Social Justice Research, Vol. 19, No. 2, June 2006 (© 2006)
DOI: 10.1007/s11211-006-0004-x

Some Thoughts on the Adaptive Function of Inequity Aversion: An Alternative to Brosnan's Social Hypothesis

M. Keith Chen[1,3] and Laurie R. Santos[2]

In this commentary, we review and question Brosnan's hypothesis that inequity aversion (IA) evolved as a domain-specific social mechanism. We then outline an alternative, domain-general, account of IA. As opposed to Brosnan's social hypothesis, we propose that IA evolved from more general reward mechanisms. In particular, we argue reference-dependence and loss-aversion can account for the evolution of IA in primates. We discuss recent work on reference-dependence and explore how it may have given rise to inequality-averse behavior in social settings. We conclude with suggestions for future work examining the proximate mechanisms that give rise to IA.

KEY WORDS: inequality aversion; reference-dependence; loss-aversion; primate evolution; capuchins.

Making strong claims about the evolution of a particular trait is always a tricky business. To understand why a specific trait evolved, one needs to recognize how that trait functioned in its ancestral environment-specifically, whether or not it allowed the individual that possessed it to out-reproduce its rivals. Unfortunately, the hands of evolutionary time are rarely kind enough to leave any hard evidence of this sort. Though it is often easy to see how a particular trait benefits an individual *today*, a trait's present-day benefits may have little bearing on how it functioned back in the ancestral day. More often than not theorists are left with little objective historical information

[1]Yale School of Management, 135 Prospect St., Box 208200, New Haven, CT 06520, USA.
[2]Department of Psychology, Yale University, Box 208205, New Haven, CT 06520, USA.
[3]Address correspondence to: M. Keith Chen, Yale School of Management, 135 Prospect St., Box 208200, New Haven, CT 06520, USA, e-mail: keith.chen@yale.edu

about ancestral survival rates, making it difficult to empirically evaluate claims about a trait's evolutionary origins.

This caveat holds especially true in the case of cognitive traits. Unlike physical traits, like a jawbone or a wing structure, cognitive traits leave no fossil remains for researchers to study, nor hard data with which to test claims about a trait's evolutionary past. Furthermore, cognitive traits tend to be more metamorphic than physical traits-they capriciously take on assorted functional roles depending on the environment in which they find themselves. Consider, for example, the cognitive trait of spatial localization depending on what was most needed for reproductive success at the time, it could have served to: locate home, find wayward mates, keep track of offspring, forage for food, or any of a myriad of other tasks. For these reasons, it is especially challenging to determine the specific selective advantage that favored one cognitive trait over another. As such, claims about the evolutionary function of cognitive traits must always be made and evaluated with great care.

With this cautionary note in place, we turn to a discussion of Brosnan's (2006) thoughts about the evolution of our human sense of fairness. Although questions about the phylogenetic origins of fairness have fascinated scholars for centuries, the past few decades have seen a rise in both empirical and theoretical work related to these evolutionary issues (e.g., Axelrod and Hamilton, 1981; Gintis *et al.*, 2004; Sober and Wilson, 1998; Hauser, 2006; Trivers, 1971). In the present volume, Brosnan adds to a growing theoretical foment, nicely articulating one hypothesis about the evolution of one critical aspect of human notions of fairness-our inequity aversion (IA). We begin by briefly reviewing Brosnan's arguments about the evolution of IA, with special emphasis on her assumptions about its function in the ancestral environment, or the reason that IA evolved in the first place. We then suggest an alternative to her view that IA began in the social domain, where it now functions. Instead, we propose a more domain-general mechanism by which IA could have come about, through the evaluation of prospects relative to a moving reference-point. We conclude with some empirical predictions generated by this alternative to Brosnan's hypothesis.

INEQUITY AVERSION: NATURAL SELECTIONS GIFT TO THE SOCIAL MIND?

Brosnan (2006) begins her argument that any response to inequality is unlikely to have evolved *de novo* in the human primate. Instead, it is most likely to have evolved in stages, each of which must have been fitness-increasing. Brosnan then argues (largely structurally and convincingly) that the likely candidates for these steps are:

(1) An organism evolves the ability to recognize the rewards and payoffs of others (and to realize that this can differ from his own payoff.)
(2) An organism evolves the ability to respond to these now-perceived differences.
(3) And, finally, an organism evolves more specific inequality-adverse structures, such as the willingness to sacrifice in order to lower the payoffs of those whose fortunes have been better than ours.

Up to this point, Brosnan is quite convincing; each step seems to necessitate the prior evolution of the previous, and thus we agree with this initial analysis.

What we disagree with, however, is Brosnan's assertion of the *specificity* of these cognitive structures. Brosnan contends that each of these proposed mechanisms served a social function-that is, they were for recognizing and avoiding inequity in the social domain. To take one example, consider the first of Brosnan's proposed stages-recognizing that rewards and payoffs differ across individuals. Here, Brosnan articulates a cognitive structure with a decidedly domain-specific flavor-it is hypothesized to apply *only* in a very specific context: that of rewards and payoffs that occur in the context of other individuals. The same criticism holds true for all the steps leading to IA (which, by its very definition, must have a social basis).

We, however, see little evidence at present to propose that the mechanisms that give rise to IA are specific to the problems of social reasoning *per se*. Instead, we would argue that primates (and probably other animals) naturally develop expectations about rewards in a variety of different contexts and possess mechanisms to detect whenever these expectations are violated relative to that expectation. In support of our view, there is a well-known literature demonstrating that many primates develop expectations about non-social rewards and react negatively when those expectations are violated in a negative way. For example (and cited by Brosnan), Tinklepaugh (1928) showed that individual monkeys reacted negatively when they received a reward that was smaller than the one they had seen hidden inside a food container (see Santos *et al.*, 2002 for a more recent version of this hiding expectancy task). As such, primates seem to react negatively in many situations in which their expectations about rewards are violated, not just those involving inequity across individuals.

AN ALTERNATIVE TO SOCIAL LEARNING: REFERENCE-POINT EVALUATION

For these reasons, we favor the view that human IA falls out of more domain-general capacities for making expectations about and evaluating

rewards relative to some initial reference-point. Note that this reference-point evaluation system has been studied extensively in the human economic literature under the rubric of prospect theory (see Kahneman and Tversky, 1979; Tversky and Kahneman, 1991). In a variety of contexts, humans behave in ways that serve to avoid perceived losses relative to some reference-point. Such lose-averse decisions can be observed across a variety of different problem domains: for example, when stock investors trade (displaying a reluctance to realize losses; Odean, 1998), homeowners sell their homes (unwilling to sell below buying price; Genesove and Mayer, 2001), and shoppers evaluate prices (asymmetrically more sensitive to price increases than decreases in numerous markets; Hardie *et al.*, 1993).

Recently, we (Chen *et al.*, 2006) demonstrate that this reference-dependent behavior is not unique to the human species. Capuchin monkeys (*Cebus apella*)-in many ways the model species for work on the evolution of fairness and IA (e.g., Brosnan and de Waal, 2003, 2004; Fragaszy *et al.*, 2004; de Waal and Davis, 2003)-also exhibit loss-aversion; that is, they often act at the expense of material rewards to minimize their perceived losses. To demonstrate this, we presented capuchin subjects with the opportunity to trade tokens with one of two human experimenters who would deliver different kinds and amounts of food rewards (see Brosnan and de Waal, 2003 for a similar trading task). The capuchins then received a choice between two experimenters that both delivered a food reward of one apple piece yet differed in an important respect-they initially showed different numbers of apple pieces to the monkeys before trading. The first experimenter began the trading task by showing the subject two apple pieces and then, upon being presented a token for a food trade, removed one of these apple pieces and delivered only the remaining piece to the subject.

In contrast, the second experimenter showed only one piece of food and always traded this single apple piece when the subject traded with her. We thus set up a situation in which trading with either experimenter delivered identical payoffs-they each gave a single apple. However, the experimenters differed in terms of the *reference-point* they initially established-the second experimenter set-up an initial offer (and thus initial reference-point) of a single apple piece, while the first experimenter established an initial offer of two pieces. For this reason, the two experimenters differed in terms of how their payoffs were framed; importantly, the first experimenter's payoff was framed as a loss of one apple piece relative to a reference-point of two. When presented with this choice, our capuchin subjects avoided the experimenter who delivered an apple piece that was framed as a loss. Like humans in a variety of contexts, capuchins avoided payoffs that were framed as losses, suggesting that they, too, develop and evaluate expectations about rewards as gains and losses relative to some previously established reference-point.

It is interesting to note that our capuchin reference-dependence study mirrors the IA trading experiments presented by Brosnan and colleagues (e.g., Brosnan and de Waal, 2003, 2004). Brosnan and de Waal (2003), for example, observed that capuchin monkeys become upset when they receive a smaller reward for trading a token than a nearby peer had received earlier. In some sense, this effect is another example of a reference-dependent framing effect-one that is mediated by a socially generated reference-point. The monkey sees what another individual received for his behavior, and this expected payoff becomes his reference-frame. Payoffs that are less than this reference-point are thus perceived as losses (and hence generate frustration). As such, a domain-general mechanism for reference-point setting could easily give rise to the IA behaviors typically observed in social settings. It is plausible, then, that a more domain-general reference-setting mechanism could also have accounted for the *evolution* of IA more generally in our primate lineage.

For these reasons, we contend that reference-point setting mechanisms can account for the IA behaviors observed in living non-human primates. But could these similar mechanisms have operated during our evolutionary past? Put another way, could domain-general reference-setting mechanisms have served to increase survival and reproductive success over evolutionary time as well? Although it is unfeasible to adequately answer these questions for the reasons we noted above, it is easy to see how reference-dependence might have conveyed an evolutionary advantage to our primate ancestors. With respect to a habit-formation type of reference-point, one can easily imagine how deviations from an individual's past payoffs would indicate that something important in the environment has changed in a relevant way. Moreover, paying attention to the payoffs of other individuals living in an organism's environment could potentially be even more useful. Simply stated, attending to the payoffs of others provides many more clues and warnings of environmental changes than simply focusing on your own experiences. This suggests that the ability to attend to the payoffs of others may confer selective advantages *even in the absence of social interactions.*

This point was made most clearly in a recent paper by economists Luis Rayo and Garry Becker (2005). They model a setting in which evolution wishes to "program" an organism with an affective summary of his current action's attractiveness, assuming the organism will seek to maximize its affective payoff. They found that in a broad set of circumstances, the optimal rule for an organism will involve a reference-point (from which happiness is measured), and that this reference-point will vary with both the past and present payoffs of both the self *and* others. Intuitively, the ability to recognize differences in what others earn and what we earn is useful whenever there are common environmental factors that affect the rewards to everyone's actions. In foraging animals, for example, the relative abundance of food sources will

affect not only the payoff an individual animal will obtain for any given effort, but will most likely influence the returns to effort of their peers in a similar fashion. As such, simply attending to the payoffs of one's peers could serve as an incredibly useful tool, a summary statistic of the returns to foraging effort given any set of environmental factors.

Reference-point determined behavior seems a more plausible evolutionary primitive than the social learning mechanisms proposed by Brosnan. Unlike a social learning account, a reference-setting model does not require the ability to acquire new skills, but only the capacity to observe others' payoffs and, from this, determine which behaviors are appropriate in different states of the world. So, for example, a foraging animal who notices that his peers are all consuming a lot of food should optimally (from an evolutionary standpoint) forage more aggressively. In terms of cognitive mechanisms, it might help for this reference-setting forager to also experience a negative envy-like effect-the fact that his neighbors are eating a lot indicates that food is relatively plentiful at present and a little envy-driven extra effort may bring large food rewards. Envy could therefore act to increase the forager's experienced foraging returns only in those situations in which those increased returns are possible. Conversely, envy could serve to reduce futile foraging efforts in bad environments, increasing the animals' foraging efficiency.

Our proposal that social reference-setting could account for the evolution of IA helps to clarify some hitherto puzzling observations concerning the main way that inequality aversion seems to vary-that is, with the stability of the social group. Under a reference-setting account, one could imagine that the payoffs of those who are not part of your stable social group (that is, who move transitorily in and out of your environment) would tell you less about that environment's stable properties; and vice versa for those who consistently live in the same environment you inhabit. Our view, therefore, makes a new prediction about the magnitude of reference-dependence across environments. Specifically, IA should be greatest in uncertain environments, where the optimal strategy is highly variable. This prediction could be tested across primate species by exploring the magnitude to IA as a function of environmental stability.

CONCLUSIONS

It is clear that the experiments of Brosnan (2006) and her co-authors are central to a growing literature on IA in non-human species. The importance of understanding the evolutionary origins of IA can hardly be understated, as demonstrated by the increased attention IA has received from biologists, psychologists, economists, and philosophers. Hopefully, our brief comment

on this latest paper can serve a useful function in this important and expanding dialogue; we offer a cautionary note and alternative hypothesis to be tested by either its formal plausibility or its empirical predictions. Testing Brosnan's hypothesis that IA evolved as a social skill *per se* will greatly shape how we understand its structure, and undoubtedly shed some much-needed light on our evolutionary past.

REFERENCES

Axelrod, R., and Hamilton, W. D. (1981). The evolution of cooperation. *Science*, 211, 1390–1396.

Brosnan, S. F. (2006). Nonhuman species' reactions to inequity and their implications for fairness. *Soc. Justice Res.*, this volume.

Brosnan, S. F., and de Waal, F. B. M. (2003). Monkeys reject unequal pay. *Nature*, 425, 297–299.

Brosnan, S. F., and de Waal, F. B. M. (2004). Socially learned preferences for differentially rewarded tokens in the Brown Capuchin Monkey (*Cebus apella*). *J. Comp. Psychol.*, 118, 133–139.

Chen, M. K., Lakshminarayanan, V., and Santos, L. (2006). How basic are behavioral biases? Evidence from Capuchin-Monkey trading behavior. *Journal of Political Economy*, 114, 517–537.

Fragaszy, D. M., Fedigan, L. M., and Visalberghi, E. (2004). *The Complete Capuchin: The Biology of the Genus Cebus*, New York: Cambridge University Press.

Genesove, D., and Mayer, C. (2001). Loss aversion and seller behavior: Evidence from the housing market. *Quart. J. Econ.*, 116, 1233–1260.

Gintis, H., Bowles, S., Boyd, Robert T., and Fehr, E. (eds.) (2004). *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life*, Cambridge, MA, MIT Press.

Hardie, B. G. S., Johnson, E. J., and Fader, P. S. (1993). Modeling loss aversion and reference-dependence effects on brand choice. *Market. Sci.*, 12, 378–394.

Hauser, M. D. (2006). *Moral Minds: The Unconscious Voice of Right and Wrong*, Harper Collins, New York.

Kahneman, D., and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263–292.

Odean, T. (1998). Are investors reluctant to realize their losses?. *J. Finance*, 5, 1775–1798.

Rayo, L., and Becker, G. (2005). *Evolutionary Efficiency and Happiness*. University of Chicago Department of Economics Working Paper.

Santos, L. R., Sulkowski, G., Spaepen, G. M., and Hauser, M. D (2002). Object individuation using property/kind information in Rhesus Macaques (*Macaca mulatta*). *Cognition*, 83, 241–264.

Sober, E., and Wilson, D. S. (1998). *Unto Others: The Evolution and Psychology of Unselfish Behavior*, Cambridge, MA: Harvard University Press.

Tinklepaugh, O. L. (1928). An experimental study of representative factors in monkeys. *J. Comp. Psychol.*, 8, 197–236.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quart. Rev. Biol.*, 46, 35–57.

Tversky, A., and Kahneman, D. (1991). Loss aversion in riskless choice: a reference-dependent model. *Quart. J. Econ.*, 106, 1039–1061.

de Waal, F. B. M., and Davis, J. M. (2003). Capuchin cognitive ecology: Cooperation based on projected returns. *Neuropsychologia*, 41(2): 221–228.