Routledge
Taylor & Francis Group

Journal of Economic Methodology

ISSN: 1350-178X (Print) 1469-9427 (Online) Journal homepage: https://www.tandfonline.com/loi/rjec20

The long-term viability of team reasoning

S.M. Amadae & Daniel Lempert

To cite this article: S.M. Amadae & Daniel Lempert (2015) The long-term viability of team reasoning, Journal of Economic Methodology, 22:4, 462-478, DOI: 10.1080/1350178X.2015.1024880

To link to this article: https://doi.org/10.1080/1350178X.2015.1024880

Journal of Economic Methodology, 2015
Vol. 22, No. 4, 462–478, http://dx.doi.org/10.1080/1350178X.2015.1024880

The long-term viability of team reasoning

S.M. Amadae[a]* and Daniel Lempert[b,1]

[a]*Ohio State University, 2126 Derby Hall, 154 N Oval Mall, Columbus, OH 43210, USA;* [b]*SUNY Potsdam, 44 Pierrepont Ave., Potsdam, NY 13676, USA*

(*Received 15 August 2014; accepted 28 August 2014*)

Team reasoning gives a simple, coherent, and rational explanation for human cooperative behavior (Bacharach 1999; Sugden 1993). This paper investigates the robustness of team reasoning as an explanation for cooperative behavior, by assessing its long-run viability. We consider an evolutionary game theoretic model in which the population consists of team reasoners and 'conventional' individual reasoners. We find that changes in the ludic environment can affect evolutionary outcomes, and that in many circumstances, team reasoning may thrive, even under conditions that, at first glance, may seem unfavorable. We also pursue several extensions that augment the basic account, and conclude that team reasoning is an evolutionarily viable mechanism with the potential to explain behavior in a range of human interactions.

Keywords: team reasoning; rational choice; replicator dynamic; individual maximization; evolutionary viability; rationality; ludic ecology; Prisoner's dilemma

1. Introduction

When a player team reasons, she plans to act so as to maximize a team payoff function – for example, the sum of all team members' individual payoffs (below, we refer to a team reasoner with the feminine pronoun and the individual reasoner with the masculine pronoun). In contrast, the individual reasoner, the conventional rational actor, acts to maximize only his own individual payoffs. This paper investigates the plausibility of team reasoning as an explanation for cooperative behavior, by analyzing an evolutionary game theoretic model in which the population consists of team reasoners and individual reasoners. We find that the degree of ludic diversity – in particular, changes in the mix of games played – can affect evolutionary outcomes; team reasoning and thus cooperation may thrive, even under circumstances that, at first glance, may seem unfavorable to each. Our results add to existing theoretical and empirical work that suggests that team reasoning is a plausible and coherent mechanism for explaining human decision-making.

2. Team reasoning

There are two simple games that seem to present a puzzle for individual reasoning-based 'orthodox' decision theory (e.g., Bacharach, 2006, pp. 35–68; Gold & Sugden, 2007, pp. 281–285). First, consider the Hi Lo (see Table 1). The intuitively compelling choice is Hi, and (Hi,Hi) is the Pareto-optimal equilibrium. But individual reasoning does not require this choice. A rational player, (one who acts to maximize his expected individual

*Corresponding author. Email: amadae.1@osu.edu
This article was originally published with errors. This version has been amended. Please see Erratum (http://dx.doi.org/10.1080/1350178X.2015.1083142).

Table 1. An example of a Hi Lo.

	Hi	Lo
Hi	3; 3	1; 1
Lo	1; 1	2; 2

Note: Payoffs are listed as (row player; column player).

payoffs) playing another rational player, under the assumption that the players' rationality is common knowledge, is only entitled to conclude that *I should choose Hi if my opponent selects Hi, and Lo if he chooses Lo* (Gold & Sugden, 2007). Second, consider the Prisoner's dilemma (PD) (see Table 2). It is well known that individual reasoning (individual payoff maximization) mandates the choice of D; yet, many people have the intuition that C is the correct choice. Team reasoning justifies these intuitions.

Bacharach (1999, 2006) and Sugden (1993, 2000) propose team reasoning as an alternative account of how people make decisions when interacting with others. (There is also a related, though not overlapping, literature in philosophy, including (Gilbert, 1989; Hollis, 1998; Hurley, 1989; Regan, 1980.) When a player team reasons, instead of asking (as in the standard account), 'what should *I* (as an individual) do?' she asks, 'what should *we* (as a team) do?' (see, e.g., Gold & Sugden, 2007) She answers the latter question by 'work[ing] out the best feasible combination of actions for all members of her team' (Bacharach, 2006). It is convenient (though not strictly required by the theory) to make the simplifying assumption that the 'best feasible combination of actions' is that which leads to the outcome maximizing the sum of team members' individual payoffs. Finally, she takes the action that the 'best feasible combination of actions' requires of her; in other words, she chooses the strategy prescribed for her in the team utility-maximizing strategy profile.

Team reasoning unambiguously leads to the choice of Hi in the Hi Lo, and the choice of Cooperate (C) in the PD, since the combination of actions that maximizes combined individual payoffs is (Hi, Hi) in Hi Lo, and (C,C) in the PD. Thus, team reasoning solves, in the Hi Lo, an equilibrium selection problem that is theoretically problematic for individual reasoning; in the PD, it leads to an outcome that is Pareto-preferable to the (D,D) equilibrium outcome that results when individual reasoners play. In addition to its theoretical appeal, team reasoning is also consistent with much of the observed cooperative behavior in the laboratory (e.g., Camerer, Loewenstein, & Rabin, 2003; Colman, Pulford, & Rose, 2008) and in the field (e.g., Heinrich et al., 2005). And evidence from social psychology (see e.g., Kramer & Brewer, 1984) shows that when players' shared social identity is primed, cooperative behavior in social dilemmas increases; this is of particular interest since Bacharach (1999) argues that social identification with a co-player will cause one to engage in team reasoning.

Despite its theoretical soundness and supporting evidence, there are still grounds for some skepticism about team reasoning. One ground is that it seems to require behavior that is potentially self-sacrificial. Is the behavior implied by team reasoning viable in the long run? Scholars across disciplines have relied on evolutionary theory and models to contest the notion that behavior consistent with team reasoning is truly viable, or, at the

Table 2. An example of a PD.

	Cooperate	Defect
Cooperate	3; 3	1; 4
Defect	4; 1	2; 2

Note: Payoffs are listed as (row player; column player).

least, they have used these tools to cast doubt on team reasoning as a mechanism for bringing about such behavior.[2] In response, we analyze an evolutionary game theoretic model that tests (and ultimately shows) the viability of team reasoning as a 'strategy.' First, we explain our modeling decisions.

3. Reasoning, evolutionary models, and the ludic ecology

In our baseline model, we will consider two types of reasoners: the individual reasoner and the team reasoner. Does it make sense to consider 'types of reasoners' in an evolutionary game-theoretic context? We suggest, relying on arguments made and reviewed by Bacharach (2006), that it does. Bacharach points out that, in addition to specifying which *traits* will be evolutionarily selected, an evolutionary model should also speak to the *mechanism* that will be favored. A mechanism explains a 'repertoire of dispositions,' that is, a set of traits – one for each decision-making context. This leads to a second notable feature of our model. Our players navigate a 'ludic ecology' (Bacharach, 2006) that consists of both a common-interest game (Hi Lo) and a social dilemma (the PD). As Bacharach (2006) notes, this is in contrast to the 'standard models in bio-evolutionary game theory,' which consider one game at a time, and thus have minimal ludic diversity [though, of course, we still will not capture the full range of social interactions that humans engage in, the Hi Lo and the PD are representative of many important decision-making contexts (see e.g., Bacharach, 2006)[3]]. Team reasoning and individual reasoning can serve as mechanisms for choice in both of these contexts; and importantly, they are both *simple* mechanisms. Bacharach (2006) notes that, in general, parsimonious ('low-cost') mechanisms will be favored over those that are more complex. Thus, in our baseline model, we will take our competing mechanisms to be team reasoning and individual reasoning.[4] (In an extension, we will explicitly assess the viability of a more complex, and costly, mechanism.)

Before formalizing the model, we emphasize one more feature of our analysis. We will consider a population where interactions are one-shot, and random. Given this absence of assortative pairing, group selection cannot operate (e.g., Sober & Wilson, 1998). This is interesting because Bacharach (2006) hypothesizes that group selection is the means by which team reasoning may thrive (see also closely related discussion in Caporael, 2007). By avoiding the contested concept of group selection, we will be able to place team reasoning on firmer (or at least alternative) footing.[5] We turn now to the model, for which Table 3 summarizes notation.

Table 3. Definition of symbols used in baseline model.

Symbol	Definition
h	Proportion of time individual reasoners play Hi in Hi Lo
p	Proportion of individual reasoners in population
p^*	Equilibrium proportion of individual reasoners, where $W(I) = W(T)$
Δp	Change in proportion of individual reasoners between time periods
$V(i\|j)$	Average payoff to generic type i interacting with type j
$W(I)$	Average payoff to individual reasoner
$W(T)$	Average payoff to team reasoner
w_0	Baseline payoff shared by each type
x	Proportion of games that are Hi Lo

4. Team reasoning and individual reasoning: an evolutionary analysis

Consider a population that consists of two types of players: team reasoners and individual reasoners. We refer to these below simply as 'types.' Suppose that members of the population are randomly paired to play one-shot variants of two games: Hi Lo, a pure coordination game with payoffs as given in Table 4, is played x proportion of the time, while the (additive) PD, as in Table 5, is played $1 - x$ proportion of the time. Recall that because the team reasoner asks 'what should *we* do given *our* standard of success?' she plays Cooperate in the PD and Hi in the Hi Lo. The individual reasoner (who asks, 'what should *I* do given *my* standard of success?') always defects in the PD. However, since (Lo, Lo) and (Hi, Hi) are equally valid equilibria for the individual reasoner in Hi Lo, and thus neither Hi nor Lo are mandated as strategies, suppose that the individual reasoner plays Hi with probability h. Table 6 then gives the expected payoffs, $V(i|j)$, for each pairwise interaction (for example, $V(I|T)$ is the expected payoff to an individual reasoner interacting with a team reasoner).

To assess how changes in the set of games played by the population impacts evolutionary outcomes, we follow standard practice in evolutionary game theory, and use the replicator dynamic (or proportional fitness rule; see e.g., Boyd & McElreath, 2007). This formula describes how the proportion of two competing strategies in the population changes from one time period to the next, as a function of their respective payoffs and initial proportions in the population.[6] In general, for types A and B, where p is the proportion of A types in the population at time t and $W(i)$ is the average payoff to type i, the

Table 4. Hi Lo: played x proportion of the time.

	Hi	Lo
Hi	$\beta;\beta$	$-\gamma;-\gamma$
Lo	$-\gamma;-\gamma$	$0;0$

Note: $\beta > 0,\ \gamma > 0$. Payoffs are listed as (row player; column player).

Table 5. PD: played $1 - x$ proportion of the time.

	Cooperate	Defect
Cooperate	$b-c;b-c$	$-c;b$
Defect	$b;-c$	$0;0$

Note: $b > c > 0$. Payoffs are listed as (row player; column player).

Table 6. Expected payoffs for four types of interactions.

Pairing	$V(i\|j)$: expected payoff
$(\boldsymbol{T}\|\boldsymbol{I})$	$-c(1-x)+x(\beta h-(1-h)\gamma)$
$(\boldsymbol{T}\|\boldsymbol{T})$	$(b-c)(1-x)+\beta x$
$(\boldsymbol{I}\|\boldsymbol{T})$	$b(1-x)+x(\beta h-(1-h)(\gamma))$
$(\boldsymbol{I}\|\boldsymbol{I})$	$x(\beta h^2-2\gamma h(1-h))$.

Note: The table gives payoff to i for interaction $i|j$. For example, the first row gives the expected payoff for a team reasoner interacting with an individual reasoner.

change in the proportion of A types between t and $t+1$, Δp, is given by the *difference equation*:

$$\Delta p = p(1-p)\frac{W(A)-W(B)}{pW(A)+(1-p)W(B)}. \tag{1}$$

The average payoffs to our types, $W(I)$ and $W(T)$, are as follows. Let p stand for the proportion of individual reasoners in the population. Then, since interactions are random, p is the probability of interacting with an individual reasoner and $1-p$ is the probability of interacting with a team reasoner. Therefore, the average payoff for each type is the sum of its two pairwise payoffs, weighted by the probability that each type of interaction occurs, plus a common baseline fitness (w_0): $W(I) = p(V(I|I) + (1-p)V(I|T) + w_0$ and $W(T) = pV(T|I) + (1-p)V(T|T) + w_0$.

Equilibria exist where $\Delta p = 0$ – where the proportion of each type is constant from one period to the next. Clearly, $\Delta p = 0$ when $p = 0$ or $p = 1$ – i.e., when the population consists of only team reasoners or only individual reasoners. The interesting question is how the variation in the set of games played impacts the *stability* of these equilibria. An equilibrium is called *stable* if, when the equilibrium mix of types is disturbed slightly, it returns to the equilibrium value. For example, to assess whether (and when) the all-team reasoner equilibrium ($p = 0$) is stable, we need to analyze what happens when a rare individual reasoner enters the population of team reasoners. Substantively, the key is to notice that whether an 'invasion' of individual reasoners is successful depends only on the two types' relative success against team reasoners. Thus, the all-team reasoner equilibrium is stable where $V(T|T) > V(I|T)$. Solving

$$(b-c)(1-x) + \beta x > b(1-x) + x(\beta h - (1-h)(\gamma))$$

for x, the proportion of games that are Hi Lo, we find that the inequality holds where

$$x > \frac{c}{(\beta+\gamma)(1-h)+c}.$$

For relatively high values of x, then, the all-team reasoning equilibrium is stable, and individual reasoners cannot invade.

We analyze the stability of the all-individual reasoner equilibrium similarly, by assessing the two types' relative performance against individual reasoners. Where $V(I|I) > V(T|I)$, the all-individual reasoner equilibrium ($p = 1$) is stable. The inequality holds where

$$\begin{cases} x < \frac{c}{\beta(1-h)h+\gamma(1-h)(2h-1)+c} & \text{if} \quad h \geq \frac{\gamma}{\beta+2\gamma} \\ \forall\, x & \text{if} \quad h < \frac{\gamma}{\beta+2\gamma} \end{cases}$$

Thus, for relatively low x values – where the PD is played relatively frequently – a population of individual reasoners can resist invasion by team reasoners. Also, note that for very low values of h – where individual reasoners play Lo frequently – there is no value of x for which team reasoners can invade. Intuitively, this is because when the population is coordinating (mostly) on a low-payoff equilibrium, a team reasoner cannot gain advantage in the Hi Lo by playing her role in a higher payoff equilibrium. We can see another reason, then, that the choice of strategy in Hi Lo is nontrivial for individual reasoners: choosing Lo often can actually be advantageous in some evolutionary contexts, as it forestalls the possibility of invasion by team reasoners for all values of x.

The third case in which $\Delta p = 0$ is where the numerator in the difference equation, $W(I) - W(T)$, equals zero. In this case, since the average fitness of each type is the same, the proportion of each type does not change. Such an equilibrium, where a mix of both types exists, is called an internal (or polymorphic) equilibrium. Solving the equation $W(I) = W(T)$ for p – the proportion of individual reasoners in the population – yields a unique solution:

$$p^* = \frac{(1-h)(\beta+\gamma) + c - (c/x)}{(1-h)^2(\beta+2\gamma)}.$$

Since p^* is a proportion, it must be between zero and one. (Whenever p^* is $\notin (0,1)$, there is no meaningful equilibrium where $W(I) = W(T)$.) For what values is this the case? First, note that because the denominator is always positive for $h < 1$ and $\beta, \gamma > 0$, p^* is greater than or equal to zero when the numerator is non-negative. This holds if $(1-h)(\beta+\gamma) + c \geq c/x$, or solving for x, where $x \geq c/[(1-h)(\beta+\gamma)+c]$. The second requirement is that $p^* \leq 1$. Solving for x, we find the inequality holds where

$$x \leq \frac{c}{\beta(1-h)h + \gamma(1-h)(2h-1) + c}.$$

Note that these constraints are the same as those that determined the stability of the equilibria at $p = 0$ and $p = 1$. Precisely, there exists an internal equilibrium (i.e., where $W(I) = W(T)$ and $p^* \in (0,1)$) if and only if both the all-individual reasoner and the all-team reasoner equilibria are stable: when

$$\frac{c}{(1-h)(\beta+\gamma)+c} < x < \frac{c}{\beta(1-h)h + \gamma(1-h)(2h-1) + c}.$$

An internal equilibrium thus exists for relatively moderate values of x. The three cases described are graphed in Figure 1.

In all cases the internal equilibrium is unstable. To see this, note that an internal equilibrium only exists if the equilibria at $p = 0$ and $p = 1$ are stable (and so Δp is decreasing around $p = 0$, where it takes on the value of 0, and Δp is increasing around $p = 1$, where it takes on the value of 0). Because Δp is continuous in p, the intermediate value theorem implies that Δp is negative for $p \in (0, p^*)$, and positive for $p \in (p^*, 1)$. This means that the internal equilibrium is unstable: intuitively, if a few 'extra' individual reasoners beyond the equilibrium value enter the population ($p > p^*$), then this leads to further increases in the proportion of individual reasoners ($\Delta p > 0$); similarly, if a few 'extra' team reasoners enter the population, the proportion of team reasoners increases further. An unstable equilibrium is of interest because it defines the *basins of attraction* of two stable equilibria – here, the stable equilibria at $p = 0$ and $p = 1$. For initial p values greater than p^*, the population tends toward the $p = 1$ (all-individual reasoner) equilibrium; for values of p less than p^*, the population tends toward the equilibrium at $p = 0$. Below, we discuss the interpretation of this baseline model, and consider a few extensions.

5. Discussion and extensions

In our baseline model, our central result is that whether team reasoning can emerge over the long run depends on the relative frequency of the PD and Hi Lo in the following way. There are three intervals of interest in the frequency that the PD is played. In the first

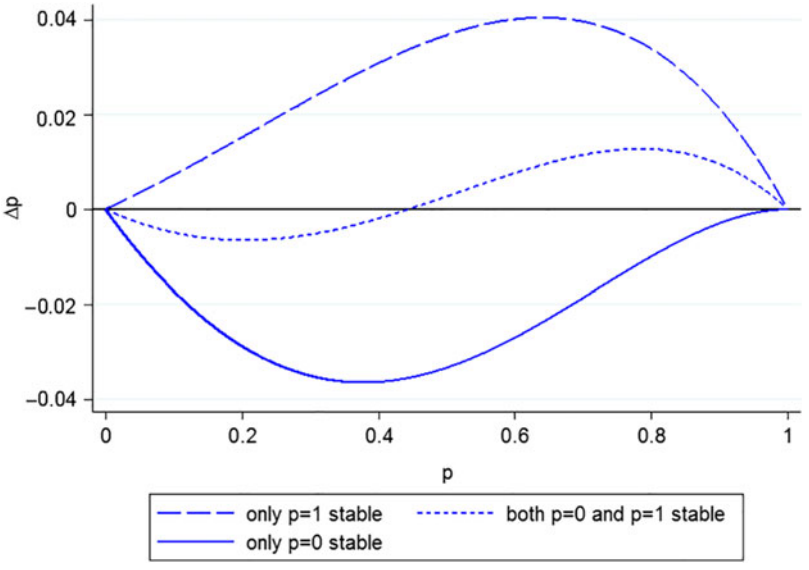

Figure 1. Variation in the set of games played can impact the stability of equilibria. The graph shows Δp as a function of p, for three values x: (1) where only the all-team reasoner equilibrium ($p = 0$) is stable, (2) where only the all-individual reasoner equilibrium is stable ($p = 1$), and (3) where both the all-team reasoner ($p = 0$) and the all-individual reasoner ($p = 1$) equilibria are stable. The specific values used to generate the graphs are $w_0 = 2$, $b = 2\,c = 1$, $\beta = 1$, $\gamma = 1$, $h = .5$ and $x = .4$ (only $p = 1$ stable), $x = .6$ ($p = 0$ and $p = 1$ stable), $x = .8$.

interval, in which the PD is played 'frequently,' individual reasoners will, in the long run, become the sole type in the population, regardless of their initial proportion in the population. In the second interval, in which the PD is played 'somewhat frequently,' whether individual reasoners emerge as the sole dominant type depends on whether they exceed an initial threshold proportion. If the initial proportion of individual reasoners exceeds the threshold proportion, they will emerge as the sole dominant type. If the initial proportion of individual reasoners is below the threshold proportion, then team reasoners will emerge in the long run as the sole type. Finally, in the third interval, where the PD is played infrequently, team reasoners will emerge as the sole type regardless of initial mix of strategies, with the following caveat. If individual reasoners who coordinate frequently on playing Lo in Hi Lo make up a very large percentage of the initial population, then individual reasoners will emerge as the sole dominant type in the long run.

Thus, holding all else constant, the more frequent are Hi Lo games compared to PD interactions, the better team reasoners do. In an environment where coordination is more important than competition, team reasoning is particularly advantaged. The caveat we note above does qualify this conclusion a bit. For any mix of games, if a very large percentage of the population coordinates on Lo, team reasoners will not be successful, and will be driven out of the population in the long run. Thus, an evolutionary argument suggests the nontriviality of the individual reasoner's choice in Hi Lo: only coordination on Lo by individual reasoners guarantees that team reasoners will not be able to establish a foothold in a population initially dominated by individual reasoners. We now consider how this basic account may be modified by re-interpretation and extensions of our model.

A note on the assumption that x *is fixed.* In our baseline analysis, we have treated x, the proportion of games that are Hi Lo, as fixed, and found no stable internal equilibrium. But

it is worth noting that x may change between time periods. As the baseline analysis shows, the value of x impacts the location of the basins of attraction of the equilibria at $p = 0$ and $p = 1$. It is straightforward to see that, in a system that is not at equilibrium, changing the value of x may change the sign of Δp: a system in which Δp is negative at t may have Δp positive at $t + 1$, if x decreases between t and $t + 1$. That is, (depending on changes in x) it is possible that, even in the absence of a stable internal equilibrium, p reaches neither 0 nor 1.

The h *parameter as an error rate.* Our baseline model does not consider the possibility that players make errors. Though in the abstract, the Hi Lo and the PD are simple games, we are interested in the games as models of real-life interactions. The structure of such interactions will sometimes be less transparent than our theoretical model suggests. However, we can incorporate one specific type of player error without complicating the analysis, simply by reinterpreting the h parameter in the model.

In particular, consider the possibility that, in a given interaction, players may be uncertain about payoffs. Table 7 models such a situation – the players (correctly) perceive that a is greater than 0, but the values of y and z are unclear. Note that if y and z are both less than 0, the game is a Hi Lo, but if y is greater than a and z is less than 0, the game is a PD. We suggest that the choice for the team reasoner is clearer than it is for the individual reasoner: as long as she perceives that $2a > y + z$, she will choose A. But this information is not enough for the individual reasoner, who – in addition – needs to accurately assess whether $y > a$ and $z < 0$ (in which case the game is a PD and he plays B), whether $y < 0$ and $z > a$ or $0 < y, z < a$ (in which cases he plays A), or whether $y < 0$ and $z < 0$ (in which case the game is a Hi Lo).

For a concrete example, suppose Player 1and Player 2 can each choose to (A) go out and hunt for a resource, splitting the proceeds evenly or (B) sit at home and wait to steal anything the other player brings back. As long as the effort put into hunting is outweighed by benefit accorded by securing the prey, it is clear that the strategy profile (A,A) Pareto-dominates (B,B). But Player 1 is an individual reasoner, so he calculates that he will be better off if he stays at home (expending no effort in hunting), and then steals any prey Player 2 returns with (thereby gaining the resource regardless). Upon her return from the successful hunt, however, Player 2 takes offense at Player 1's attempted theft, and, in the ensuing tussle, both players are grievously injured. Player 1 has just mistaken a Hi Lo for a PD. Note also that a team reasoning Player 2 chooses A (to hunt), even if she makes *the same* mistake in perception as Player 1.

Such mistakes can be incorporated into our baseline model simply by reinterpreting the h parameter. Even supposing that individual reasoners always *intend* to play Hi in Hi Lo, $1 - h$ can be interpreted as the proportion of time individual reasoners mis-perceive the Hi Lo as a PD.[7] Therefore, the existence of such errors implies that there is always some value of x for which team reasoners can invade a population of individual reasoners.

Cognitive load. Here, we substitute for the individual reasoner a type – which we call the flexible reasoner – who team reasons in a Hi Lo (always playing Hi), and reasons

Table 7. A game with ambiguous payoffs, $a > 0$.

	A	B
A	$a; a$	$z; y$
B	$y; z$	$0; 0$

Note: Payoffs are listed as (row player; column player).

individually in a PD (always defecting). Absent more, it is clear that such a player will do better than the team reasoner (except when all games are Hi Lo). However, the flexible reasoner's strategy, which involves switching back and forth between modes of reasoning, is more complex than that of the team reasoner, and thus places a greater cognitive load on the flexible reasoner.[8]

We incorporate the cost that this increased cognitive load takes by subtracting a cost parameter $k > 0$, weighted by $x(1-x)$, from the payoffs of the flexible reasoner. (For example, $V(F|T)$ is $b(1-x) + x\beta - kx(1-x)$.) The logic is that the more even is the mix of games, the more frequent is the need to switch modes of reasoning, and therefore, the greater the cognitive load for the flexible reasoner.

This set-up gives a very simple solution. Regardless of the proportion of the types in the population, $W(F) > W(T)$ if $x < c/k$, $W(F) = W(T)$ if $x = c/k$, and $W(F) < W(T)$ if $x > c/k$. Thus, for any initial p, the population tends toward the all-team reasoner equilibrium if $x > c/k$, and toward the all-flexible reasoner equilibrium if $x < c/k$. So, as before, the higher the proportion of games that are Hi Lo, the more advantaged are team reasoners. Here, the critical value of x is determined solely by the relative magnitudes of the c and k parameters.

When individual reasoners' play varies based on the proportion of types in the population. A relatively minor modification of our baseline model results in the *possibility* of multiple internal equilibria. In particular, suppose that individual reasoners become more likely to play Hi as the proportion of team reasoners (who always play Hi) in the population increases. For example, suppose $h = 1 - p$. Then, the numerator of the difference equation is a fifth degree polynomial in p, and thus the function can take on a value of 0 for up to three values of $p \in (0, 1)$, multiple internal equilibria may exist, at least one of which may be stable. A concrete example is shown in Figure 2. In this

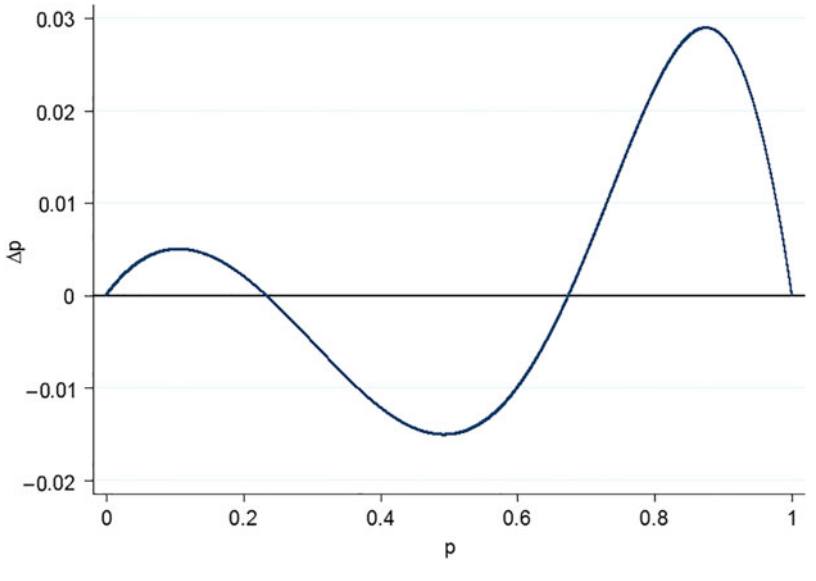

Figure 2. Multiple internal equilibria can exist when h is a function of p. The graph shows an example with both stable and unstable internal equilibria. Here, $w_0 = 2$, $b = 2$, $c = 1$, $\beta = 1$, $\gamma = 1$, $h = 1 - p$, and $x = .7$.

particular case, there are two internal equilibria, and a stable equilibrium with $p \approx .2$ exists.

Circumspect team reasoning. We now consider team reasoners who are circumspect (Bacharach, 1999, 2006). In our baseline model, team reasoners choose to play their role in the strategy profile that is best for the team, under the assumption that their co-player is also a team reasoner. A team reasoner who is circumspect, however, considers the possibility that a co-player may 'fail'(for any reason) to play his role in the strategy profile that is best for the team, and instead performs some 'default' action. The circumspect team reasoner thus calculates, *in light of probability that a team member fails*, the strategy profile that is best for the team (*including* the potentially failing team member), and executes the move that corresponds to her choice in that profile. Below, we sketch a version of the baseline model in which all team reasoners are circumspect. Our results indicate that key features of our baseline model carry over: most importantly, team reasoners perform relatively better when Hi Lo is played relatively often. At the same time, circumspect team reasoners perform (weakly) better than 'un-circumspect' team reasoners against individual reasoners – in fact, for a certain class of PDs, circumspect team reasoners can always invade a population of individual reasoners; we also find that the population can remain polymorphic indefinitely, even when there is not, technically, a stable internal equilibrium.

We take the probability that a team member fails to be p, the proportion of individual reasoners in the population; following Bacharach (1999), we assume that a circumspect team reasoner accurately assesses this probability. The default actions of a failing team member are that of the individual reasoner. In this extension we generalize the PD in our baseline model, making the payoffs from mutual cooperation $(a; a)$, with $a > 0$, $2a > b - c$ instead of $(b - c; b - c)$ (see Table 8).[9]

The circumspect team reasoner determines the strategy profile that maximizes the sum of the individual payoffs, given the failure probability p, and chooses the strategy it prescribes for her. If the failure probability is 0, utility-maximizing strategy profile – from the perspective of the team reasoner – is (C,C), as in our baseline model (since, by definition, in the PD $2a > b - c$, 0). Now, suppose first that $b - c > 0$. Then, the team reasoner ranks the profiles (C,C) > (C,D) = (D,C) > (D,D). This implies that for any failure probability, the team reasoner cooperates, in an attempt to implement (C,C); even in the event of certain failure on the part of the co-player, the team is better off if the team reasoner cooperates. Thus, if $b - c > 0$, the team reasoner's behavior in the PD is no different than in our baseline model. But suppose that $b - c < 0$. Then, the team reasoner ranks the profiles (C,C) > (D,D) > (C,D) = (D,C). The circumspect team reasoner calculates the value of p at which implementing (C,C) yields greater utility for the team than implementing (D,D), i.e., where $2p(1 - p)(b - c) + (1 - p)^2 2a > 0$. Solving for p,

Table 8. A generalized PD.

	Cooperate	Defect
Cooperate	$a; a$	$-c; b$
Defect	$b; -c$	0; 0

Note: $a, b, c > 0$, $2a > b - c$, payoffs are listed as (row player; column player).

the circumspect team reasoner cooperates iff

$$p < \frac{a}{a-(b-c)} \equiv p_{pd}.$$

The circumspect team reasoner proceeds similarly for Hi Lo, calculating the value of p at which implementing (H,H) yields greater utility than implementing (L,L), and thus plays Hi for all p, for $h > .5$, and for $h < .5$, iff

$$p < \frac{\beta}{(\beta+2\gamma)(1-2h)} \equiv p_{hl}.$$

To facilitate exposition of the evolutionary model, define $f(p) \equiv W(I) - W(T)$, the numerator on the right side of the difference Equation (1); note that $f(p)$ determines the sign of the right-hand side of (1) for $p \in (0,1)$ (which, in turn, determines the stability of equilibria). The formulas for $f(p)$ differ based on the play of the team reasoner, and are as follows. When the team reasoner plays Cooperate and Hi,

$$f(p) = p[c(1-x) - x(1-h)(\beta h + 2\gamma h - \gamma)] + (1-p)[(b-a)(1-x) - x(1-h)(\beta+\gamma)]. \tag{2}$$

When the team reasoner plays Defect and Hi,

$$f(p) = px(1-h)(\gamma - 2\gamma h - \beta h) - (1-p)x(1-h)(\beta+\gamma). \tag{3}$$

When the team reasoner plays Cooperate and Lo,

$$f(p) = p[c(1-x) + xh(\beta h + 2\gamma h - \gamma)] + (1-p)[(b-a)(1-x) - xh\gamma]. \tag{4}$$

When the team reasoner plays Defect and Lo,

$$f(p) = xh[p(\beta h + 2\gamma h - \gamma)) - (1-p)(\gamma)]. \tag{5}$$

Case 1: $p_{pd} \in (0,1)$, but $p_{hl} \notin (0,1)$ (implying that $h > .5$ and $c > b$).

On $[0, p_{pd}]$, $f(p)$ is given by (2). Let x^* be the value of x for which $f(0) = 0$.[10] Because $f(p)$ is increasing in p, $f_{x^*}(p') > 0$ for any $p' \in (0, p_{pd})$. Set $x = x^* + \varepsilon$ (while holding other parameters constant). Since $f(p)$ is decreasing in x, $f_{x^*+\varepsilon}(0) < 0$; because $f(p)$ is continuous in x, $f_{x^*+\varepsilon}(p') > 0$, for sufficiently small ε. Thus, there exists a set of parameter values for which a stable equilibrium at $p = 0$, and an unstable internal equilibrium at $p^* < p_{pd}$ exist. Similarly, setting $x = x^* - \varepsilon$ gives a set of parameter values for which an unstable equilibrium at $p = 0$ exists, and no internal equilibrium $\in (0, p_{pd})$ exists. Finally, observe that $f(1) < 0 \,\forall\, x > c/[c + (1-h)(\beta h + 2\gamma h - \gamma)]$, so (since certainly $p_{pd} < 1$), there exists a set of parameter values for which a stable equilibrium at $p = 0$ exists, and no internal equilibrium $\in (0, p_{pd})$ exists.

On $(p_{pd}, 1)$, $f(p)$ is given by (3). This is always negative, so the equilibrium at $p = 1$ is unstable.

Thus, for Case 1, there are the following possibilities. If the all-team reasoner equilibrium is unstable, then p will 'oscillate' around p_{pd}, increasing until it exceeds p_{pd}, and then decreasing until it becomes smaller than p_{pd}; thus, the outcome is, in practice, similar to an internal equilibrium at p_{pd} (see Figure 3.) If the all-team reasoner equilibrium is stable, but there is an unstable equilibrium at $p^* \in (0, p_{pd})$, then for any initial $p < p^*$, the equilibrium outcome is $p = 0$, and for any initial $p > p^*$, p will ultimately oscillate

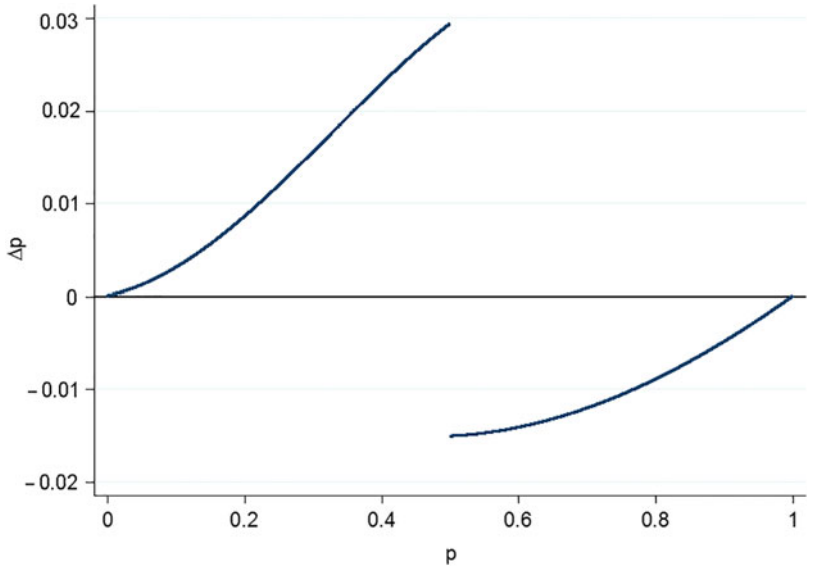

Figure 3. When team reasoners are circumspect, both the $p = 0$ and $p = 1$ equilibria may be unstable, and p may oscillate around p_{pd}. Here, $w_0 = 5$, $a = 1$, $b = 2$, $c = 3$, $\beta = 2$, $\gamma = 2$, $h = .8$, and $x = .5$. See Case 1 in text for details.

around p_{pd}. Finally, if the all-team reasoner equilibrium is stable and there is no internal equilibrium $\in (0, p_{pd})$, then the equilibrium outcome is $p = 0$, for any initial p.

Case 2: $0 < p_{hl} < p_{pd} < 1$ (implying $h < .5$ and $c > b$).

On $(0, p_{hl})$, $f(p)$ is given by (2). Here too, $f(p)$ is increasing in p. An analysis parallel to that of Case 1 shows that sets of parameter values exist such that (1) there is a stable equilibrium at $p = 0$, and an unstable equilibrium $\in (0, p_{hl})$; (2) there is an unstable equilibrium at $p = 0$ and no internal equilibrium $\in (0, p_{hl})$. Finally, solving $f(p_{hl}) < 0$ for x gives

$$x > \frac{\beta c + 2(\beta h + 2\gamma h - \gamma)(a - b)}{\beta c + 2(\beta h + 2\gamma h - \gamma)(a - b) - (\beta h + 2\gamma h - \gamma)(1 - h)(\beta + 2\gamma)}.$$

Thus, for such x, there is a stable equilibrium at $p = 0$, and no internal equilibrium $\in (0, p_{hl})$.[11]

On $f(p_{hl}, p_{pd})$, $f(p)$ is given by (4). This is again increasing in p. An analysis parallel to that of Case 1 shows that sets of parameter values exist such that (1) $f(p_{hl}) < 0$ but there is an unstable internal equilibrium $\in (p_{hl}, p_{pd})$; (2) $f(p) > 0$ $\forall p \in (p_{hl}, p_{pd})$.[12] And since $f(p_{pd}) < 0$, for $x > b(c - b + a)/[b(c - b + a) + h[\gamma(c - b) - a(\beta h - \gamma + 2\gamma h)]]$, there exist a set of parameter values for which $f(p) < 0$, $\forall\, p \in (p_{hl}, p_{pd})$.

On $f(p_{pd}, 1)$, $f(p)$ is given by (5). Whenever $p_{hl} \in (0, 1)$, this is certainly negative, so the equilibrium at $p = 1$ is unstable.

Lastly, observe that $f(p)$ is increasing in p on $[0, p_{pd}]$. (To see this, note that the right side of (5) is greater than the right side of (2), for $p = p_{hl}$, iff $\beta h < \gamma - 2\gamma h$, which always holds for $p_{hl} \in (0, 1)$.)

Therefore, the substantive interpretation of the model is the same as for Case 1: if the all-team reasoner equilibrium is unstable, then p will oscillate around p_{pd}; if the all-team

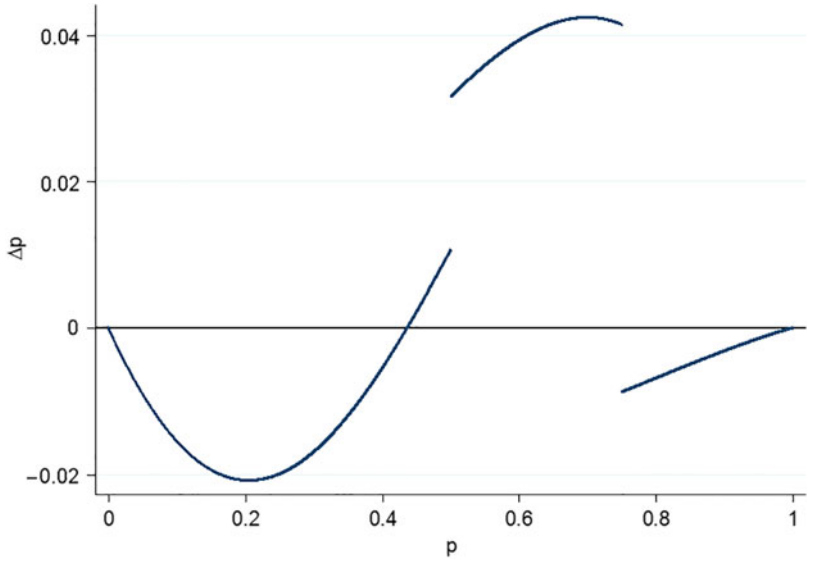

Figure 4. Another possibility when team reasoners are circumspect. Here, the equilibrium at $p = 0$ is stable, the equilibrium at $p = 1$ is unstable, there is an unstable internal equilibrium, and oscillation around p_{pd}. Here, $w_0 = 5$, $a = 1.5$, $b = 2$, $c = 2.5$, $\beta = 2$, $\gamma = 4$, $h = .3$, and $x = .4$. See Case 2 in text for details.

reasoner equilibrium is stable, but there is an unstable equilibrium $p^* \in (0, p_{pd})$, then for any initial $p < p^*$, the equilibrium outcome is $p = 0$, and for any initial $p > p^*$, p will ultimately oscillate around p_{pd} (see Figure 4); finally, if the all-team reasoner equilibrium is stable and there is no internal equilibrium $\in (0, p_{pd})$, then the equilibrium outcome is $p = 0$.

Case 3: $0 < p_{pd} < p_{hl} < 1$ (again implying $h < .5$ and $c > b$).

On $[0, p_{pd}]$, $f(p)$ is given by (2). Here too, $f(p)$ is increasing in p. An analysis parallel to that of Case 1 shows that sets of parameter values exist such that (1) there is a stable equilibrium at $p = 0$, and an unstable equilibrium $\in (0, p_{pd})$; (2) there is an unstable equilibrium at $p = 0$ and no internal equilibrium $\in (0, p_{pd})$. Finally, the analysis of Case 2 on $(0, p_{hl})$ shows that there is a set of parameter values for which $f(p_{pd}) < 0$, since p_{pd} must be less than p_{hl}.

On (p_{pd}, p_{hl}), $f(p)$ is given by (3), and, again, increasing in p. To see that $f(p_{hl}) < 0$ (and thus $f(p) < 0$, $\forall p \in (p_{pd}, p_{hl})$), write $f(p)$ as $x(1-h)/[(\beta + 2\gamma)(1-2h)]$ $(\gamma(1-2h) - \beta h)(\beta - 2(\beta + \gamma))$; observe that $\gamma(1-2h) - \beta h$ is positive whenever $p_{hl} \in (0, 1)$, so $f(p_{hl}) < 0$.

On $(p_{hl}, 1)$, $f(p)$ is given by (5), and so whenever $p_{hl} \in (0, 1)$, $f(p) < 0$, $\forall p \in (p_{hl}, 1)$.

To summarize Case 3, the all-individual reasoner equilibrium is never stable. If the all-team reasoner equilibrium is unstable, then p will oscillate around p_{pd}; thus, the outcome is, again, in practice, similar to an internal equilibrium at p_{pd}. If the all-team reasoner equilibrium is stable, but there is an unstable equilibrium at $p^* \in (0, p_{pd})$, then for any initial $p < p^*$, the equilibrium outcome is $p = 0$, and for any initial $p > p^*$, p will ultimately oscillate around p_{pd}. Finally, if the all-team reasoner equilibrium is stable and there is no internal equilibrium $\in (0, p_{pd})$, then the equilibrium outcome is $p = 0$, for any initial p.

Case 4: $p_{hl} \in (0, 1)$, but $p_{pd} \notin (0, 1)$ (implying $c < b$ and $h < .5$). On $(0, p_{hl})$, $f(p)$ is given by (2). Although $f(p)$ is decreasing in p for $b > a + c + [x(1 - h)^2(\beta + 2\gamma)]/(1 - x)$, $f(1)$ (and so certainly for such b, $f(p_{hl})$) is never less than 0 – thus, there can be no stable internal equilibrium $\in (0, p_{hl})$. (In other words, no internal equilibrium exists for b such that $f(p)$ is decreasing.[13]) If b is such that $f(p)$ is increasing in p, then, an analysis parallel to that of Case 2 on $(0, p_{hl})$ shows that sets of parameter values exist such that (1) there is a stable equilibrium at $p = 0$, and an unstable equilibrium $\in (0, p_{hl})$; (2) there is an unstable equilibrium at $p = 0$ and no internal equilibrium $\in (0, p_{hl})$; (3) there is a stable equilibrium at $p = 0$, and no internal equilibrium $\in (0, p_{hl})$.

On $(p_{hl}, 1)$, $f(p)$ is given by (4). Note that, on $(p_{hl}, 1)$, $f(p)$ is continuous and decreasing in x; also, it is decreasing in p iff $b > a + c + [xh^2(\beta + 2\gamma)]/(1 - x)$. Let $(0 <) b_* < a + c + [xh^2(\beta + 2\gamma)]/(1 - x) < b^* (< 2a + c)$, and let $c/[c - h(\beta h + 2\gamma h - \gamma)] \equiv x^*$ be the value of x for which $f(1) = 0$. Clearly, for any $p' \in (p_{hl}, 1)$, $f_{b^*,x^*}(p') > 0$ and $f_{b_*,x^*}(p') < 0$. Now, set $x = x^* + \varepsilon$ (holding a,b^*, b_*, β,c, γ, h, p' fixed). Because $f_{b,x}(p)$ is continuous in x, $f_{b^*,x^*+\varepsilon}(p') > 0$ and $f_{b_*,x^*+\varepsilon}(p') < 0$ for sufficiently small ε. At the same time, since $f_{b,x}(p)$ is decreasing in x, $f_{b^*,x^*+\varepsilon}(1) < 0$ and $f_{b_*,x^*+\varepsilon}(1) < 0$.

Thus, there exists a set of parameter values for which there is an unstable equilibrium at $p = 1$ and a stable internal equilibrium at $p^* > p_{hl}$, and a set of values such that there is an unstable equilibrium at $p = 1$ and no internal equilibrium $\in (p_{hl}, 1)$. The parallel analysis for $x^* - \varepsilon$ establishes that there is a set of parameter values for which there is a stable equilibrium at $p = 1$ and an unstable internal equilibrium at $p^* > p_{hl}$, and a set of values such that there is a stable equilibrium at $p = 1$ and no internal equilibrium $\in (p_{hl}, 1)$. Lastly – since, for every set parameter values (given that $p_{hl} \in (0, 1)$), $f(p_{hl} + \varepsilon) > f(p_{hl} - \varepsilon)$ – there will be no oscillation around p_{hl}.

To summarize Case 4, the all-team reasoner equilibrium can be stable or unstable; an internal equilibrium $p^* < p_{hl}$, if it exists, will be unstable. The all-individual reasoner equilibrium can also be stable or unstable, and an internal equilibrium $p^* > p_{hl}$, if it exists, can be either stable or unstable. So, if the all-team reasoner equilibrium is unstable, then for any initial p, the population will end up at the stable internal equilibrium $p^* > p_{hl}$, if it exists, and at the (stable) all-individual reasoner equilibrium, if it does not. If the all-team reasoner equilibrium is stable, and the all-individual reasoner equilibrium is unstable, then – if no stable internal equilibrium exists – the system ends up at the all-team reasoner equilibrium for any initial p; but – if a stable equilibrium does exist – then the endpoint of the all-team reasoner equilibrium's basin of attraction will be at the unstable equilibrium $p^* < p_{hl}$, if such p^* exists, and at p_{hl}, if it does not. If both the all-team reasoner and the all-individual reasoner equilibria are stable, their basins of attraction are defined by the unstable internal equilibrium $p^* < p_{hl}$, if it exists, and by p_{hl} otherwise.

6. Conclusion

We have shown that team reasoners, and thus cooperative behavior, can thrive even in an environment that appears hostile to such behavior; even in one-shot interactions with random pairings of players, cooperation can be sustained. The key is variation in the ludic ecology. True, the mechanism of individual reasoning is successful relative to team reasoning in social dilemmas like the PD. However, it is relatively unsuccessful in the Hi Lo (and we have suggested that this is particularly likely to be true outside the laboratory.) We have shown that these facts can be important for understanding evolutionary

outcomes. In an environment where common interest games are prevalent, team reasoning is the only evolutionarily stable strategy; at the same time, if the ludic ecology consists mainly of social dilemmas, individual reasoning is favored.

We have also pursued several extensions that augment this basic account. We have noted that over-time changes in the proportion of games that are Hi Lo may allow both types of reasoners to persist in the system longer than would otherwise be expected. Second, we have argued that individual reasoners may be more susceptible to certain errors of perception; this may be a further evolutionary advantage for team reasoners. Third, we have discussed how a more complex, and thus costly, mechanism fares against team reasoners. Fourth, we have noted that a stable internal equilibrium, with both team and individual reasoners, may exist, when individual reasoners' play in the Hi Lo responds to the proportion of team reasoners in the population. Last, in considering team reasoners who are circumspect, we have shown that there are scenarios in which individual reasoners and circumspect team reasoners coexist indefinitely, and that – whenever the parameters of the PD are such that circumspection potentially affects behavior – circumspect team reasoners can always invade a population of individual reasoners.

Though we have shown that team reasoning is a evolutionarily viable strategy against individual reasoning, we do not argue that other explanations of cooperative behavior should be discounted. Cooperation, in many contexts, is compatible with individual reasoning. However – as we have discussed – team reasoning explains cooperative behavior in many (seemingly) simple interactions, and it does so in a coherent and parsimonious fashion. And it can explain cooperation even in the absence of reciprocity (conditional cooperation) or assortative interaction (nonrandom pairing). Surely, further empirical investigation regarding the prevalence of, and mechanisms underlying, team reasoning is warranted; in particular, scholars will have to design and implement critical tests in which the behavioral implications of team reasoning differ from those of competitor theories (Faillo, Smerilli, & Sugden, 2013). But, until such evidence accumulates, we argue that team reasoning should not be dismissed as implausible on evolutionary-theoretic grounds; it is a viable approach to basic human interactions.

Acknowledgements

For helpful comments and suggestions, we thank Jim Alt, Marilyn Brewer, Ákos Lada, Rohit Parikh, Ken Shepsle, Robert Sugden, and the participants in the Harvard Department of Government Research Workshop in Political Economy.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes

1. Email: lemperds@potsdam.edu.
2. We point out a few representative examples, without implying that these views are unanimously held in any discipline or subfield. In biology, see (Hamilton, 1964), (Trivers, 1971), and (Wilson, 1975); in economics, see (Binmore, 2004; Gintis, Bowles, Boyd, & Fehr, 2003); in political science, see (Axelrod, 1984; Bendor & Swistak, 1997); in philosophy, see (Pettit, 2000; Skyrms, 2003); in psychology, see the discussion in (Caporael, Dawes, Orbell, Van de kragt, & van de Kragt, 1989) and the responses therein, especially by Houston and Hamilton, Krebs, Liebrand, Rachlin, Tooby and Cosmides, and Vine.
3. Substituting the Stag Hunt for the Hi Lo does not qualitatively change the results we present in our baseline model.

4. We note that we slightly diverge from Bacharac (2006) here; he calls the mechanism leading to the choice of Hi in Hi Lo and Cooperate in PD 'group identification.' Group identification is said to activate team reasoning, which, in turn, leads a player to exhibit the traits Hi and Cooperate. Other theorists are more agnostic about the role of group identification in bringing about team reasoning (Gold & Sugden, 2007). In any case, our model is not affected if we replace 'team reasoning' with 'identifying with one's co-player as a group member' and 'individual reasoning' with 'not identifying with one's co-player as a group member.'
5. We do not mean to imply that we are ourselves skeptical of group or multilevel selection; we simply note the concept is contested (on this point, see Wilson & Sober, 1994).
6. See (Bendor & Swistak, 1997, pp. 295–296) for a behavioral interpretation of the replicator dynamic.
7. The model could be complicated by considering other kinds of errors, but we do not pursue that extension here.
8. Bacharach (2006, p. 108) discusses closely related considerations.
9. As we show below, circumspect team reasoning is interesting in the PD only if $b - c < 0$.
10. $x^* = b - a/[(1 - h)(\gamma + \beta) + b - a]$.
11. Note this $x \in (0, 1)$, since $p_{hl} \in (0, 1)$ implies $\beta h + 2\gamma h - a < 0$.
12. Here, $f(p_{hl}, x^*) = 0$ for $x^* = [\beta c + (\beta h + 2\gamma h - \gamma)(a - b)]/[\beta c + (\beta h + 2\gamma h - \gamma)(a - b) - (\beta h - \gamma + 2\gamma h)(\beta h + \gamma h)]$; note that this $x^* \in (0, 1)$.
13. It is worth observing that this is conditional on $p_{hl} \in (0, 1)$. Otherwise, a stable internal equilibrium can exist for such b. The reason that this is not a possibility in our baseline model is because in the additive PD, $a = b - c$, meaning $f(p)$ is increasing in p. Thus, the possibility of a stable internal equilibrium is not tied to team reasoners being circumspect; however, it is conditional on the parameters of the PD being such that circumspect team reasoners' behavior is not always the same as that of noncircumspect team reasoners.

References

Axelrod, R. (1984). *The evolution of cooperation*. New York: Basic Books.

Bacharach, M. (1999). Interactive team reasoning: A contribution to the theory of cooperation. *Research in Economics*, *53*, 117–147. doi:10.1006/reec.1999.0188

Bacharach, M. (2006). *Beyond individual choice: Teams and frames in game theory*. Princeton, NJ: Princeton University Press.

Bendor, J., & Swistak, P. (1997). The evolutionary stability of cooperation. *American Political Science Review*, *91*, 290–307. doi:10.2307/2952357

Binmore, K. (2004). Reciprocity and the social contract. *Politics, Philosophy, and Economics*, *3*, 5–35. doi:10.1177/1470594X04039981

Boyd, R., & McElreath, R. (2007). *Modeling the evolution of social behavior: A guide for the perplexed*. Chicago, IL: University of Chicago Press.

Camerer, C. F., Loewenstein, G., & Rabin, M. (Eds.). (2003). *Advances in behavioral economics*. Princeton, NJ: Princeton University Press.

Caporael, L. R. (2007). Evolutionary theory for social and cultural psychology. In W. Arie, E. Kruglanski, & T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (2nd ed., pp. 3–18). New York: Guildford Press.

Caporael, L. R., Dawes, R. M., Orbell, J. M., & Van de kragt, A. J. C. (1989). Selfishness examined: Cooperation in the absence of egoistic incentives. *Behavioral Brain Sciences*, *12*, 683–739. doi:10.1017/S0140525X00025292

Colman, A. M., Pulford, B. M., & Rose, J. (2008). Collective rationality in interactive decisions: Evidence for team reasoning. *Acta Psychologica*, *128*, 387–397. doi:10.1016/j.actpsy.2007.08.003

Faillo, M., Smerilli, A., & Sugden, R. (2013). *The roles of level-k and team reasoning in solving coordination games*. University of Trento Cognitive and Experimental Economics Laboratory Working Paper 6-13. Trento: University of Trento.

Gilbert, M. (1989). *On social facts*. London: Routledge.

Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior*, *24*, 153–172. doi:10.1016/S1090-5138(02)00157-5

Gold, N., & Sugden, R. (2007). Theories of team agency. In F. Peter & H. Bernhard Schmid (Eds.), *Rationality and commitment* (pp. 280–312). Oxford, UK: Oxford University Press.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. II. *Journal of Theoretical Biology*, *7*(1), 17–52. doi:10.1016/0022-5193(64)90038-4

Heinrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., et al. (2005). Economic man in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*, *28*, 795–815.

Hollis, M. (1998). *Trust within reason*. Cambridge: Cambridge University Press.

Hurley, S. (1989). *Natural reasons*. New York: Oxford University Press.

Kramer, R. M., & Brewer, M. (1984). Effects of group identity on resource use in a simulated commons dilemma. *Journal of Personality and Social Psychology*, *46*, 1044–1057. doi:10.1037/0022-3514.46.5.1044

Pettit, P. (2000). Rational choice, functional selection and empty black boxes. *Journal of Economic Methodology*, *7*, 33–57. doi:10.1080/135017800362239

Regan, D. (1980). *Utilitarianism and co-operation*. Oxford: Oxford University Press.

Skyrms, B. (2003). *The stag hunt and the evolution of social structure*. Cambridge, UK: Cambridge University Press.

Sober, E., & Wilson, D. S. (1998). *Unto others* (2nd ed.). Cambridge, MA: Harvard University Press.

Sugden, R. (1993). Thinking as a team: Towards an explanation of nonselfish behavior. *Social Philosophy and Policy*, *10*, 69–89. doi:10.1017/S0265052500004027

Sugden, R. (2000). Team preferences. *Economics and Philosophy*, *16*, 175–204. doi:10.1017/S0266267100000213

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

Wilson, D. S., & Sober, E. (1994). Reintroducing group selection to the human behavioral sciences. *Behavioral and Brain Sciences*, *17*, 585–608. doi:10.1017/S0140525X00036104

Wilson, E. O. (1975). *Sociobiology*. Cambridge, MA: Harvard University Press.