p_{Pd}^{M}. \tag{7}$$

(Note that eq. 7 is equivalent to $\pi_{M,M}(\eta) > \pi_{S,M}(\eta)$.)

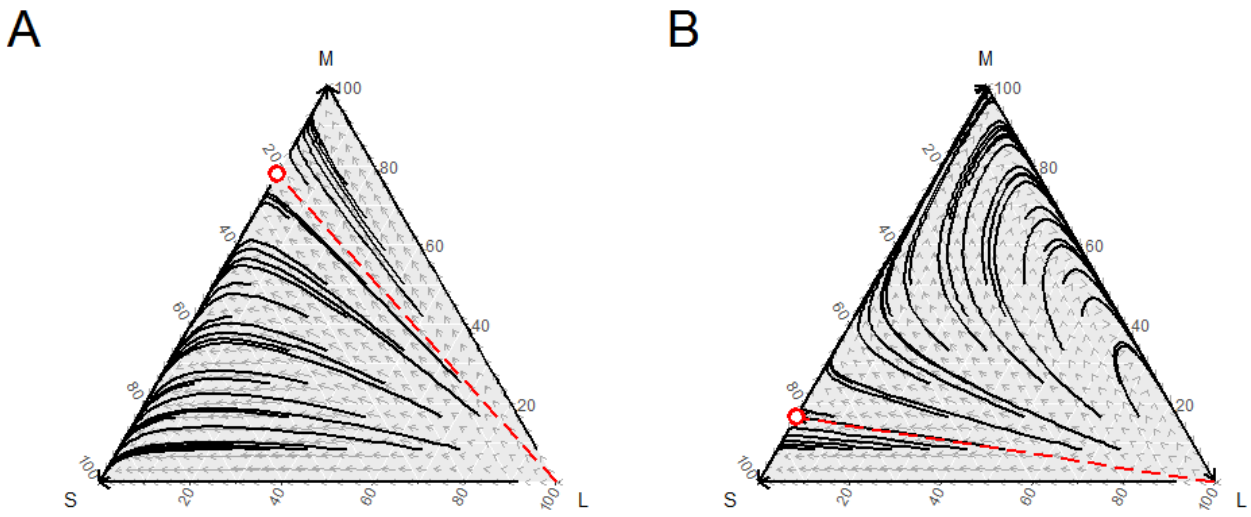

Figure 1: *Two illustrative dynamics.*
Panel A: $p_{As} = p_{Ba} = p_{Hr} = p_{Pd} = p_{Sh} = \frac{1}{12}$;
Panel B: $p_{As} = p_{Ba} = p_{Hr} = p_{Sh} = \frac{1}{10}, p_{Pd} = \frac{4}{10}$

Figure 1 illustrates evolutionary dynamics for two niches. Panel A of Fig. 1 shows the dynamics for $p_i = \frac{1}{12}, \forall i \in \gamma$, implying $p_{Pd}^{M} < p_{Pd} < p_{Pd}^{S}, p_{Pd}^{M/L}$, i.e. both S and M are stable (and $s^* = \frac{2}{9}$). Parameters in panel B are $p_{As} = p_{Ba} = p_{Hr} = p_{Sh} = \frac{1}{10}$, and $p_{Pd} = \frac{4}{10}$, implying $p_{Pd}^{M}, p_{Pd}^{M/L} < p_{Pd} < p_{Pd}^{S}$, i.e. S is stable, M is unstable (and $s^* = \frac{5}{6}$).

4.2 Finite populations

We have just seen that many niches exist in which S-types can 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 respective condition, $p_{Pd}^M < p_{Pd} < p_{Pd}^S$, 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. For these, 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 $p_{Pd} < p_{Pd}^M, p_{Pd}^S$ and $s^* < 1/3$. The latter condition holds if

$$p_{Pd} < \frac{6p_{Ba} - 2p_{As} + 3p_{Hr}}{4} =: p_{Pd}^{1/3}. \tag{8}$$

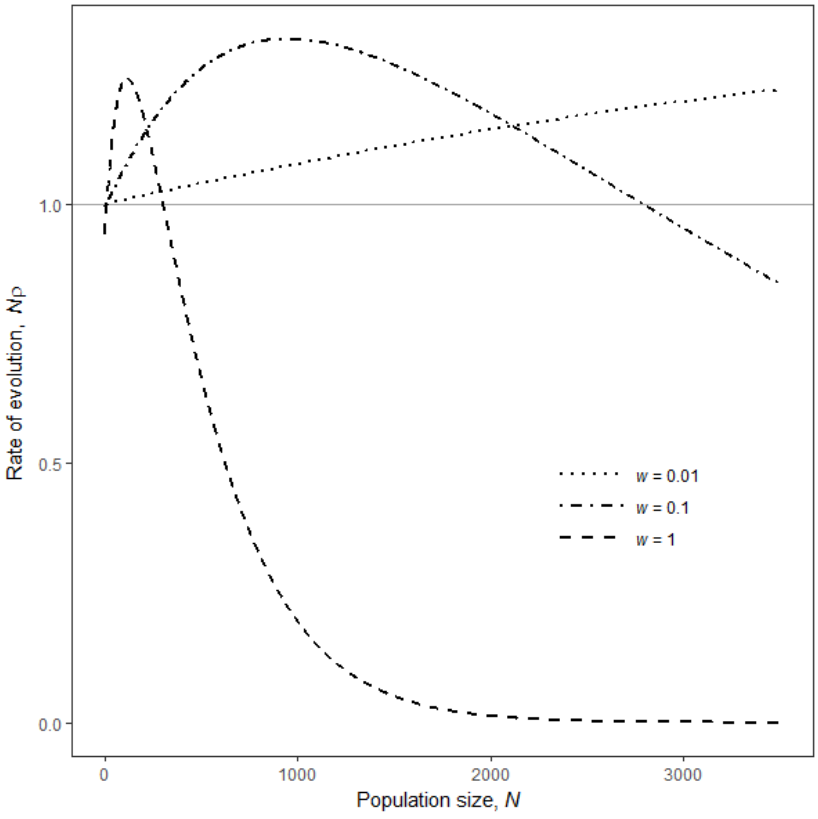

Figure 2: *Rates of evolution in finite populations of size N.* Parameters: $p_{As} = p_{Ba} = p_{Hr} = p_{Pd} = p_{Sh} = \frac{1}{12}$

More generally, for any strength of selection $w \in [0,1]$ and population size N, we can use the methods of Nowak 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

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

wherein $f_i = 1 - w + w\left[\pi_{S,S}(\eta)(i-1) + \pi_{S,M}(\eta)(N-i)\right]/\left[N-1\right]$ and $g_i = 1 - w + w\left[\pi_{M,S}(\eta)i + \pi_{M,M}(\eta)(N-i-1)\right]/\left[N-1\right]$. 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 Nth-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 $p_i = \frac{1}{12}, \forall i \in \gamma$. 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 $13 \leq N \leq 302$ 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, information about opponents' type or past behavior nor any of the other previously studied factors favoring the evolution of cooperativeness (see, e.g.: Nowak 2006). 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). The model presented here reconfirms this focus. In niches that are 'too dilemmatic', i.e. whenever p_{Pd} exceeds certain thresholds, collaboration does not evolve. However, the model also shows that there are 'quite dilemmatic' niches in which it still does (e.g.: for appropriate N and w, collaboration can evolve in finite populations inhabiting the niche characterized by $p_{As} = 0.05$, $p_{Ba} = 0.35$, $p_{Hr} = 0.01$, $p_{Sh} = 0.08$, and $p_{Pd} = 0.5$, i.e. a niche in which every second social interaction is a Pd).

Two, it may be deemed a weakness of collaboration that it cannot evolve in niches that are too dilemmatic. However, when we return to our opening 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 psychology and economics 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). The evidence collected in these studies points more in the direction of the latter conjecture, rendering maxims a promising formal tool for modeling choice behavior of this kind.

Finally, the model presented here has several limitations, including the following. One, only symmetric 2×2-games were studied. Two, the baseline payoff for the case of no social interaction was exogenously fixed at 1. Three, players were assumed to be unable to opt out of social interactions once payoffs have realized. Four, players' fitness was assumed to be approximated sufficiently closely by expected payoffs. Five, maxims were assumed to be inherited without mutations. Removing these limitations represents a promising task for future research.

References

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., et al. 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.

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.

Hauert, C., et al. 2007. "Via Freedom to Coercion: The Emergence of Costly Punishment". *Science* 316 (5833): 1905–1907. doi:10.1126/science.1141588.

Hauert, C., et al. 2002. "Volunteering as Red Queen mechanism for cooperation in public goods games". *Science (New York, N.Y.)* 296 (5570): 1129–1132. doi:10.1126/science.1070582.

Karpus, J., and M. Radzvilas. Forthcoming. "Team Reasoning and a Measure of Mutual Advantage in Games". *Economics and Philosophy.*

Newton, J. 2012. "Coalitional stochastic stability". *Games and Economic Behavior* 75 (2): 842–854. doi:10.1016/j.geb.2012.02.014.

— . 2017. "Shared intentions: The evolution of collaboration". *Games and Economic Behavior* 104:517–534. doi:10.1016/j.geb.2017.06.001.

Nowak, M. A. 2012. "Evolving cooperation". *Journal of Theoretical Biology* 299:1–8. doi:10.1016/j.jtbi.2012.01.014.

— . 2006. "Five rules for the evolution of cooperation". *Science* 314 (8 December): 1560–1563. doi:10.1126/science.1133755.

Nowak, M. A., et al. 2004. "Emergence of cooperation and evolutionary stability in finite populations". *Nature* 428 (6983): 646–650. doi:10.1038/nature02414.

Pennisi, E. 2005. "How Did Cooperative Behavior Evolve?" *Science* 309 (5731): 93. doi:10.1126/science.309.5731.93.

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.

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., et al. 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.

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, et al. 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.

Tomasello, M. 1999. *The Cultural Origins of Human Cognition*. Cambridge: Harvard University Press.

— . 2009. *Why we cooperate*. Cambridge: MIT Press.

Tomasello, M., et al. 2012. "Two Key Steps in the Evolution of Human Cooperation". *Current Anthropology* 53 (6): 673–692. doi:10.1086/668207.

Tomasello, M., et al. 2005. "Understanding and sharing intentions: The origins of cultural cognition". *Behavioral and Brain Sciences* 28 (5): 675–691. doi:10.1017/S0140525X05000129.

West, S. A., C. El Mouden, and A. Gardner. 2011. "Sixteen common misconceptions about the evolution of cooperation in humans". *Evolution and Human Behavior* 32 (4): 231–262. doi:10.1016/j.evolhumbehav.2010.08.001.

West, S. A., A. S. Griffin, and A. Gardner. 2007. "Evolutionary Explanations for Cooperation". *Current Biology* 17 (16): R661–R672. doi:10.1016/j.cub.2007.06.004.

Supplements

S1: The twelve strict symmetric ordinal 2×2-games

Ch	L	R
U	$\underline{\lVert 3,3 \rVert}$	2, 4
D	4, 2	1, 1

Cm	L	R
U	$\underline{\lVert 3,3 \rVert}$	4, 2
D	2, 4	1, 1

Co	L	R
U	$\underline{\lVert 4,4 \rVert}$	2, 1
D	1, 2	3, 3

Dl	L	R
U	$\underline{\lVert 3,3 \rVert}$	4, 1
D	1, 4	2, 2

Ha	L	R
U	$\underline{\lVert 4,4 \rVert}$	3, 2
D	2, 3	1, 1

Nc	L	R
U	$\underline{\lVert 4,4 \rVert}$	2, 3
D	3, 2	1, 1

Pc	L	R
U	$\underline{\lVert 4,4 \rVert}$	3, 1
D	1, 3	2, 2

As	L	R
U	$\lVert 4,4 \rVert$	1, 2
D	2, 1	$\underline{3,3}$

Ba	L	R
U	$\underline{2,2}$	$\lVert 3,4 \rVert$
D	$\lVert 4,3 \rVert$	1, 1

Hr	L	R
U	$\underline{2,2}$	$\lVert 4,3 \rVert$
D	$\lVert 3,4 \rVert$	1, 1

Pd	L	R
U	$\lVert 3,3 \rVert$	1, 4
D	4, 1	$\underline{2,2}$

Sh	L	R
U	$\lVert 4,4 \rVert$	1, 3
D	3, 1	$\underline{2,2}$

Table S1: Overview of the 12 strict symmetric ordinal 2×2-games; underlined profiles are reached by M-types, profiles in norm dashes ($\lVert \bullet, \bullet \rVert$) are reached by S-types

S2: 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 $\gamma = \{As, Ba, Hr, Pd, Sh\}$. Take the example of the Pd. It realizes with probability p_{Pd}. S-types obtain 3 when playing against other S-types and 1 when playing against M-types. Conversely, M-types obtain 4 when matched with an S-type and 2 when matched with another M-type. Payoffs for the other games in γ are calculated analogously to the Pd example just given,

resulting in

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

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

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

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