Hierarchy Maintenance, Coalition Formation, and the Origins of Altruistic Punishment

Yasha Rohwer†‡

Game theory has played a critical role in elucidating the evolutionary origins of social behavior. Sober and Wilson (1999) model altruism as a prisoner's dilemma and claim that this model indicates that altruism arose from group selection pressures. Sober and Wilson also suggest that the prisoner's dilemma model can be used to characterize punishment; hence, punishment too originated from group selection pressures. However, empirical evidence suggests that a group selection model of the origins of altruistic punishment may be insufficient. I argue that examining dominance hierarchies and coalition formation in chimpanzee societies suggests that the origins of altruistic punishment may be best captured by individual selection models. I suggest that this shows the necessity of coupling of game-theoretic models with a conception of what our actual social structure may have been like to best model the origins of our own behavior.

1. Introduction. Sober and Wilson (1999), Skyrms (1994), and Kitcher (1999) have recently shown how game theory can help make clear the Darwinian origins of certain social behaviors. In addition, some have called for a better understanding of the actual social dynamics of the societies in which these hypothetical games would be played.

Skyrms suggests in his book, *The Evolution of the Social Contract,* that "in bargaining situations between more than two people, coalitions may play a crucial role" (Skyrms 1994, 108). Kitcher concurs with Skyrms, in his commentary on Skyrms' book, saying "we cannot ignore attention to coalitions" (1999, 228). Kitcher even suggests that we examine chimpanzee societies for this information, and goes so far as to outline a model of coalition formation.

I agree with these authors. We should pay attention to the social dy-

†To contact the author, please write to: Department of Philosophy, University of Maryland, Skinner Building, College Park, MD 20742; e-mail: yrohwer@umd.edu.

‡I would like to thank Zachary Ernst and Emma Marris for their many helpful comments which greatly improved this paper.

Philosophy of Science, 74 (December 2007) pp. 802–812. 0031-8248/2007/7405-0021$10.00

namics of the societies where these games would have been played. I also accept, with Kitcher, that chimpanzee societies are the best place to look for this information. However, game theorists have not taken advantage of the characterization of chimpanzee social structure that exists in primatology literature.

Dominance hierarchies are the overriding social structure of chimpanzee societies (de Waal 1982, 1989; Goodall 1986, 1992; Smuts 1987; Boehm 1999; Falk 2000). The details of how dominance hierarchies are formed and maintained in chimpanzee societies can offer superior explanations of certain behaviors whose origins may be misdiagnosed by traditional game-theoretic models.

To show this, I will examine altruistic punishment. Altruistic punishment is punishment that indirectly benefits others at a cost to the punisher, as in the case of punishment of a free rider. I will focus on Sober and Wilson's explanation of the origin of altruistic punishment from their book *Unto Others.* I will argue that, although Sober and Wilson offer a plausible group selection explanation of the origins of altruistic punishment, empirical data on chimpanzee behavior suggests an alternative hypothesis.

By appealing to the dominance hierarchy and its maintenance, I think it can be shown that certain behaviors that can be covered by the definition of altruistic punishment need not have arisen from group selection pressures. Examining hierarchy maintenance shows that in certain cases altruistic punishment benefits the individual that punishes more than the rest of the group, and that these benefits significantly outweigh the costs to the punisher. I will also suggest that dominance hierarchies can be used to understand coalition formation, especially coalitions that form specifically to punish. I believe that individual selection pressures could have produced the phenomenon of altruistic punishment.

2. Punishment as Altruism. If one individual punishes another, the individual who punishes incurs a cost by expending time, energy, and perhaps taking on considerable risk, while other members of the group reap the benefits, by having the undesirable behavior of the punished curtailed or discouraged. Individual selection theory suggests that this type of behavior would not evolve, and yet social animals do engage in punishing behavior. Sober and Wilson, in their book *Unto Others* (1999), offer a solution. They claim that punishment is a form of altruism, and must have arisen by the mechanism of group selection.

Sober and Wilson believe that altruism is best modeled as a prisoner's dilemma game. In this game, there are two strategies that can be employed when interacting with the other: one can be selfish or altruistic. If you act altruistically, you raise the other individual's payoff by 4 while taking

a loss of 1.[1] If you act selfishly you do not raise another individual's payoff and you incur no loss. So, if an altruist interacts with another altruist, they both get a payoff of 3 (+4 from the other's action and −1 from their own action). If an altruist and a selfish individual interact, then the selfish individual gets a payoff of 4 while the altruist takes a loss of 1. And if two selfish individuals interact, they both get 0, receiving no payoff and taking no loss. Altruism is not a winning strategy within a mixed group of two individuals, because altruists take a loss of 1 while the selfish individuals get a payoff of 4. However, altruism is a winning strategy between groups, because a group of two altruists receives a net payoff of 6, and a group composed of a selfish individual and an altruist receives a net payoff of 3, while a group of two egoists receive no payoff at all. Thus, Sober and Wilson note that "[altruism] can be exploited within groups but nevertheless evolves because groups of altruists do better than groups of exploiters" (1999, 86).

Sober and Wilson see punishment as a form of altruism. Punishment is altruistic not toward the punished, but toward the other members of the group, since the undesirable action of the punished is curtailed.[2] A punisher incurs a loss while other members of the group receive a payoff, presumably giving the group at large an advantage against other groups. Therefore, punishment, understood as a form of altruism, can be modeled as a prisoner's dilemma as well. Since punishment is a form of altruism, and can be modeled by the same game, Sober and Wilson argue that it evolved by the same mechanism: group selection.

Sober and Wilson use the case of meat sharing to illustrate their proposed group-selectionist origin of punishment. They begin their analysis by characterizing—and then objecting to—the predominant individual selectionist conception of meat sharing. Imagine a group of hunter-gatherers, who, when sharing the meat after the hunt, do not favor family or those who could reciprocate. The meat is shared in a strictly egalitarian fashion, regardless of who was the best hunter. Even so, we find that the best hunters reap some rewards when they share out their meat. They avoid punishments associated with *not* sharing, and they end up siring more children. So in the end it appears that the hunters are egoists. This standard analysis can be expanded to explain the behavior of the punishers as well. They too appear to be egoists, because in punishing those that do not share, they receive meat.

1. The exact numbers used do not matter; however, they must reflect the asymmetric relationship between the altruist and the egoist.

2. The individual being punished is taking a loss, as is the individual who is punishing. This interaction has been characterized as spite (Wilson 1979). Sober and Wilson refer to it as altruism because of the indirect benefit to others.

Although this characterization of meat sharing in hunter-gathering societies may be predominant, Sober and Wilson believe "that the facts don't fit comfortably within the framework of individual selection theory" (1999, 143). They think that although punishment in meat sharing may seem to be egoistic, when you do the numbers, it is clearly a case of altruism, and so must be reanalyzed through the lens of group selection.

Following Ellickson (1991), Sober and Wilson's interpretation begins by designating sharing as a "primary behavior" and punishing as a "secondary behavior." "By itself, the primary behavior [sharing] increases the fitness of the group and decreases the relative fitness of the hunter within the group" (1999, 143). Alone, the primary behavior of sharing meat convincingly looks to them as though it arose through group selection.

The same analysis holds for the secondary behavior, punishment. Forcing an individual to share by threatening punishment—or by punishing—indirectly increases the fitness of the group, while incurring a cost to the individual who punishes. As Sober and Wilson state, "the secondary behavior increases the fitness of the group by causing the primary behavior to be expressed and decreases relative fitness within the group, if there is any time, effort, or risk associated with performing the secondary behavior" (1999, 145). Therefore, punishing and sharing both benefit the group at a cost to the individual who punishes or shares, and so both seemingly must have originated through between group selection pressures.

Sober and Wilson's analysis of punishment as altruism has been widely cited in recent literature on cooperation (Fehr and Gächter 2002; Boyd et al. 2003; Fehr and Fischbacher 2003; de Quervain et al. 2004). Although Sober and Wilson never explicitly use the terminology 'altruistic punishment', I will refer to their group selectionist account as the model of the origins of altruistic punishment.

Sober and Wilson's model of the origins of altruistic punishment is plausible, and they are correct to take seriously empirical research on the behavior. However, an analysis of the actual behavior of our closest relative, the chimpanzee, suggests an alternative model. If we take seriously the notion that chimpanzee behavior can help us elucidate the origins of our own behavior, then we must analyze chimpanzee altruistic punishing behavior when theorizing on the origins of our own punishing behavior. In chimpanzee societies, I will note that most altruistic punishments can indeed be made to fit Sober and Wilson's model. However, an analysis of the social structure suggests that altruistic punishment has the function of keeping the top raking male, or coalition of males, on top, or preserving the troop-level macrocoalition that disproportionately serves the interests of those on top. This observation suggests an explanation of altruistic punishment that is closer to the traditional individual selection account.

3. The Linear Dominance Hierarchy. Primatology literature has an abundant supply of excellent descriptions of the workings of chimpanzee society. Chimpanzee communities are usually composed of 50–100 animals. Females are able to join other communities, but males stay in their birth community. Males are organized in a "hierarchy of levels" (Bygott 1979; Goodall 1986; Boehm 1999), known as a 'linear dominance hierarchy'. Boehm (1999, 24–25) explains, "the adults are aligned in a simple, linear dominance hierarchy with the alpha outranking everyone, the number two outranking all but the alpha, and so on down the line."

Signals, both vocal and nonvocal, help identify the alpha male in a given chimpanzee society. The alpha male maintains his head position by constantly displaying (de Waal 1982, 1989; Goodall 1986; Smuts 1987; Boehm 1999; Falk 2000). Boehm explains that displaying involves acts of aggression, such as, "running, stamping, slapping the ground, swinging on vines, dragging branches, uprooting small trees, and scooping stones and large rocks into the air" (1999, 20). Boehm continues by noting that these acts are usually met by signs of supplication by lesser males, such as "crouching, bobbing, presenting one's rump, sleeking the hair, and exhibiting a fear-grin" (1999, 24), along with a variety of supplication sounds.

When a young male chimp reaches the appropriate age, he seems to be overcome by social ambition (de Waal 1982; Goodall 1986; Boehm 1999). The young chimp enters the hierarchy by first displaying to females. Although he may encounter setbacks along the way, the male eventually becomes dominant to all the females and begins to display to low ranked males.

Climbing the hierarchy is highly advantageous. As Boehm explains, "a dominant position leads to better access to food resources, and for the typically promiscuous males, high rank confers better mating opportunities" (1999, 26). Smuts (1987, 388) notes that "A large number of studies indicate that high rank is often associated with increased mating activity." Goodall mentions that alpha males have the power to "monopolize" a female during her fertile period (1986, 451). De Waal also notes the astounding benefits secured by the alpha male in his captive chimps; "[w]hen Yeroen was the alpha male he alone was responsible for about three-quarters of all matings. Not counting sexual intercourse with young females (who arouse less rivalry) his share was almost 100 percent" (1982, 167–168). These observations, especially de Waal's, show that, from a Darwinian perspective, the alpha has much greater fitness than lower ranking males.

4. Distinguishing Punishment and Aggression. Chimpanzees can be very aggressive. Therefore, before analyzing altruistic punishment in chimpan-

zee societies, it is necessary to make a distinction between aggression and punishment. For humans this distinction can be made easily; punishment is usually associated with morality (Feinberg 1994). Punishments act as moral equalizers (they right wrongs), as moral education, as expressions of moral disapproval, and as deterrents for future immoral actions. As far as we know, the moral element is not present in chimps. But, since we are discussing the very origins of punishment, the absence of moral factors should not be surprising.

Punishment does more than morally condemn; it also aims to change the future actions of the individual being punished, as when a child is punished by her parents. In order to distinguish acts of aggression from acts of punishment in chimpanzees, I propose that behavior having the function of modifying another individual's behavior can be considered an instance of punishment, while behavior that does not have that function should not.

When an alpha male joins a group of individuals from his band he will typically charge at the group screaming, hair bristled, sending them running in fear. In this instance, it seems that the alpha male's behavior serves only to maintain the subordinate behavior of the others. But when the alpha male charges a specific individual who has challenged his dominance, the alpha is attempting to modify that individual's behavior, in order to return the individual to a subordinate position. This distinction allows us to differentiate punishment from mere aggression.

5. The Traditional Model of Altruistic Punishment in Chimpanzee Societies. Once we have distinguished punishment from aggression, it is possible to propose how the dominance hierarchy helps explain some forms of altruistic punishment. For example, alpha and high ranking males have been known to break up fights between individuals and punish the participants.

Males must move up the hierarchy, and this is done by challenging others and becoming dominant to them. These challenges can erupt into serious fights. Goodall (1986) notes that a chimp vying to become the alpha male was found dead the next morning by his keepers. Even if the fights do not end in death, they can result in very serious wounds.

However, not all fights are allowed to continue. Boehm notes that "not infrequently, at Gombe, the alpha male assumes a 'control role' to stop significant fights that break out" (1999, 26). The alpha male usually displays in front of the two fighting males and then sits in between the two to prevent them from resuming their conflict. Dominant males have even been known to physically pry apart two fighting males and then beat them

(Goodall 1986). Only dominant males intervene and the punishment is usually inflicted on both parties.[3]

This type of intervention and punishment could be analyzed using Sober and Wilson's model of the origins of altruistic punishment. It might be thought that the alpha male's action is altruistic because it helps preserve the macrocoalition, which serves the interest of the entire group.[4]

All the males of the group form a macrocoalition, which routinely patrols the territory that their community lives in (Goodall 1986; Boehm 1999). These troop-level coalitions are also known to infiltrate other territories in search of food. If two groups from different communities meet, the reaction from both parties is usually to menacingly howl at one another and withdraw into their own territory. Withdrawing is not the only possible outcome; they do have scuffles and chimps can be killed, especially if one macrocoalition significantly outnumbers the other.

The macrocoalition protects the group's territory not only from the attacks of other groups, but also from predators. Boehm tells of chimps banding together to scare away pythons and leopards—dangerous predators, but no match for a large coalition of chimps.

Macrocoalitions help the entire group, but in order to be effective there must be a significant number of healthy males participating. Fights that occur between males in the hierarchy can sometimes result in serious injury, or death. But males are constantly trying to move up in the hierarchy, making these fights inevitable. Using Sober and Wilson's model it would seem that, since it is the dominant male that expends time and energy intervening when fights become "significant" or "serious," as Boehm puts it, and punishes the two participants, the dominant males' action fits the traditional model of altruistic punishment.

6. The Missing Factors in the Traditional Model of Altruistic Punishment. The analysis from the previous section, although partially correct, leaves out important factors that can help explain the alpha male's behavior. It is the alpha male who is performing the altruistic punishment, but it is also the alpha male who, in preserving the macrocoalition, benefits the most. It is the alpha male who gets first access to the food in the protected territory. And more importantly, because of his status he has the most mating opportunities with females on the protected territory.

3. Conflict regulation by high ranking males has also been observed in pigtailed macaques (Flack, de Waal, and Krakauer 2005).

4. Alpha males are known to break up fights of females as well (Goodall 1986; Boehm 1999). But males are far more aggressive than females, on account of their attempts to move up the hierarchy, so I am considering only fight regulation and punishment involving males.

Because the alpha male has the most mating opportunities it is likely that he has sired the majority of the troop's offspring. So, in preserving the macrocoalition, the alpha male also benefits because his progeny are protected by the coalition from the dangers of the jungle.

Unlike the potential punisher in Sober and Wilson's meat sharing case, who would expend time and energy so that everyone in the community would have an equal share, the actual punisher in chimpanzee communities spends time and energy so that he can have the first choice of the food resources and priority at mating opportunities. But this does not mean that the punishment was not altruistic. The alpha male's action does benefit the group. However, of the entire group it is the alpha male who benefits the most. Maintaining the macrocoalition by punishing fighting individuals would therefore be an important priority for the alpha male and the other high ranking males in the community.

Also, it should be noted that altruistic punishment is not a very risky behavior for alpha or high ranking males. Because of their status in the dominance hierarchy, it is easy to modify the behavior of others. The benefits of being the alpha male, though, are great. The alpha male's action may cost him time and energy but the risk is minimal. The benefits accrued undoubtedly outweigh the slight risk involved. The minimal cost incurred by the alpha further illustrates that the altruistic punishment is not well captured by the traditional game-theoretic model of altruism.

7. The Dominance Hierarchy and Coalitions That Perform Altruistic Punishments. As we have seen, altruistic punishment in chimpanzee communities is usually executed by the alpha male, and this punishment directly serves his interest by maintaining the macrocoalition that disproportionately benefits him. However, Boehm does mention other instances where a microcoalition of chimpanzees has succeeded in altering the behavior of an alpha male. For example, Boehm mentions a case in which the younger chimps banded together to keep an ex–alpha male from returning to his former position (Boehm 1999). I believe that this behavior can be captured by the traditional model of altruistic punishment. However, as in the last case, a better model is supplied by understanding the dominance hierarchy.

In chimpanzee society, usurping an alpha male can involve a serious fight. Boehm describes a fight between the alpha male Goblin, and a rival—Wilkie—that ended with Goblin suffering severe wounds (Boehm 1999). Goblin would have most likely have died if there had not been a veterinarian in the anthropologists' camp at the time. He did, however, lose the alpha position in the group, and in his absence, every other male in the hierarchy moved up a notch. When Goblin was healed he returned to his community and tried to regain his former position as alpha male,

but was gang-attacked by a coalition (Goodall 1986, 1992). Boehm describes the encounter, originally described by Goodall: "when Goblin tried to enter . . . most of the chimpanzee males united as a single large coalition and collectively dominated their former leader when his intention was to dominate them again" (1999, 160).

Having a microcoalition punish an individual would seem to lend itself nicely to Sober and Wilson's explanation of altruistic punishment. Goodall notes that "Goblin had been a very tempestuous alpha male, continually disrupting peacefully grooming or resting chimpanzees with his repeated and vigorous charging displays" (1992, 139). The microcoalition's actions certainly fit Sober and Wilson's model because, by not allowing Goblin to return to his alpha position, the males were preventing a "tyrant" from returning and therefore benefiting the group as a whole. By supporting the new alpha they were also acting altruistically towards Wilkie. However, we can only infer that the act primarily functioned to benefit others because we do not know who participated and what their rank was. But, Boehm does tell us that Goblin was eventually let back into the community at a very low rank (Boehm 1999). The altruistic punishment of Goblin is better explained by each participating chimp's desire to maintain his own new position in the hierarchy.

Goodall's previous observation about Goblin's first fall from power supports this conjecture. As has been noted, alpha males can lose their rank and fall significantly. In order to reclimb the hierarchy, they must dominate all males ranked above them. Goodall notes that when Goblin first lost his alpha position and tried to regain the top seat, "the *senior males* repeatedly and enthusiastically supported one another against Goblin" (1986, 435, emphasis mine). Hence their participation was presumably to maintain their newly acquired rankings. It was the participants of the coalition in this situation that benefited the most from the altruistic punishment.

But the same holds for the second fall from power, where Goblin was gang-attacked. Younger males participated in the attack, but their participation helped them become dominant to Goblin, because, as Goodall notes, when he returned "he now [occupied] a very subordinate position" (1992, 141). Status maintenance, especially of a newly acquired place in the hierarchy, is therefore the best explanation for coalitions that form to punish.

8. Conclusion. These field reports suggest that altruistic punishment need not have originated by group selection, as the Sober and Wilson model assumes. Seen through the lens of the linear dominance hierarchy, it is reasonable to suspect that altruistic punishment may have originated primarily through individual selection pressures; for although chimp pun-

ishment does indirectly benefit others in the group, the punishing behavior most directly serves the interests of the individual or individuals who are executing the punishment.

This reexamination of the origins of altruistic punishment is not necessarily opposed to Sober and Wilson's group selection explanation. Indeed, their analysis seems to be the appropriate one in a wide range of cases, given the conditions laid out in their meat sharing example. Unfortunately, in our closest relatives, we do not see these conditions. The distribution of benefits is not 'fair' in the hierarchy, and those who punish do so in order to reinforce the 'unfair' benefits that they control. The origins of our behavior seem likely to have evolved in the context of a dominance hierarchy. However, Sober and Wilson's analysis may be correct in the context of current human behavior. This is why there needs to be a coupling of game theory models with what our actual social structure most likely resembled.

Skyrms and Kitcher are correct to suggest that including social dynamics is the next step in formulating game theory models. However, we need to understand the likely social structure before framing our models. An understanding of the social structure will sometimes help us ground the game theory model of certain behaviors, or it could suggest an alternative solution, as in the case of altruistic punishment. I propose that we follow Skyrms and Kitcher's advice. Further examination of coalition behavior and hierarchy maintenance will help shed light on the Darwinian origins of our behavior.

REFERENCES

Boehm, Christopher (1999), *Hierarchy in the Forest: The Evolution of Egalitarian Behavior*. Cambridge, MA: Harvard University Press.

Boyd, Robert, et al. (2003), "The Evolution of Altruistic Punishment", *Proceedings of the National Academy of Sciences* 100: 3531–3535.

Bygott, J. David (1979), "Antagonistic Behavior, Dominance, and Social Structure in Wild Chimpanzees of the Gombe National Park", in David A. Hamburg and Elizabeth R. McCown (eds.), *The Great Apes*. Menlo Park, CA: Benjamin & Cummings, 405–427.

de Quervain, Dominique J.-F., et al. (2004), "The Neural Basis of Altruistic Punishment", *Science* 305: 1254–1258.

de Waal, Frans (1982), *Chimpanzee Politics: Power and Sex among Apes*. New York: Harper & Row.

——— (1989), *Peacemaking among Primates*. Cambridge, MA: Harvard University Press.

Ellickson, Robert C. (1991), *Order without Law: How Neighbors Settle Disputes*. Cambridge, MA: Harvard University Press.

Falk, Dean (2000), *Primate Diversity*. New York: Norton.

Fehr, Ernst, and Urs Fischbacher (2003), "The Nature of Altruism", *Nature* 425: 785–795.

Fehr, Ernst, and Simon Gächter (2002), "Altruistic Punishment in Humans", *Nature* 415: 137–140.

Feinberg, Joel (1994), "The Expressive Function of Punishment", in Antony Duff and David Garland (eds.), *A Reader on Punishment*. Oxford: Oxford University Press, 73–91.

Flack, Jessica C., Frans de Waal, and David C. Krakauer (2005), "Social Structure, Ro-

bustness, and Policing Cost in a Cognitively Sophisticated Species", *American Naturalist* 165: E126–E139.
Goodall, Jane (1986), *The Chimpanzees of Gombe: Patterns of Behavior*. Cambridge, MA: Harvard University Press.
——— (1992), "Unusual Violence in the Overthrow of an Alpha Male Chimpanzee at Gombe", in Toshisada Nishida, William C. McGrew, Peter Marler, Martin Pickford, and Franz de Waal (eds.), *Topics in Primatology*, vol. 1, *Human Origins*. Tokyo: University of Tokyo Press, 131–142.
Kitcher, Philip (1999), "Games Social Animals Play: Commentary on Brian Skyrms's Evolution of the Social Contract", *Philosophy and Phenomenological Research* 59: 221–228.
Skyrms, Brian (1994), *The Evolution of the Social Contract*. Cambridge: Cambridge University Press.
Smuts, Barbara B. (1987), "Sexual Competition and Mate Choice", in Barbara B. Smuts, Dorothy L. Cheney, Robert M. Seyfarth, Richard W. Wrangham, and Thomas T. Struhsaker (eds.), *Primate Societies*. Chicago: University of Chicago Press, 385–399.
Sober, Elliott, and David Sloan Wilson (1999), *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.
Wilson, David Sloan (1979), *The Natural Selection of Populations and Communities*. Menlo Park, CA: Benjamin & Cummings.