Journal of Theoretical Biology 253 (2008) 261–270

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Human uniqueness-self-interest and social cooperation

Daijiro Okada [a,b,*,1], Paul M. Bingham [c]

[a] Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540, USA
[b] Department of Economics Rutgers, The State University of New Jersey, New Brunswick, NJ 08901-1248, USA
[c] Department of Biochemistry and Cell Biology and College of Human Development, Stony Brook University, Stony Brook, NY 11794-5215, USA

ARTICLE INFO

Article history:
Received 18 July 2007
Received in revised form
23 December 2007
Accepted 22 February 2008
Available online 13 March 2008

Keywords:
Kinship-independent cooperation
Scale of cooperation
Cost of coercion
Synchronous attackers
Elite throwing
Human fossil record

ABSTRACT

Humans are unique among all species of terrestrial history in both ecological dominance and individual properties. Many, or perhaps all, of the unique elements of this nonpareil status can be plausibly interpreted as evolutionary and strategic elements and consequences of the unprecedented intensity and scale of our social cooperation. Convincing explanation of this unique human social adaptation remains a central, unmet challenge to the scientific enterprise.

We develop a hypothesis for the ancestral origin of expanded cooperative social behavior. Specifically, we present a game theoretic analysis demonstrating that a specific pattern of expanded social cooperation between conspecific individuals with conflicts of interest (including non-kin) can be strategically viable, *but only* in animals that possess a highly unusual capacity for conspecific violence (credible threat) having very specific properties that dramatically reduce the costs of coercive violence. The resulting reduced costs allow preemptive or compensated coercion to be an instantaneously self-interested behavior under diverse circumstances rather than in rare, idiosyncratic circumstances as in actors (animals) who do not have access to inexpensive coercive threat.

Humans are apparently unique among terrestrial organisms in having evolved conspecific coercive capabilities that fulfill these stringent requirements. Thus, our results support the proposal that access to a novel capacity for projection of coercive threat might represent the essential initiating event for the evolution of a human-like pattern of social cooperation and the subsequent evolution of the diverse features of human uniqueness. Empirical evidence indicates that these constraints were, in fact, met only in our evolutionary lineage. The logic for the emergence of uniquely human cooperation suggested by our analysis apparently accounts simply for the human fossil record.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

On genetic grounds, humans are plainly one of four African great ape species. Yet, we exercise a level of ecological dominance never before seen in the ca. 3.75 billion year history of life on this planet. Moreover, we have a series of individual properties (including complex symbolic speech and cognitive virtuosity) likewise apparently unprecedented. A credible scientific understanding of the origins and implications of these unique human attributes requires active collaboration between the natural sciences and the social sciences.

We suggest, with others, that all the unique properties of humans might be interpretable as elements and effects of our novel pattern of social cooperation. A logically coherent and empirically verisimilar explanation of this cooperation, and the evolved human behaviors that sustain it, has been and remains an active subject of research (Alexander, 1987; Axelrod, 1984; Darwin, 1871; Maynard Smith and Szathmâary, 1995; Sober and Wilson, 1998; Williams, 1966; Wilson, 1975b). Existing theoretical and empirical work has provided important local insights (Gavrilets and Vose, 2006; Gurerk et al., 2006; Nowak, 2006; Rockenbach and Milinski, 2006; Sethi and Somanathan, 2003). However, most currently popular approaches either implicitly presuppose human uniqueness or explain it in ways that are empirically doubtful or untestable (Section 3).

Our goal here is to examine the logic of social cooperation from a new perspective. This perspective, in turn, suggests a potentially fruitful, verisimilar proposal for the evolution of uniquely human social cooperation.

We analyze a novel strategic basis for social cooperation between individuals with conflicts of interest. To understand the possible evolutionary implications of our analysis it is useful to recall how the economists' concept of "conflicts of interest" maps onto biological social behavior. Kin-selection theory (Hamilton,

* Corresponding author at: Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901-1248, USA. Tel.: +1 732 932 7797; fax: +1 732 932 7416.
E-mail addresses: okada@econ.rutgers.edu (D. Okada), pbingham@notes.cc.sunysb.edu (P.M. Bingham).

[1] Daijiro Okada gratefully acknowledges the financial support from the Richard B. Fisher Membership at Institute for Advanced Study.

0022-5193/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2008.02.041

1964a, b; Williams and Williams, 1957) has had many successes in accounting for non-human animal social behavior and diverse human behaviors we often think of as constituting our "private" lives.

Specifically, social behaviors by one animal that enhance the replication, through reproduction, of design information shared by closely related animals—especially offspring, siblings and parents—evolve by natural selection. Such kin-selected behaviors serve the "interests" of design information (Dawkins, 1976; Hamilton, 1996). A major remaining challenge is to understand the logic of social cooperation among conspecific animals who are *not* close kin, i.e., *kinship-independent* social cooperation. In this "public" cooperation between non-kin conspecifics, natural selection is expected to produce individuals each of whom behaves as if he/she is the unit of interest—though sometimes also acting on behalf of close kin. Equivalently, the individual organism can be treated as the unit of interest in the public arena. It is this public cooperation between non-kin that ultimately concerns us here.

We interpret available evidence concerning social behavior in non-human animals (Dugatkin, 1997; Koenig, 1988; Krebs and Davies, 1993; Taylor and McGuire, 1988; Wilkinson, 1988) to provide insights into the dilemma presented by the rare, special occurrence of uniquely massive kinship-independent social cooperation in humans. Specifically, we argue that this body of evidence supports three fundamental claims.

First, non-kin members of the same species (conspecifics) almost always live in a crowded "Malthusian" world requiring competition for resources necessary for individual evolutionary success. As a result, natural selection produces individual animals who behave as if they understand their *interests* in obtaining resources even at the expense of other non-kin conspecifics. Thus, kinship-independent social cooperation is strictly limited by short-term, and recurring, *conflicts of interest* between conspecific individuals.

Second, kinship-independent cooperation *can* occur between non-human individuals with conflicts of interest when circumstances permit cost-effective suppression of conflicts of interest through *active coercion*. This issue has received relatively little attention from investigators focusing on human social behavior; however, a large, diverse body of elegant work in social insects strongly supports this claim (Ratnieks et al., 2006).

Third, this active coercion must represent *immediately self-interested* behavior on the part of coercing *individuals*. Again, this claim is strongly supported by the work in social insects (Ratnieks et al., 2006). However, this inference will be controversial to advocates of some approaches to understanding human social cooperation, especially advocates of "strong reciprocity" and related group selection models (Sections 3.1, 3.3). The non-cooperative game we analyze here provides a basic strategic logic for our hypothesis that immediate individual self-interest might be the dominant consideration in non-kin conspecific social cooperation.

In our game cooperation between individuals with conflicts of interest is sustained by the purely self-interested application of coercive threat. Our analysis indicates that this strategy is viable only for individuals who can project coercive threat remotely, that is, from a distance of many body diameters. Humans are apparently the first and only animal with an innate biological capacity for elite remote attack against adult conspecifics. Thus, our analysis suggests the potentially useful proposal that humans evolved their uniquely massive kinship-independent social cooperation initially[2] as a result of cooperative logic similar to that analyzed in our game (Section 3.2).

[2] Whether other factors, such as cultural transmission of norms etc., play fundamental roles in subsequent further expansion of human cooperation needs to be closely examined.

2. A game of cooperation and preemptive or compensated coercion

In this section we develop and analyze a two-stage game that captures important features of one form of social cooperation. A useful mental image for the game might be a cooperative hunting/power-scavenging episode in which the cost of hunting/scavenging generates a benefit in the form of a carcass to which players may have access or from which they may be coercively excluded (ostracized). All proofs can be found in Supporting Online Material.

2.1. Description of the game

There are n players in total. First, players simultaneously choose C (cooperate) or NC (not cooperate). How the game proceeds hereafter depends on how many players chose C. Let k be the number of players who chose C (henceforth C players), so $n-k$ is the number of players who chose NC (NC players).

(a) If $k=0$ (no player chose to cooperate) then the game ends and each player receives a baseline payoff normalized to be 0.
(b) If $k=n$ (all players chose to cooperate) then the game ends and each player receives $b-c$ where b is the per capita benefit of cooperation and c is the per capita cost of cooperation. We assume that $b>c$.
(c) If $1\leqslant k\leqslant n-1$, then the game continues to the second stage, denoted by $G(k)$. In this subgame each of the C players chooses F (fight) or NF (not fight).[3] Let ℓ represent the number of C players who chose F (C/F players) so that $k-\ell$ is the number of C players who chose NF (C/NF players). Payoffs are then specified as follows.
(c-1) If $\ell=0$, the total benefit kb is shared equally among all players. Thus each C player receives $(k/n)b-c$ and each NC player receives $(k/n)b$.
(c-2) If $1\leqslant\ell\leqslant k$, then combat takes place between the C/F players and the NC players. We assume that one of two events takes place as a result of the combat: either the C/F players successfully chase away the NC players or this attempt is unsuccessful. For simplicity we will assume that the former event occurs if the C/F players are at least as numerous as the NC players and the latter event occurs otherwise.[4] Fighting costs $f(\ell, n-k)$ to *each* C/F player and $a(\ell, n-k)$ to each NC player. We assume that f is decreasing in ℓ and increasing in $n-k$. Similarly, a is assumed to be increasing in ℓ and decreasing in $n-k$. Regardless of the outcome of the fight, C/NF players are assumed to lose a portion of the benefit.[5]

Specifically, if the C/F players are unsuccessful in fighting off the advancing NC players (this happens when $\ell<n-k$), then the total

[3] The NC players are assumed to advance toward the C cooperators in an attempt to seize a portion of the benefit from cooperation. One could allow each NC player to choose between F and NF. However, this introduces additional complexity into the analysis without a qualitative difference in conclusion.

[4] In practice, animals have different fighting abilities and the group with the higher total fighting ability would dominate the opposing group. Our assumption here is simply a special case of this where the animals possess equal fighting ability, except the tie-breaking rule employed. Alternative tie-breaking rules, however, do not change the qualitative outcome of our analysis.

[5] The realistic behavior corresponding to this constraint might be as follows. Non-fighting cooperators (C/NF) are expected to leave the scene with their initial shares to avoid becoming ensnared in costly combat (also see Section 2.5 below). Any new consumable shares generated by subsequent ostracism of non-cooperating individuals would then be distributed only to the fighting cooperators.

benefit kb is divided equally among all players. Thus the payoffs to the three types of players are as follows:

$$\begin{aligned} \text{C/F player} &: \frac{k}{n}b - c - f(\ell, n-k) \\ \text{C/NF player} &: \frac{k}{n}b - c \\ \text{NC player} &: \frac{k}{n}b - a(\ell, n-k) \end{aligned} \tag{1}$$

If the C/F players successfully chase away the NC players (this happens when $\ell \geqslant n-k$), then we assume that the C/F players take the share that otherwise could have gone to the NC players.[6] Note that the share of the benefit that could have gone to the NC players is $(n-k)(k/n)b$. This is divided equally among the C/F players so that each C/F player receives an augmented benefit $(k/n)b+((n-k)/\ell)(k/n)b$ and pays the cost of cooperation plus the cost of fighting. Each NC player receives no benefit but pays the cost of engaging in combat. Hence the payoffs in this case are as follows:

$$\begin{aligned} \text{C/F player} &: \frac{k}{n}b + \left(\frac{n-k}{\ell}\right)\frac{k}{n}b - c - f(\ell, n-k) \\ \text{C/NF player} &: \frac{k}{n}b - c \\ \text{NC player} &: -a(\ell, n-k) \end{aligned} \tag{2}$$

Observe that a C/NF player's payoff is same regardless of the outcome of the combat. Becoming a non-fighter is a decision to give up a portion of the benefit, $(1-(k/n))b$, rather than incurring the cost of fighting.

Note that fighting here has the effect of *preempting* access to the benefits of cooperation, thereby generating an immediate benefit (compensation) that can be distributed among coercing players. However, this potential benefit is in no way guaranteed and comes only at the cost (of fighting) and the risk (of losing a fight). This is in contrast to a number of earlier models in which coercion is conceived as *post facto* "punishment".[7]

Lastly, we make an additional assumption that the number of participating players n is at least modestly large so that $b < nc$.[8]

[6] Roughly equal shares are the expected outcome of simple scramble competition for the shared resource. One way to visualize the assumptions in our model is as follows. After a kill, all parties (C/F, C/NF and NC) tear a roughly equal share of the meat from the carcass. The NC individuals come under attack by the C/F individuals and drop their shares and withdraw in order to end the attack when the C/F individuals are in the majority. The remaining C/F individuals "distribute" the NC leavings by further scramble competition. Thus, the C/NF individuals obtain a "per capita" share on basis of all players, while C/F players obtain the extra return from redistributing the shares of the NC players. See Section 2.5 for how this logic works in the "remote attacking" animals that will concern us here. In such animals, C individual cannot "stay around" to take an extra share without becoming a target of coercive violence from NC individuals. Under these conditions, the C player remaining on site without projecting threat toward the NC target enhances her/his own cost be extending the duration of fire from NC individuals. Thus, the only rational strategies are to leave the scene (C/NF as stipulated) or remain on the scene, project threat and participate in scramble competition for the recaptured NF resources (the C/F strategy as stipulated).

[7] Analysis of non-human animal behavior indicates that animals quickly evolve to assess what the risks and costs of fighting (ultimate costs) would be and behave on the basis of those anticipated costs (Maynard Smith, 1982). These likely costs are sometimes assessed by interactions that entail some, generally modest, costs in themselves and that allow the ultimate costs of full conflict to be reliably assessed. However, the animal's evolved behavior is apparently predicated on the assessed likely ultimate costs and our analysis focuses on these costs.

[8] This assumption is very likely to be valid for the real biological cases that ultimately concern us. Evolution of cooperation between close kin individuals is partially driven by *confluent* (genetic) interests. For most large mammals, the number of adults involved in such cooperation would generally be of the order of 3–10. In contrast, kinship-independent cooperation can be extended to indefinitely large number of individuals. Thus, n will generally be of the order of 10 individuals or more. Under these conditions the costs of cooperation would have to be unusually small (less than 10% of gross benefit) for this assumption to be violated. We suggest that there are relatively few such extravagantly remunerative

2.2. Analysis of the second-stage games

In this section we examine strategic (Nash) equilibria of the second-stage game. Results of this section will be used in the next section to derive conditions under which an outcome is a robust equilibrium outcome in the first-stage game. Recall that k represents the number of C players. A typical outcome of the subgame $G(k)$ is denoted by $(\ell, k-\ell)$ where ℓ is the number of C/F players.

We claim that, if less than half of the players chose to cooperate in the first stage, the only equilibrium outcome of the second-stage game is $(0, k)$, all C players stand back and let NC players take shares of cooperative benefit.

Proposition 1. *For $1 \leqslant k < n/2$, the only equilibrium outcome of $G(k)$ is $(0, k)$.*

In contrast, suppose that C players are in majority in the first stage. In this case, additional equilibrium outcomes where some cooperators choose to fight emerge under certain conditions.

Proposition 2.

(a) *For $n/2 \leqslant k < n-1$, $(0, k)$ is always an equilibrium outcome of $G(k)$.*

(a)′ *$(0, n-1)$ is an equilibrium outcome of $G(n-1)$ if $((n-1)/n)b \leqslant f(1,1)$.*[9]

(b) *For $n/2 \leqslant k \leqslant n-1$, $(\ell, k-\ell)$ is an equilibrium outcome of $G(k)$ if $n-k \leqslant \ell < k$ and*

$$\left(\frac{\ell}{n-k}\right) f(\ell, n-k) \leqslant \frac{k}{n}b \leqslant \left(\frac{\ell+1}{n-k}\right) f(\ell+1, n-k). \tag{3}$$

(c-2) *For $n/2 \leqslant k \leqslant n-1$, $(k, 0)$ is an equilibrium outcome of $G(k)$ if*

$$\frac{k}{n-k} f(k, n-k) \leqslant \frac{k}{n}b. \tag{4}$$

The following fact, which is a straightforward consequence of Proposition 2, will be useful in the next section.

Corollary 1. *Suppose that $n/2 \leqslant k \leqslant n-1$. If $(\ell, k-\ell)$, where $n-k \leqslant \ell < k$, is an equilibrium outcome of $G(k)$, then a C/F player's payoff is at least a C/NF player's payoff.*

In sum, when cooperators are in the minority, the only viable option for them in the stage two subgame is to stand back and allow non-cooperators to partake of the benefit of cooperation (Proposition 1). When cooperators are in the majority, however, a subset of fighters, or enforcers, can emerge who hold the non-cooperators in check provided that the fighters outnumber the non-cooperators[10] and, in addition, the cost of fighting is relatively small (Propositions 2 and 3).

(*footnote continued*)
cooperative behaviors. Note that consistent availability to highly cost-effective cooperative behaviors would be expected to drive the evolution of large kin-selected units to exploit them. Few, if any, such units are seen in large mammals.

[9] Note that $(0, n-1)$ is not unconditionally an equilibrium outcome of $G(n-1)$. This is because, unlike the case where $k < n-1$, there is only one NC player in this case and so, if a C/NF player switched to C/F, he will win the fight and take an additional share of the benefit but at a cost of fighting, $f(1,1)$.

[10] This conclusion depends on our assumption of equal fighting abilities among all the players. However, the fundamental qualitative outcomes are not affected if there are differences in fighting abilities. Also see Section 2.5.

2.3. Robustness of the first-stage outcomes

We now derive conditions for various outcomes of the first-stage game to be *robust outcomes*. The idea is that, in deciding a role to play in the first stage (C or NC), each player anticipates some equilibrium outcome to be played in the second stage. A first-stage outcome is said to be robust if no player can gain by unilaterally switching his role given anticipated equilibrium outcomes in the second-stage games.[11] Thus, whether a first-stage outcome is robust or not depends on what equilibrium outcomes are anticipated in the second-stage games induced by the behavior of a player whose incentive is under consideration.

To make the definition of robust outcome more precise, denote a typical first-stage outcome by $[k, n-k]$ where k is the number of C players. Consider a C player. If he remains a C player, then the subgame $G(k)$ will be played. If he switches his role, then the induced subgame is $G(k-1)$. So a C player has no incentive to become a NC player so long as his payoff at an anticipated equilibrium outcome in $G(k)$ is at least a NC player's payoff in an anticipated equilibrium outcome in $G(k-1)$. A NC player makes a similar comparison of payoffs in $G(k)$ and $G(k+1)$. The outcome $[k, n-k]$ is robust if neither a C player or a NC player has an incentive to switch his role given other players' roles in $[k, n-k]$.

As is the case with virtually all models of cooperation based on prisoner's dilemma or its n-player extension ("tragedy-of-commons" models), the outcome in which no player cooperates is robust.

Proposition 3. *The first-stage outcome* $[0, n]$ *is robust.*

Next we examine the robustness of the first-stage outcomes in which some cooperators are present: $[k, n-k]$ where $1 \leqslant k \leqslant n$. The following lemma helps us narrow down the candidates for robust outcomes.

Lemma 1. *A first-stage outcome* $[k, n-k]$, $1 \leqslant k \leqslant n-1$, *cannot be robust if* (i) $(\ell, k-\ell)$ *with* $\ell < k$ *is the anticipated equilibrium outcome in* $G(k)$ *and* (ii) $(0, k-1)$ *is the anticipated equilibrium outcome in* $G(k-1)$.

An immediate consequence of Lemma 1 is that no first-stage outcome where C players are minority can be robust.

Corollary 2. *A first-stage outcome* $[k, n-k]$ *with* $1 \leqslant k < n/2$ *is not robust.*

We now turn to the outcomes $[k, n-k]$ where $n/2 \leqslant k \leqslant n-1$. We will derive the conditions for $[k, n-k]$ to be robust which depend on the anticipated equilibrium outcomes in the second-stage games $G(k)$, $G(k-1)$ (induced by a C player's deviation) and $G(k+1)$ (induced by a NC player's deviation). We will see that the analysis is substantially simplified by virtue of Lemma 1 above. We first exhibit the payoffs that a player receives if he switched his role in $[k, n-k]$. They will be compared with the payoffs that the player receives in $[k, n-k]$ in order to derive the conditions under which he has no incentive to make a switch and hence $[k, n-k]$ is a robust outcome of the first-stage game. (Recall payoff specifications (1) and (2)).

(C→NC) Suppose that a C player switched his role in $[k, n-k]$ and became a NC player in $[k-1, n-k+1]$. If the (anticipated equilibrium) outcome of $G(k-1)$ is

(a) $(0, k-1)$, then his payoff will be $((k-1)/n)b$,
(b) $(\ell', k-1-\ell')$, where $n-k+1 \leqslant \ell' \leqslant k-1$, then his payoff will be $-a(\ell', n-k+1)$.

(NC→C) Suppose that a NC player switched his role in $[k, n-k]$ and became a C player in $[k+1, n-k-1]$. If the (anticipated equilibrium) outcome of $G(k+1)$ is

(a) $(0, k+1)$, then his payoff will be $((k+1)/n)\, b-c$,
(b) $(\ell'', k+1-\ell'')$, where $n-k-1 \leqslant \ell'' \leqslant k+1$, then his payoff will be

$$\frac{k+1}{n}b + \left(\frac{n-k-1}{\ell''}\right)\frac{k+1}{n}b - c - f(\ell'', n-k-1)$$

if he is one of the ℓ'' C/F players, or $((k+1)/n)b-c$ if he is one of the $k+1-\ell''$ C/NF players.

The next three propositions provide conditions under which $[k, n-k]$ is a robust outcome depending on the anticipated equilibrium outcome in $G(k)$, the second-stage game that will be played if no player switched a role in $[k, n-k]$.

Proposition 4. *Suppose that* $(0, k)$ *is the anticipated equilibrium outcome of* $G(k)$.

(a) *If* $(0, k-1)$ *is anticipated in* $G(k-1)$, *then* $[k, n-k]$ *is not robust regardless of an anticipated outcome in* $G(k+1)$.
(b) *If* $(\ell', k-1-\ell')$, *where* $n-k+1 \leqslant \ell' \leqslant k-1$, *is anticipated in* $G(k-1)$, *and* $(0, k+1)$ *is anticipated in* $G(k+1)$, *then* $[k, n-k]$ *is robust provided that*

$$\frac{n}{k}(c - a(\ell', n-k+1)) \leqslant b. \tag{5}$$

(c) *If* $(\ell', k-1-\ell')$, *where* $n-k+1 \leqslant \ell'' \leqslant k-1$, *is anticipated in* $G(k-1)$, *and* $(\ell'', k+1-\ell'')$, *where* $n-k-1 \leqslant \ell'' \leqslant k+1$, *is anticipated in* $G(k+1)$, *then* $[k, n-k]$ *is a robust outcome provided that*

$$\frac{n}{k}(c - a(\ell', n-k+1)) \leqslant b \leqslant \frac{n\ell''}{\ell'' + (n-k-1)(k+1)}(c + f(\ell'', n-k-1)). \tag{6}$$

Proposition 5. *Suppose that* $(\ell, k-\ell)$, *where* $n-k \leqslant \ell < k$, *is anticipated in* $G(k)$.

(a) *If* $(0, k-1)$ *is anticipated in* $G(k-1)$, *then* $[k, n-k]$ *is not robust regardless of an anticipated outcome in* $G(k+1)$.
(b) *If* $(\ell', k-1-\ell')$, *where* $n-k+1 \leqslant \ell' \leqslant k-1$, *is anticipated in* $G(k-1)$, *and* $(0, k+1)$ *is anticipated in* $G(k+1)$, *then* $[k, n-k]$ *is robust provided that*

$$\frac{n}{k}(c - a(\ell', n-k+1)) \leqslant b \leqslant \frac{n}{k+1}(c - a(\ell, n-k)). \tag{7}$$

(c) *If* $(\ell', k-1-\ell')$, *where* $n-k+1 \leqslant \ell' \leqslant k-1$, *is anticipated in* $G(k-1)$, *and* $(\ell'', k+1-\ell'')$, *where* $n-k-1 \leqslant \ell'' \leqslant k+1$, *is anticipated in* $G(k+1)$, *then* $[k, n-k]$ *is robust provided that*

$$\frac{n}{k}(c - a(\ell', n-k+1)) \leqslant b \leqslant \frac{n\ell''(c + f(\ell'', n-k-1) - a(\ell, n-k))}{(\ell'' + n-k-1)(k+1)}. \tag{8}$$

[11] Thus, our solution concept is that of subgame-perfect (or backward induction) equilibrium. We employ this "rational actor" solution concept for two reasons. First, there seems to be no widely accepted notion of evolutionarily stable equilibrium (ESS) for asymmetric games in extensive form. Rather than using a novel solution concept, we chose a well-studied and uncontroversial notion of subgame perfection that we believe is also a minimum stability requirement in multi-stage interactive situation. Second, the logical basis on which the popularity of ESS rests seems to be that a wide range of dynamic evolutionary processes, e.g., replicator dynamics, converges to it. An important question of what type of dynamic process over our extensive form game has which (if any) subgame-perfect equilibrium as a stable state will be addressed in the future article. In addition, note that we discuss only pure strategy equilibria. There may be a mixed strategy equilibrium. For example, each player's choosing F with probability p (NF with $1-p$) is an equilibrium of $G(k)$ if p satisfies.

Proposition 6. *Suppose that* $(k, 0)$ *is anticipated in* $G(k)$.

(a) *If* $(0, k-1)$ *is anticipated in* $G(k-1)$, *then* $[k, n-k]$ *is not robust regardless of an anticipated outcome in* $G(k+1)$.
(b) *If* $(\ell', k-1-\ell')$, where $n-k+1 \leqslant \ell' \leqslant k-1$, *is anticipated in* $G(k-1)$ *and* $(0, k+1)$ *is anticipated in* $G(k+1)$, *then* $[k, n-k]$ *is robust provided that*

$$c + f(k, n-k) - a(\ell', n-k+1) \leqslant b \leqslant \frac{n}{k+1}(c - a(k, n-k)). \quad (9)$$

(c) *If* $(\ell', k-1-\ell')$, where $n-k+1 \leqslant \ell' \leqslant k-1$, *is anticipated in* $G(k-1)$ and $(\ell'', k+1-\ell'')$, where $n-k-1 \leqslant \ell'' \leqslant k+1$, *is anticipated in* $G(k+1)$, *then* $[k, n-k]$ *is robust provided that*

$$c + f(k, n-k) - a(\ell', n-k+1) \leqslant b \leqslant \frac{n\ell''(c + f(\ell'', n-k-1) - a(k, n-k))}{(k+1)(\ell'' + n - k - 1)}. \quad (10)$$

Corollary 3. *The first-stage outcome* $[n/2, n/2]$ *is not robust.*

Propositions 4–6 cover the boundary cases where $k = n-1$ or n. Indeed, for the first-stage outcome $[n-1, 1]$, a NC player's deviation leads to $[n, 0]$ upon which the game ends and every player receives $b-c$. If we set $k = n$, then all payoffs appearing in (NC→C)(a) and (b) become identically $b-c$.

In the first-stage outcome $[n, 0]$, the game ends and every player receive $b-c$. In addition, there is no NC player present in $[n, 0]$ so we need not consider a NC player's incentive. Thus for the case $k = n$, we deduce the following conclusion from Propositions 4–6.

Corollary 4.

(a) *If* $(0, n-1)$ *is anticipated in* $G(n-1)$, *then* $[n, 0]$ *is not robust.*
(b) *If* $(\ell', n-1-\ell')$, *where* $1 \leqslant \ell' \leqslant n-1$, *is anticipated in* $G(n-1)$, *then* $[n, 0]$ *is robust.*

2.4. Cost of coercion: synchronous vs. asynchronous attackers

The central results of our analysis will be that (1) the costs of coercion are a limiting variable in determining the occurrence of cooperation among individuals with conflicts of interest, and (2) that these costs of coercion are profoundly affected by available strategies for conspecific violence. It is useful to consider two extreme cases, "synchronous" and strictly "asynchronous" attackers as follows.[12]

Asynchronous attackers can only attack one another one at a time. Adult male wild sheep that attack by ramming might be good approximations of such a case. When multiple animals of such a species attack a single target, only one of the multiple attackers is engaged with that target at any moment in time. We argue that all non-human animals are adequately approximated as asynchronous attackers for our purposes.

Synchronous attackers are animals in which multiple attackers can *simultaneously* engage (inflict violence upon) a target animal. As far as we are aware, the only kind of animal who can achieve substantially synchronous attack in relatively large numbers is an animal who can inflict serious violence from a distance. Such an animal might be called a "stand-off" or "remote" attacker. Multiple remote attackers need not interfere with one another (taking separate remote positions) allowing multiple individuals to inflict ongoing damage on the target at the same instant. Note that humans are apparently the first and only animal in Earth's biological history to possess a truly remote or stand-off attacking strategy effective against adult conspecifics—elite aimed throwing (DISCUSSION).

The difference between synchronous and asynchronous attackers has profound effects on the costs of self-interested coercion and, apparently, on viable structures of strategic social interaction (see Section 2.5 below).

To quantify costs of fighting, $f(\ell, n-k)$ and $a(\ell, n-k)$, it is convenient to introduce a parameter, r, corresponding to the cost a fighter would sustain before electing to opt out of further combat, e.g., the amount of injury that would begin to degrade the ability to continue to fight effectively. Alternatively, it might be the amount of injury producing a risk of subsequent death from infection, say. We assume that r is sufficiently high so that $b < r$.[13]

2.4.1. Asynchronous attackers

Using this definition of r it is straightforward to compute each player's costs of an episode of fighting/combat/violence as follows. For asynchronous attackers, when an individual is on the minority side (whether cooperator or free rider), he (or she) will break off conflict after absorbing r units of cost. During this time an individual member of the minority side will have projected r units of cost against the members of the majority side. If these costs are uniformly, randomly distributed among the individual members of the two sides the following expressions describe the costs of each member on each.[14] Thus asynchronous attackers are characterized by the following costs of fighting:

$$f(\ell, n-k) = \begin{cases} r\left(\dfrac{n-k}{\ell}\right) & \text{if } \ell \geqslant n-k, \\ r & \text{if } \ell < n-k. \end{cases} \quad (11)$$

$$a(\ell, n-k) = \begin{cases} r & \text{if } \ell \geqslant n-k, \\ r\left(\dfrac{\ell}{n-k}\right) & \text{if } \ell < n-k. \end{cases} \quad (12)$$

With these cost functions, and the assumption $b < r$, we obtain the following result concerning the equilibrium outcomes of the second-stage games.

Lemma 2. *For asynchronous attackers, the only equilibrium outcome of the second-stage game* $G(k)$, $1 \leqslant k \leqslant n-1$, *is* $(0, k)$, *all C players stand back and let NC players take their shares.*

Recall that $[0, n]$ is a robust outcome of the first-stage game. (Proposition 3) A first-stage outcome $[k, n-k]$ is not robust if $1 \leqslant k < n/2$ (Corollary 2) or if $n/2 \leqslant k \leqslant n$ but $(0, k-1)$ is the anticipated equilibrium outcome of $G(k-1)$, regardless of the anticipated outcomes in $G(k)$ and $G(k+1)$ (Proposition 4(a),

[12] It is likely that real animals will sometimes be intermediate between these two extremes in their capabilities for conspecific violence. However, we will also argue that humans are a very (uniquely) extreme example of a synchronous attacking species. Thus, the dichotomous comparison here probably captures the crucial biological distinctions. It is possible to extend our analysis to encompass much less proficient non-human semi-synchronous killing animals—and their possibly small level of kinship-independent cooperation.

[13] We believe this assumption is very likely to be valid. Serious personal injury (or death, even if rare) can result from fighting. That is, the entire adaptive lifetime of the animal is at risk. In contrast, a typical cooperative enterprise or behavior is likely to yield a much more modest proportional return—a day's food or a night's access to a sleeping territory, for example.

[14] This can be visualized by recognizing that each member of the minority side is engaged with one member of the majority side at all times. In contrast, each member of the majority side may either be engaged or not engaged at any moment. For example, if 10 attack a single target, the target will acquire r units of cost and project r units of cost, leaving each of the 10 attacker with 1/10th r unit of cost. This is a *10-fold* difference in the costs to each attacker relative to the single target.

Proposition 5(a) and Proposition 6(a), and Corollary 3). The next theorem thus follows from Lemma 2.

Theorem 1. *For asynchronous attackers, the only robust equilibrium outcome of the first-stage game is* $[0, n]$, *every player chooses to be a NC player.*

In other words, no cooperation between individuals with conflicts of interest at any scale is possible if the mode of coercive violence available to an animal prohibits simultaneous projection of threats in any degree against potential non-cooperators.[15]

2.4.2. Synchronous attackers

Analogously to the asynchronous case above, we can quantify the costs of fighting for a synchronously attacking animal as follows. Again, when an individual is on the minority side, he will break off conflict after absorbing r units of cost. However, in contrast to asynchronous attackers above, this minority synchronous attacker will be under ongoing, continuous attack from *all* members of the majority side at all instants in time. Thus, a minority individual will accrue costs at a rate proportional to the number of attacker on the majority side and hence the pull-out-triggering cost level r will be reached more quickly than in an asynchronously attacking animal, shortening the combat. As a result, the minority individual will dole out less cost to individuals on the majority side in comparison to the numerically equivalent asynchronous case.[16] If these costs are uniformly, randomly distributed among the individual members of the two sides the following expressions describe the costs to each member on each side.

$$f(\ell, n-k) = \begin{cases} r\left(\dfrac{n-k}{\ell}\right)^2 & \text{if } \ell \geqslant n-k, \\ r & \text{if } \ell < n-k. \end{cases} \tag{13}$$

$$a(\ell, n-k) = \begin{cases} r & \text{if } \ell \geqslant n-k, \\ r\left(\dfrac{\ell}{n-k}\right)^2 & \text{if } \ell < n-k. \end{cases} \tag{14}$$

In addition to the assumption $b<r$, we make the following assumption concerning the size of r relative to b:

$$r \leqslant \frac{(n-1)^2}{n} b.$$

For example, for a modest group size of $n = 10$, this assumption requires that $r \leqslant (8.1)b$. For $n = 20$, it is $r \leqslant (18.05)b$.[17]

Unlike asynchronous attackers, cost-effective coercion reflected in (13) and (14) allows emergence of expanded cooperation among synchronous attackers. In order to make this statement precise, we first analyze equilibrium outcomes of the second-stage games. In particular, we will see that $(k, 0)$, all C players elect to fight, can be an equilibrium outcome of G (k for all k above a certain level which depends on the parameters n, b, and r. Let us rewrite Proposition 2 with the specific cost functions (13) and (14).

Proposition 2′.

(a) $(0, k)$ *is always an equilibrium outcome of* $G(k)$ *for* $1 \leqslant k < n-1$.
(a)′ $(0, n-1)$ *is an equilibrium outcome of* $G(n-1)$ *if* $((n-1)/n)b \leqslant r$.
For $n/2 \leqslant k \leqslant n-1$,

(b) $(\ell, n-k)$ *where* $1 \leqslant \ell < k$ *is an equilibrium outcome of* $G(k)$ *if* $n-k \leqslant \ell < k$ *and* $((n-k)/\ell)r \leqslant (k/n)b \leqslant ((n-k)/(\ell+1))r$,
(c) $(k, 0)$ *is an equilibrium outcome of* $G(k)$ *if*

$$\frac{n(n-k)}{k}\frac{r}{b} \leqslant k. \tag{15}$$

As in the case of asynchronous attackers, $(0, k)$ is an equilibrium outcome of $G(k)$ for $1 \leqslant k \leqslant n-1$. It is clear that the last set of inequalities in (b) is impossible, and hence $(\ell, k-\ell)$, $1 \leqslant \ell < k$, cannot be an equilibrium outcome of $G(k)$. As for $(k, 0)$, it is an equilibrium outcome of $G(k)$ so long as k is large enough to satisfy (15). Fig. 1 at the end of the Supplemental Appendix depicts the relation ship between k and $((n-k)/k)(r/b)$ for some specific values of n and r/b. The threshold value above which $(k, 0)$ becomes an equilibrium outcome can be found by solving the quadratic inequality (a rearrangement of (15)): $k^2 + n(r/b)k - n^2(r/b) \geqslant 0$. This yields

$$k \geqslant \Psi\left(n, \frac{r}{b}\right) \equiv \frac{n}{2}\left(-\frac{r}{b} + \sqrt{\frac{r}{b}\left(4 + \frac{r}{b}\right)}\right). \tag{16}$$

Lemma 3.

(a) *Under the assumption* $1 \leqslant r/b \leqslant (n-1)^2/n$, *we have* $n/2 < \Psi(n, (r/b)) < n-1$.
(b) $\Psi(n, (r/b))$ *is increasing in* n *and* r/b.

The next lemma is an immediate consequence of Proposition 1, Proposition 2′ and Lemma 3.

Lemma 4.

(a) *For* $1 \leqslant k < \Psi(n, (r/b))$, *the only equilibrium outcome of* $G(k)$ *is* $(0, k)$.
(b) *For* $\Psi(n, (r/b)) \leqslant k \leqslant n-1$, *equilibrium outcomes of* $G(k)$ *are* $(0, k)$ *and* $(k, 0)$.

Using the results of Section 2.3 and Lemma 4, we now exhibit the results on the robustness of the first-stage outcomes $[k, n-k]$, $n/2 < k \leqslant n-1$.

Theorem 2. *The first-stage outcome* $[k, n-k]$, *where* $1 \leqslant k \leqslant n-1$, *is robust provided that* $\Psi(n, r/b) < k \leqslant n$, $(0, k)$ *is anticipated in* $G(k)$, *and either*

(a) $(k-1,0)$ *is anticipated in* $G(k-1)$ *and* $(0, k+1)$ *is anticipated in* $G(k+1)$, *or*
(b) $(k-1,0)$ *is anticipated in* $G(k-1)$, $(k+1,0)$ *is anticipated in* $G(k+1)$, *and*

$$b \leqslant \frac{n}{n-k}\left(c + r\left(\frac{n-k-1}{k+1}\right)^2\right). \tag{17}$$

[15] Our treatment assumes equality of fighting ability for simplicity. Larger dominant asynchronously killing animals might be able to enforce some small-scale, limited cooperation among smaller conspecifics, of course.

[16] As with the asynchronous case, a simple numerical example may aid visualization. If 10 synchronous attacker engage a single target, the single target accrues r units of cost before breaking off the conflict/standing down. This requires that each of the 10 attackers has projected 1/10th r units. During the same time interval, the single target will also have projected 1/10th r units of cost. This 1/10th r unit of cost is distributed among the 10 attackers for an individual cost to each attacker 1/100th r units of cost. This is a *100-fold* difference in the costs to each attacker relative to the single target. Contrast this with the *10-fold* difference in costs between a single target and ten asynchronous attackers.

[17] Though empirical/field studies are necessary to determine the circumstances under which this condition is valid, we believe that it will commonly be correct. Note that r reflects the point at which the player elects to break off further participation in violent conflict. In general, animals will evolve to make this choice based on cost–benefit considerations. Thus, incurring r vastly in excess over b (is unlikely to be an adaptive pattern of behavior.

A few remarks are in order. The first concerns the veracity of the condition (17). For $k = n-1$, this inequality can be reduced to $b \leqslant nc$ which is one of the assumptions. Hence this inequality is likely to hold when k is sufficiently close to $n-1$.[18]

Next, the reason that $[k, n-k]$ fails to be robust when $(k, 0)$ is anticipated in $G(k)$ is that a NC player's incentive is violated. This is understandable since, when all k cooperators are ready to punish and k is relatively large, a NC player stands to lose and hence would be better off by switching his role thereby inducing the subgame $G(k+1)$ even if he were to become a C/NF player there. Thus, once a critical number of C players are present, it may cause a cascade of role switching by NC players, ultimately leading to a robust outcome $[n, 0]$, a total cooperation. (See the next theorem). This highlights a sharp difference between asynchronous and synchronous attacking animals.

Theorem 3. *The first-stage outcome* $[n, 0]$ *is robust provided that* $(n-1, 0)$ *is anticipated in* $G(n-1)$.

By the second remark following Theorem 2, it is a reasonable conjecture that $[n, 0]$—all players cooperate—is nearly always a robust outcome for synchronous attackers. Thus, under the appropriate conditions and *only* in synchronous attacking animals, cooperation among individuals with conflicts of interest can be a robustly stable outcome.

2.5. Remote attacking and the higher-order free rider problem

Our analysis indicates that synchronous (remote) attack capability is a prerequisite to expansive kinship-independent social cooperation. However, we believe this fundamental result to be even more robust than is immediately apparent.

First and most importantly, we assume that coercion is costly (including injury from attack by opposing individuals). This assumption creates an implicit higher order free rider problem. Specifically, cooperators who don't leave the field when coercive violence begins (explicitly play C/NF) but pretend to participate in coercion (without subjecting themselves to all or any of its costs) can achieve a higher net return than those who faithfully play C/F.

We suggest that the capacity for effective synchronous attack might uniquely allow social cooperation to be strategically structured so as to obviate this higher-order free rider problem. Specifically, in asynchronously (proximally) attacking animals, conjoint coercive ostracism requires that individuals rotate into and out of ostensible contact with a target in ways that may make monitoring the efficacy of this coercive contact difficult. This creates the well-known, potentially unmanageable regress of higher-order free riding.

In contrast, for a synchronously (remotely) attacking animal, avoiding the costs of coercion (pretending to play C/F) is probably impossible under many circumstances including the specific social situation considered in our model. In particular, when a *potentially* remotely attacking cooperator remains in the presence of the target (does not leave the field) he (or she) remains a threat and will draw fire from the target. To see this, consider an elite throwing ancestral human in the context of "ammunition" (throwing stones) on the ground. Remaining on the field, but refraining from throwing, is analogous to bringing an empty gun to a gun fight. The empty gun toter does not avoid fire (from a target who cannot know his gun is empty). Indeed this individual, *increases* his own risk by not firing and thereby extending the time to suppression of fire from the target. The opportunity for strategically viable higher order free riding does not arise.

Second, our game assumes similar fighting abilities among all players. However, in reality, different animals will generally have somewhat different fighting abilities. In asynchronous attacking animals, this reduces the incentive for less gifted fighters to coercively ostracize (play C/F), thereby increasing the incentive for superior fighters to avoid the costs of cooperation (play NC). This effect further exacerbates the already inadequate return on coercive ostracism in asynchronous attacking animals (see Section 2.4.1 in the text). In other words, our analysis is consistent with the empirical observation that dominant individuals often control the local social environments in non-human (asynchronously attacking) animals.

In contrast, in a synchronous attacker, the increased cost of ostracizing an individual with above-average fighting ability can still be low enough to make ostracism a selfishly enforceable strategy for many or most individuals. This results from the "square effect" in remote attackers (see Eqs. (13) and (14)). Thus, for example, when 10 synchronously attacking C/F player confront a target of comparable fighting ability, their costs are 1/100th of his (or hers). However, if a target NC player has a fighting ability twice as high as the average of the 10 attacking C/F players (a very large difference for a fellow adult conspecific), each C/F individual still incurs only 1/50th of the cost accrued by the target. Taking differential fighting abilities into account in the analysis of Section 2.4.2, it is straightforward to show that $(k, 0)$ (all play C/F) is an equilibrium outcome of the subgame $G(k)$ for sufficiently large k and, in particular, the total cooperative outcome $[n, 0]$ remains robust. In overview, dominance behavior on the basis of individual strength and skill becomes much less important in a synchronously attacking animal (like a human) than an asynchronously attacking animal.

Thus, we suggest that synchronous, remote attacking capability may be decisive, both for the cost-benefit reasons revealed by our analytical treatment above and because of its impact on the achievable strategic logic of social cooperation. If these considerations are correct, the requirement for remote, synchronous attack represents a potentially general, universal law for social cooperation among organisms with conflicts of interest. This possibility will be important to our evolutionary proposal below.

3. Discussion

3.1. Overview

Our investigation of the specific game we analyze here suggests a potentially general rule for systematic, extensive cooperation in the face of conflicts of interest. Specifically, it suggests that such cooperation might be strategically viable only when access to inexpensive coercive threat—provided by synchronous (remote) attacking capability—is available to the players.

The following elements of empirical evidence suggest that this proposal might be of value in understanding both the origin and logic of uniquely human social cooperation. First, humans are the first and only animal in Earth's history to show extensive and intensive kinship-independent conspecific social cooperation—cooperation in spite of conflicts of interest. Second, humans are the first and only animal to possess the innate biological capacity to project coercive threat remotely (synchronously) with high effectiveness against conspecifics. This derives from our unprecedented capacity to throw with sufficient accuracy and violence to kill adult conspecifics with thrown projectiles (e.g., stones) out to distances of at least 10 m. Third, remote projection of coercive threat remains central to our social cooperation through the

[18] The exact threshold value for k above that holds can be calculated by solving a cubic inequality. But the algebraic expression of the solution is too complicated to be illuminating. We also note here that, besides values close to $n-1$, can hold for some values of k close to (but larger than) $n/2$.

present instant. Consider, for example, the pervasive role of gunpowder projectile weapons in law enforcement throughout the contemporary world.

The *empirical* description of our unique social cooperation continues to grow in diverse disciplines. In contrast, our *theoretical* grasp of the origin of this phenomenon has remained substantially incomplete (Fehr and Gachter, 2003; Johnson et al., 2003; Sethi and Somanathan, 2003).

Traditional kin-selection, conventional reciprocity (direct, indirect, network), and group-selection theories—recently embodied in five "rules" for the evolution of cooperation (Nowak, 2006), for example—have the virtue of simplicity and logical coherence. However, their application and implications are sensitive to the assumptions made in abstracting the underlying problem. If these assumptions lack biological verisimilitude, the resulting theory will lack useful relationship to the phenomena ostensibly described. Of these five rules, only the long-established principle of kin-selection has clear empirical content (robust ability to predict animal social behavior)[19] and none of these rules credibly accounts for human uniqueness (below).

For example, a generation of field studies ensuing from a seminal work (Trivers, 1971) has demonstrated that reciprocal altruism between non-kin conspecific animals occurs only in rare, narrowly defined circumstances of uncertain evolutionary logic—except in humans (Koenig, 1988; Krebs and Davies, 1993; Taylor and McGuire, 1988; Wilkinson, 1988). Thus, we need an explanation of the unique scale of reciprocal altruism in humans.

Similarly, (trait) group selection models, pioneered by Wilson (1975a, 1976, 1977) for the evolution of group-beneficial traits in general and strong reciprocity or altruistic punishment in particular (Bowles and Gintis, 2002; Gintis, 2000, 2003), often require a very special set of circumstances: frequent inter-group conflict, small groups, and low migration rates (Wade, 1978). To sustain the vast, pervasive adaptive biological changes associated with the evolution of uniquely human traits (elite symbolic speech, for example), these special conditions would have to recur consistently over extremely long periods of time—and do so uniquely in the human lineage. It will be of great interest to see if the empirical evidence of human uniqueness, in general, and the details of the human paleoanthropological and historical records, in particular, can be deployed to test these requirements.

Culture, in general, and efficient transmission of specific kinds of information (e.g., behavioral norms), in particular, undoubtedly plays an important role in human evolution. (Boyd and Richerson, 2000, 2005; Henrich and Boyd, 2001) The remaining challenge is to understand why expansive, kinship-independent and manipulation-proof transmission of fitness-relevant information became available and adaptive uniquely in the human lineage.

Punishment (post-facto imposition of costs) of offenders has been considered as a factor to enhance cooperation in various earlier models. (Boyd and Richerson, 1992; Hirshleifer and Rasmusen, 1989; Sethi and Somanathan, 1996). We believe that this body of work contains important first steps toward an understanding of the origin of human social cooperation. We suggest that, in order to make further progress toward robust answers to the questions of human uniqueness, it will be important to confront the issues of the evolution of coercive capability and of realistic cost structures associated with specific coercive means.

Our work suggests a framework for understanding the origin of uniquely human social cooperation; it may arise from enforcement of non-kin cooperation through suppression of conflicts of interest as a by-product of immediately, individually self-interested behavior, as suggested originally in (Bingham, 1999). In other words, our analysis suggests that expanded kinship-independent social cooperation can arise *solely* as a result of the instantaneous pursuit of individual self-interest by animals who possess a sufficiently sophisticated capacity for synchronous (remote) projection of coercive threat. Members of a synchronously attacking species are expected to be adapted to living in local groups because of the *individually adaptive* advantages of cooperation as a by-product of ongoing *individually self-interested* coercive threat conjointly with other members (preemptive or compensated coercion), on this view. As a direct result of this coercive threat, each individual will display public behaviors that can be construed as beneficial to other coalition members.[20]

3.2. Empirical evidence for a central role of coercion in human uniqueness

On the basis of our analysis, we propose that the systematic, large-scale employment of self-interested, compensated coercion is a credible candidate for the fundamental unique human property in the sense that flight is the fundamental unique property of birds. Several details of the empirical evidence are consistent with the hypothesis that uniquely human social cooperation *emerges from* the acquisition of the capacity for inexpensive coercion, as strictly required by this proposal.

If expanded kinship-independent social cooperation in humans arose as the result of evolving a self-interested coercive strategy related to the one we model here, it is expected to have emerged rapidly *after* the novel evolution of elite human aimed throwing. The hominid fossil record strongly supports this prediction.

Before assessing the relevant record we note that new brain expansion in the human lineage is very likely to be a reliable symptom of expanded social cooperation, reflecting, in part, extensive, socially supported changes in life history (Bogin and Smith, 1996; Key, 2000). It apparently 'takes a village to raise a [uniquely human] child'. Specifically, human life history supporting brain expansion involves at least three novel stages that arguably could have evolved only in the context of the uniquely extensive, kinship-independent social support as follows. Human newborns are substantially enlarged (relative to maternal body mass), creating an especially challenging late-term pregnancy requiring social support. Moreover, rapid, fetal-like brain growth in human newborns continues through ca. one year of age (unlike non-human animals), creating a behaviorally helpless baby requiring constant monitoring and protection. Finally, human children are weaned long before brain growth ceases (unlike non-human animals) requiring socially provided elite foods (marrow or fat-rich meats, for example) to substitute for mother's milk.

Thus, we can apparently identify the onset of expanding human social cooperation in the fossil record through enlarged cranial capacity in fossil skulls. Our evolutionary proposal requires that access to synchronous (remote) attacking capability must precede this symptom of expanded social cooperation.

[19] We note that kin-selection theory can be construed as a theory of the locus of interest in social behavior rather than as a theory of social "cooperation" *sensu stricto*.

[20] Extrapolating from our game theoretic model, ostracizing cooperation would apparently arise (or evolve) initially in portions of a synchronous attacking population where ostracizing cooperators are, serendipitously, a local majority. However, it is important to notice that each individual in this local population is pursuing instantaneous self-interest; no altruism occurs at any point. Moreover, no individual derives any additional adaptive benefit—beyond the individual benefit from cost-effective ostracism—merely in consequence of being present in this local population or "group". We argue that this process should not be construed as "group selection".

Pre-human hominids (australopiths; ca. 2.4–5 million years ago or mya) showed little or no brain expansion (Aiello and Dean, 1990) and probably could not throw with elite human skill as assessed by shoulder, pelvis, foot, shoulder and hand anatomy, among other criteria (our unpublished analysis of extensive published fossil evidence). Thus, these animals show the trait patterns our proposal requires—no elite throwing and no cranial expansion.

In contrast, the first members of our genus as traditionally defined (*Homo rudolfensis* and African *ergaster/erectus*; ca. 1.8 mya) showed significant brain expansion (Aiello and Dean, 1990) and could apparently throw with elite human skill as assessed by as assessed by shoulder, pelvis, foot, shoulder and hand anatomy (our unpublished analysis of extensive fossil evidence). For example, the shoulder joint is central to elite human throwing and this joint is substantially redesigned in all members of *Homo* relative to australopiths, including in the earliest known humans (Aiello and Dean, 1990).

Finally, two pieces of evidence provide direct support for the prediction that elite human throwing *briefly* preceded the relatively explosive emergence of substantially expanded social cooperation (brain expansion) as our evolutionary proposal requires. In one case, fragmentary evidence indirectly suggests that very late pre-human African hominids (ca. 2.3 mya) evolved elite throwing as part of a professional hunting and/or "power scavenging" adaptation before significant new brain expansion (Asfaw et al., 1999; Leaky, 1979; Potts, 1988). This new foraging adaptation represents the most likely adaptive opportunity leading to initial selection for remote attack capability in the immediate pre-human ancestor.

In a second case, recent, exciting findings from the ongoing excavation at Dmanisi (Georgia, Asia) strongly suggest that these very early members of *Homo* have postcranial skeletons redesigned for elite throwing (from shoulder, femur and first metatarsal morphology) and used elite throwing for power scavenging or hunting (from importation of a large number of missile-sized manuports into a setting with hominid butchering). In contrast to later humans, these earliest members of *Homo* show only modest, variable new brain expansion (Fischman, 2005; Lordkipanidze et al., 2007).

One last issue is noteworthy before leaving the question of empirical support for coercion as the limiting factor in the emergence of human social cooperation. Recent historical increases in the scale of human social cooperation—reflecting management of the conflict of interest problem on a new scale—are associated with prior acquisition of a new coercive technology. Examples include the bow and "agricultural civilizations" of pre-colonial contact North America (Blitz, 1988) and gunpowder weaponry and the "modern state" (reviewed in Porter, 1994; Hall, 1997). It will be of interest to determine if the approach illustrated by our model can be modified and generalized to account for these empirical correlations in detail.

3.3. Evolved human psychology

The human mind is expected to consist of evolved proximate psychological devices often generating strategically viable, species—typical behavior under the appropriate circumstances. Our evolutionary proposal above suggests that humans may be adapted to social cooperation based on the self-interested projection of coercive threat. If so, our minds should be highly adapted to a very specific pattern of pre-emptive or compensated coercion in the presence of other non-kin conspecifics (in "public"). Though it is beyond the scope of this manuscript to review them in detail, diverse features of our ethical psychology and social/political/economic behavior support these expectations. See, for example, (Price et al., 2002). In one especially simple, yet salient example, contemporary humans take to the streets conjointly with projectile weapons (including thrown stones) in defense of individual political/economic interests shared with large numbers of others. This behavior is cross-culturally universal. Humans are also expected to construe their interests as congruent with coercive groups of which they are members, as they commonly do.

However, we wish to concentrate here on several specific details from experimental social psychology/economics that appear, superficially, to contradict the predictions of our evolutionary proposal, as these particular results have received attention recently.

In diverse, elegant laboratory studies (Fehr and Gachter, 2002) humans consistently punish altruistically—not in an immediately self-interested way. Altruistic punishment can be logically viewed as a group-selected trait and these observations are consistent with a group selection hypothesis (Boyd et al., 2005). As argued above, however, the empirical-biological verisimilitude of such models needs to be closely examined. Alternatively, formally altruistic punishment in these studies can be interpreted as misapprehension of an adaptively novel, artificial setting by the experimental subjects. Several recent discussions of these studies raise technical concerns very similar to ours. (Hagen and Hammerstein, 2006; Haley and Fessler, 2005; Levitt and List, 2007).

Specifically, to borrow a familiar example, this misapprehension is potentially similar to the well-understood male sexual response to two-dimensional representations of reproductive females (the *Playboy effect*). The Playboy effect is quite illuminating when properly interpreted; however, manifestly, it does not require us to assume adaptive ancestral male sexual activity with two-dimensional images of females.

It has long been recognized that responses in the experimental psychology laboratory may be vulnerable to a more subtle, but analogous effect. Such responses can, nevertheless, also be highly informative *when properly interpreted*. Such an interpretation may require us to be aware that the ancestral human mind (which we inherit) is adapted to social circumstances in which anonymous interactions occurred only extremely rarely in a social world dominated by "public" interactions among familiar individuals. "Modern" anonymous, isolated situations were probably not an important adaptive challenge throughout most of the vast human evolutionary past. Thus, humans may "interpret" artificial experimental situations as the adult male mind "interprets" two-dimensional images of reproductive females—as something other than the situation literally in view.

In view of these considerations, the re-interpretation of these experimental results suggested by our evolutionary proposal has two parts as follows. First, the "post facto" quality of the observed "punishment" may be misleading. Specifically, punishing experimental subjects are (unconsciously) misinterpreting the (relatively trivial) non-cooperative experimental behaviors of other subjects as *expressions of intent* to engage in more substantial free riding or cheating in the immediate future. Thus, "punishing" subjects are actually attempting to engage (unconsciously) in pre-emptive coercion of these impending, more substantive defections—or signaling their willingness to so engage.

Second, the altruistic quality of this ostensibly posto facto punishment or retribution likewise may result from the (unconscious) misapprehension of the social situation. The experimentalist and subjects "know" (consciously) that the subjects are alone and anonymous; however, the subjects are incapable of (unconsciously) internalizing this evolutionarily novel context. Instead, subjects behave as if they are (actually or potentially) in

the presence of others—including, of course, the experimentalist—even when the design of the experiment creates technically impenetrable subject anonymity and/or individual isolation.

Given these two potential misapprehensions, a subject's "altruistic punishment" might actually represent a public (unconscious) expression of the intent to profitably coerce in concert with others (Fehr and Gachter, 2003; Johnson et al., 2003; Price et al., 2002).

New theoretical approaches reveal the need for new experimental design. Even the most oblique cues of being "watched" influence an experimental subject's behavior (Haley and Fessler, 2005). Thus, humans may never be able to behave as if they are truly and completely alone and anonymous in an experiment. The very fact of being a living human may (unconsciously) imply the risk of being watched by strategically salient others. It will be challenging, but not necessarily impossible, to design approaches that would support/falsify this suggestion.

Acknowledgments

The authors would like to thank many colleagues for valuable conversations during the long course of this work, including George C. Williams, the late William D. Hamilton, Robert Trivers, Sandro Brusco, David S. Wilson, Philip Rightmire, Philip Tobias, Leslie Aiello, Robert Blumenschine and John Shea. P.B. is also especially grateful to Joanne Souza and Zuzana Zachar for many vital discussions.

Appendix A. Supporting Information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jtbi.2008.02.041.

References

Aiello, L.C., Dean, C., 1990. An Introduction to Human Evolutionary Anatomy. Academic Press, London.
Alexander, R.D., 1987. The Biology of Moral Systems (Foundations of Human Behavior). Aldine de Gruyter, New York.
Asfaw, B., White, T., Lovejoy, O., Latimer, B., Simpson, S., Suwa, G., 1999. *Australopithecus garhi*: a new species of early Hominid from Ethiopia. Science 284, 629–635.
Axelrod, R.M., 1984. The Evolution of Cooperation. Basic Books, New York.
Bingham, P.M., 1999. Human uniqueness: a general theory. Q. Rev. Biol. 74, 133–169.
Blitz, J.H., 1988. Adoption of the bow in prehistoric North America. North Am. Archaeol. 9, 123–145.
Bogin, B., Smith, B.H., 1996. Evolution of the human life cycle. Am. J. Hum. Biol. 8, 706–716.
Bowles, S., Gintis, H., 2002. Homo reciprocants. Nature 415, 137–140.
Boyd, R., Richerson, P.J., 1992. Punishment allows the evolution of cooperation (or anything else) in sizable groups. Ethol. Sociobiol. 13, 171–195.
Boyd, R., Richerson, P.J., 2000. Solving the puzzle of human cooperation. In: Levinson, S. (Ed.), Evolution and Culture. MIT Press, Cambridge, MA.
Boyd, R., Richerson, P.J., 2005. Not by Genes Alone: How Culture Transformed Human Evolution. University of Chicago Press, Chicago.
Boyd, R., Gintis, H., Bowles, S., Richerson, P.J., 2005. The evolution of altruistic punishment. In: Gintis, H., et al. (Eds.), Moral Sentiments and Material Interests. MIT Press, Cambridge, MA, pp. 215–227.
Darwin, C., 1871. The Descent of Man, and Selection in Relation to Sex. J. Murray, London.
Dawkins, R., 1976. The Selfish Gene. Oxford University Press, USA, New York.
Dugatkin, L.A., 1997. Cooperation among Animals: An Evolutionary Perspective. Oxford University Press, USA, New York.
Fehr, E., Gachter, S., 2002. Altruistic punishment in humans. Nature 415, 137–140.
Fehr, E., Gachter, S., 2003. The puzzle of human cooperation—reply. Nature 421, 912.
Fischman, J., 2005. Family ties. Natl Geogr. 207, 16–27.
Gavrilets, S., Vose, A., 2006. The dynamics of Machiavellian intelligence. Proc. Natl Acad. Sci. USA 103, 16823–16828.
Gintis, H., 2000. Strong reciprocity and human sociality. J. Theor. Biol. 206, 169–179.
Gintis, H., 2003. The Hitchhiker's guide to Altruism: gene-culture coevolution, and the internalization of norms. J. Theor. Biol. 220, 407–418.
Gurerk, O., Irlenbusch, B., Rockenbach, B., 2006. The competitive advantage of sanctioning institutions. Science 312, 108–111.
Hall, B.S., 1997. Weapons and Warfare in Renaissance Europe: Gunpowder, Technology, and Tactics. Johns Hopkins University Press, Baltimore, MD.
Hagen, E.H., Hammerstein, P., 2006. Game theory and human evolution: a critique of some recent interpretations of experimental games. J. Theor. Popul. Biol. 69, 339–348.
Haley, K.J., Fessler, D.M.T., 2005. Nobody's watching? Subtle cues affect generosity in an anonymous economic game. Evol. Hum. Behav. 26, 245–256.
Hamilton, W.D., 1964a. The genetical evolution of social behaviour I. J. Theor. Biol. 7, 1–16.
Hamilton, W.D., 1964b. The genetical evolution of social behaviour II. J. Theor. Biol. 7, 17–52.
Hamilton, W.D., 1996. Narrow Roads of Gene Land. W.H. Freeman, New York.
Henrich, J., Boyd, R., 2001. Why people punish defectors-weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. J. Theor. Biol. 208, 79–89.
Hirshleifer, D., Rasmusen, E., 1989. Cooperation in repeated prisoners' dilemma with ostracism. J. Econ. Behav. Organ. 12, 87–106.
Johnson, D.C.D., Stopka, P., Knights, S., 2003. The puzzle of human cooperation. Nature 421, 911–912.
Key, C.A., 2000. The evolution of human life history. World Archaeol. 31, 329–350.
Koenig, W.D., 1988. Reciprocal altruism in birds: a critical review. Ethol. Sociobiol. 9, 73–84.
Krebs, J.R., Davies, N.B., 1993. An Introduction to Behavioural Ecology. Blackwell Scientific Publications, Oxford.
Leaky, M.D., 1979. Olduvai Gorge: My Search for Early Man. Collins, London.
Levitt, S.D., List, J.A., 2007. What do laboratory experiments measuring social preferences reveal about the real world? J. Econ. Perspect. 21, 153–174.
Lordkipanidze, D., Jashashvili, T., Vekua, A., de Leon, M.S.P., Zollikofer, C.P.E., Rightmire, G.P., Pontzer, H., Ferring, R., Oms, O., Tappen, M., Bukhsianidze, M., Agusti, J., Kahlke, R., Kiladze, G., Martinez-Navarro, B., Mouskhelishvili, A., Nioradze, M., Rook, L., 2007. Postcranial evidence from early Homo from Dmanisi, Georgia. Nature 449, 305–310.
Maynard Smith, J., 1982. Evolution and the Theory of Games. Cambridge University Press, Cambridge, UK.
Maynard Smith, J., Szathmâary, E., 1995. The Major Transitions in Evolution. W.H. Freeman Spektrum, New York.
Nowak, M.A., 2006. Five rules for the evolution of cooperation. Science 314, 1560–1563.
Porter, B.D., 1994. War and the Rise of the State, first ed. The Free Press, New York.
Potts, R., 1988. Early Hominid Activities at Olduvai. Foundations of Human Behaviour, Aldine de Gruyter, New York.
Price, M.E., Cosmides, L., Tooby, J., 2002. Punitive sentiment as an anti-free rider psychological device. Evol. Hum. Behav. 23, 203–231.
Ratnieks, F.L.W., Foster, K.R., Wenseleers, T., 2006. Conflict resolution in insect societies. Annu. Rev. Entomol. 51, 581–608.
Rockenbach, B., Milinski, M., 2006. The efficient interaction of indirect reciprocity and costly punishment. Nature 444, 718–723.
Sethi, R., Somanathan, E., 1996. The evolution of social norms in common property resource use. Am. Econ. Rev. 86, 766–788.
Sethi, R., Somanathan, E., 2003. Understanding reciprocity. J. Econ. Behav. Organ. 50, 1–27.
Sober, E., Wilson, D.S., 1998. Unto Others: The Evolution and Psychology of Unselfish Behavior. Harvard University Press, Cambridge, MA.
Taylor, C.E., McGuire, M.T., 1988. Reciprocal altruism: 15 years later. Ethol. Sociobiol. 9, 67–72.
Trivers, R.L., 1971. The evolution of reciprocal altruism. Q. Rev. Biol. 46, 35–57.
Wade, M.J., 1978. A critical review of the models of group selection. Q. Rev. Biol. 53, 101–114.
Wilkinson, G.S., 1988. Reciprocal altruism in bats and other mammals. Ethol. Sociobiol. 9, 85–100.
Williams, G.C., 1966. Adaptation and Natural Selection—A Critique of some Current Evolutionary Thought. Princeton University Press, Princeton, NJ.
Williams, G.C., Williams, D.C., 1957. Natural-selection of individually harmful social adaptations among sibs with special reference to social insects. Evolution 11, 32–39.
Wilson, D.S., 1975a. A theory of group selection. Proc. Natl Acad. Sci. USA 72, 143–146.
Wilson, E.O., 1975b. Sociobiology: The New Synthesis. Harvard University Press, Cambridge, MA.
Wilson, D.S., 1976. Evolution on the level of communities. Science 192, 1358–1360.
Wilson, D.S., 1977. Structured demes and the evolution of group-advantageous traits. Am. Nat. 111, 157–185.