Altruistic Punishment: A Sweet Dish
By Jeffrey K. Parrott
Today's Science On File, October 2004

Most people have a strong desire to see social wrongdoers punished, whether the violation is relatively minor, like cheating at a game or cutting in line, or very serious, like robbery or murder. This fundamental human emotion is expressed in every human culture, and frequently finds its way into popular art. It is the feeling that motivates acts of personal revenge (think of the Marvel Comics character The Punisher), well-intentioned vigilantism (think of the DC Comics character Batman) or the operation of police and legal systems (think of the TV show *Law and Order* and its many spin-offs). As University of Nebraska political scientist John Hibbing told *Forbes* magazine, "The notion that a bad guy is going to get it is really important to humans."

Evolutionary biologists define "altruism" as any behavior that is costly to the performing organism while benefiting other organisms. Understood in this narrow sense, punishment of social wrongdoers by humans is virtually always an altruistic behavior. An act of punishment can often be costly to the people who carry out the punishment, and punishment almost never gives any direct material benefit to the punisher. If a punished wrongdoer is successfully deterred from committing a future violation, only strangers (unknown and unrelated to the punisher) are likely to experience any direct benefits. For example, suppose a man cuts in line (a social violation) and is harshly scolded by a woman standing in the line (an act of punishment). If this scolding persuades the man not to take cuts anymore, all the people who stand in a line with him in the future will benefit. But the scolder probably won't be among them--she may never encounter this particular line-cutter again. So the benefits of her act of punishment (the wrongdoer stops cutting in line) go to strangers and not directly to her.

It may seem obvious that punishment benefits society in general. But evolution does not operate on abstract groups like 'species' or 'societies.' Natural selection favors only those traits that directly benefit an individual organism by making it more likely to pass along its genes. Therefore, biologists have long been concerned with the evolution of altruistic behaviors in animals and humans. Altruistic punishment is of special interest to scientists because of its central role in the unique human ability to cooperate with large numbers of genetically unrelated strangers in complex groups. So why do humans feel such a strong emotional desire to punish, an altruistic behavior that gives no direct benefit to the individual?

The answer, as usual, can be found in the human brain. Recently, a team of Swiss scientists addressed the altruistic punishment question using brain imaging technology. Their study, published in the August 27 issue of *Science*, shows that thinking about meting out punishment to a wrongdoer activates a region of the brain called the caudate nucleus (part of the basal ganglia, a bundle of nerve cells located at the very core of the brain). In previous brain imaging experiments, increased activity in this region has been shown to accompany the anticipation of "nonsocial rewards such as monetary rewards

and pleasant tastes," writes Stanford University psychologist Brian Knutson in a commentary that accompanied the study in *Science*. In other words, altruistic punishment is its own reward: people want to punish wrongdoers because it feels good.

For their experiment, the University of Zurich researchers had male participants play an economic game. Participant A and participant B are each given 10 "money units" (which they were later able to exchange for real Swiss Francs). The men interact anonymously and cannot see each other. A is told that he can either keep his 10 units of money, or send it all to B. If A sends his money to B, the experimenters quadruple the amount to 40 units of money, giving B a total of 50. If B is the trustworthy sort, he will return half of the money to A so that both A and B benefit equally from the interaction. But sometimes B decides to "defect," or to betray A's trust by keeping all 50 units of the money. In this case, the experimenters tell A that he can choose to punish B by assigning "punishment points" that subtract from B's money units (but crucially do not add to A's money units).

During the 1 minute that participant A is given to decide whether and how much to punish cheater B, A's brain activity is scanned using Positron Emission Tomography (PET). For PET, a person is injected with weakly radioactive chemicals that are carried around in the bloodstream. Increased blood flow to specific parts of brain anatomy (called "activation" of that region) can then be detected by measuring the radioactive emissions of the blood-borne chemicals. "Increased cerebral blood flow in a certain brain region means more oxygen consumption and more brain activity in this region," explained University of Zurich neuroscientist Dominique de Quervain, one of the study's co-authors, to *National Geographic*.

The experiments showed that participants who were thinking of punishment had significantly increased blood flow to their caudate nucleus, which is part of a larger brain organ called the corpus striatum. (The corpus striatum is one of the several pairs of sub-cortical organs that make up the basal ganglia, which are located at the very center of the brain, underneath the cerebrum and above the spinal cord.) In previous animal and human brain imaging studies, activation of the caudate nucleus has been associated with the processing of pleasurable rewards such as money, food and sex. Drugs like cocaine and nicotine have also been observed to activate the caudate region. In their article, the authors write that "altruistic punishment provides relief or satisfaction to the punisher" and this accounts for the observed activation of "reward-related brain regions" like the caudate nucleus.

The researchers' conclusions were further supported by a modification to the basic experimental design, in which participant A is told that he must pay a unit of money for every punishment point he assigns to defector B. In this case, the researchers found a significant correlation between the degree of caudate activation in participant A when punishment is free, and the amount A is willing to pay when punishment is costly. In other words, the more satisfaction A takes in punishing B, the more he is willing to pay to punish. Ernst Fehr, director of the Institute for Empirical Research in Economics at the University of Zurich and a co-author of the study, told *National Geographic* that "the nice feature of our study is that the variation in the dorsal striatum [the top part of the

corpus striatum, where the caudate nucleus is located] predicts these differences in behavior quite well. Subjects with lower activation in the dorsal striatum punish less."

Appropriately, the Swiss study provides a neural explanation for common metaphors comparing altruistic punishment to food (e.g. 'revenge is sweet' or 'revenge is a dish best served cold'). "People seem to feel rewarded when they punish a defector," Fehr told *Science*. "There is nothing irrational about feeling rewarded about eating a chocolate. Similarly, there is nothing telling us that altruistic punishment is irrational."

However, "as with any compelling study, the findings raise additional questions for future research," writes Knutson in his *Science* commentary. Scientists would like to use different brain imaging techniques, such as functional magnetic resonance imaging (fMRI), "to make even more specific observations regarding when and where activation occurs during altruistic punishment." Because the current study examined only men, further research is needed to determine whether and how women's brains differ with regard to altruistic punishment. Also, evolutionary biologists believe that "altruistic punishment has been crucial to the evolution of cooperation in human societies," as Fehr told *Forbes*. Scientists think that the development of social institutions allowed altruistic punishment behavior to flourish by lowering its cost to individuals within a group. But further research will need to address the evolutionary and genetic origins of the pleasurable emotions associated with punishment.

As satisfying as altruistic punishment may be, and whatever its contribution to complex human societies, Fehr cautions against interpreting these results as a pass to commit acts of vengeance against social violators: "[The study] doesn't imply in any sense that we should allow private revenge. It only means we can explain why people do this."

Further Reading:

"The Neural Basis of Altruistic Punishment." Dominique J.-F. de Quervain, et al. *Science*, August 27, 2004, page 1254.

"Sweet Revenge?" Brian Knutson. *Science*, August 27, 2004, page 1246.

"The Co-evolution of Individual Behaviors and Social Institutions." Samuel Bowles, et al. *Journal of Theoretical Biology*, Volume 223, 2003, page 135.

"The Nature of Human Altruism." Ernst Fehr and Urs Fischbacher. *Nature*, October 23, 2003, page 785.

"The Evolution of Altruistic Punishment." Robert Boyd, et al. *Proceedings of the National Academy of Sciences*, Volume 100, 2003, page 3531.

"Altruistic Punishment in Humans." Ernst Fehr and Simon Gaechter. *Nature*, January 10, 2002, page 137.

Internet Resources:

"Researchers study the pleasure of punishment." (msnbc.msn.com/id/5820379/) A *Science* story, on MSNBC, about the altruistic punishment study.

"Brain Study Shows Why Revenge Is Sweet." (news.nationalgeographic.com/news/2004/08/0827_040827_punishment.html) An article about the Swiss study in *National Geographic*.

"Human brain - Wikipedia, the free encyclopedia." (en.wikipedia.org/wiki/Human_brain) An entry about the human brain--its anatomy, function and study--in the free online encyclopedia, with links and references.

Keywords for electronic searches:

altruistic punishment, evolution of behavior, evolution of cooperation, neuropsychology, basal ganglia, corpus striatum, caudate nucleus, positron emission tomography (PET), functional magnetic resonance imaging (fMRI), brain imaging