Communicative & Integrative Biology 3:6, 522-524; November/December 2010; ©2010 Landes Bioscience

# Coevolution of cooperation, causal cognition and mindreading

H. Clark Barrett,[1,*] Leda Cosmides[2] and John Tooby[3]

[1]Center for Behavior, Evolution and Culture; Department of Anthropology; University of California; Los Angeles, CA USA; [2]Center for Evolutionary Psychology; Department of Psychology and [3]Anthropology; University of California; Santa Barbara, CA USA

**The evolution of cooperation between unrelated individuals has long been a puzzle in evolutionary biology. Formal models show that reciprocal altruism is approximately as stable as kin-based altruism when cooperators can assort. Why, then, is reciprocal altruism so rare? We suggest that the key lies in the difficulty of assortment based on underlying intentions: if individuals are able to reliably detect others' cooperative intent then cooperation is stable, but detecting intentions is notoriously difficult, especially when there are incentives to deceive. For this reason, we suggest, there is likely to be a coevolutionary relationship between human cooperativeness and our skills of social causal cognition; it is not a coincidence that we are both extraordinarily social, cooperating with non-kin to a degree not seen in other species and extraordinarily good at inferring others' beliefs, intentions and motivations, a skill sometimes known as mindreading. We discuss results of a recent study that provides evidence for this coevolutionary view of cooperation and social cognition.**

**Key words:** evolution, cooperation, cheater detection, free-riding, mindreading

Submitted: 06/05/10

Accepted: 06/06/10

Previously published online: www.landesbioscience.com/journals/cib/article/12604

DOI: 10.4161/cib.3.6.12604

*Correspondence to: Clark Barrett; Email: hclarkbarrett@gmail.com

Addendum to: Cosmides L, Barrett HC, Tooby J. Adaptive specializations, social exchange, and the evolution of human intelligence. Proc Natl Acad Sci USA 2010; 107:9007–14; PMID: 20445099; DOI: 10.1073/pnas.0914623107.

Why are humans so astonishingly successful at sociality? The evolution of cooperation has been a long-standing puzzle in the study of human evolution and in evolutionary biology more generally, because cooperation suffers from a serious evolutionary stability problem: the fitness incentives to reap the benefits of others' cooperativeness while shirking oneself—also known as "cheating" or "free-riding"—are high. This means that in the absence of mechanisms for preventing cheating, cooperation quickly unravels.[1]

On the other hand, models of the evolution of cooperation via reciprocity (the trading of cooperative favors between unrelated individuals) show that it is about as easy to evolve as cooperation via kin selection, if cooperative individuals are able to assort with like-mindedly cooperative individuals, excluding cheaters from illicitly reaping the benefits of cooperation and destabilizing the system (reviewed in ref. 1–3). And yet, cooperation between unrelated individuals is thought to be far rarer in the biological world than cooperation between kin.[4] This suggests that the "assortment problem" is, in fact, what makes the evolution of cooperation difficult.

From a mathematical perspective, not only are the assortment problems in kin selection and reciprocal altruism identical, they also pose similar informational problems in that both require the inference or detection of a hidden property by prospective cooperative partners: shared genes in the case of kin selection, and shared cooperative intent (or design) in the case of reciprocal altruism. It is here, however, where the resemblance might end. One reason that kin-based cooperation is thought to be so common is that, as Hamilton[5] pointed out, descent from a common ancestor is a highly statistically reliable cue to probability of shared genes, and biologists have discovered several reliable means of inferring shared ancestry, such as direct observation of birth by mothers, experience of parental care by offspring, and coresidence during childhood by siblings.[6] In contrast, the internal states that must be detected for cooperation by assortment to be stabilized—dispositions, motivation structures and

intentions to cooperate—are notoriously difficult to reliably infer.

The informational difficulty of figuring out what another individual intends to do in strategic social contexts, including both cooperation and strategic competition, has long been recognized in the study of the evolution of intelligence.[7-9] Intentions can be both concealed and faked, and when there is an incentive to do so—as in the case of free-riding—there are formidable informational barriers to the evolution of reliable intention detectors. Unlike kinship detection, it is likely that there are no simple magic bullets for detecting the internal states of others that matter for cooperation. Instead, the evolutionary pathway to cooperative intelligence is probably long, involving many steps, each of which incrementally improves organisms' ability to react strategically to the social world via progressively more fine-tuned capacities to detect and represent its hidden causal structure.

Given these considerations, it is perhaps not a coincidence that humans are both hyper-social and that we possess a suite of cognitive capacities that appear well-suited to inferring and representing the hidden causal structure of the social world and of other people as social actors. One such set of capacities, sometimes called "mindreading" or "theory of mind," is widely thought to be an evolved adaptation for inferring the hidden internal states of others, including their knowledge, motivations and intentions.[10] More broadly, we appear to be uniquely good at causal cognition and especially, *social* causal cognition: we devote substantial mental resources to figuring out why other people do things, representing their social relationships, assessing their skills, computing their motivational structures and so on.

We suggest that this is not a coincidence: human cooperativeness and our ability to solve free-rider problems have likely coevolved with our capacities for social causal cognition and theory of mind. While this potential evolutionary synergy has long been recognized in the concept of Machiavellian intelligence, in practice the relationships between cooperation, causal cognition and theory of mind have been little-studied, both theoretically and empirically. Evolutionary game theory models, for example, typically assume an extremely impoverished cognitive interface for "mindreading," involving simple rules that look only at a few prior instances of an individual's past cooperative behavior (as opposed to, e.g., inferring their incentive structures based on individual characteristics or positions in a social network or using other cues to intent or motivation). In psychology, on the other hand, mindreading is typically studied as a skill in isolation, without asking in what socio-ecological contexts it is most likely to have evolved, nor how it is likely to interact by design with other cognitive mechanisms, such as mechanisms that stabilize cooperation (reviewed in refs. 11 and 12).

We recently conducted a study[13] that explicitly investigated the relationship between cognitive mechanisms for stabilizing cooperation—in particular, the cheater detection mechanism[14]—and mechanisms of social causal cognition. Using the Wason selection task, a standard method of experimentally studying cheater detection, we investigated three aspects of social causal cognition that we expected to interact with vigilance for cheaters in cooperative contexts, using a standard scenario in which school volunteers are sorting children into good and bad school districts and have the potential to cheat by giving children illicit access to better schools than their parents have paid for.

The first factor we investigated, which we called "benefit," refers to the ability to compute others' incentive structures in cooperative contexts: their personal incentive to cheat or free ride (i.e., the degree to which they stand to benefit from doing so). We investigated this by activating an evolved psychology of kinship: we predicted that if subjects knew that another individual could benefit his or her own offspring by free-riding, they would be more vigilant for cheating, as measured by turning over exactly those cards necessary to detect cheating in the Wason selection task. This is what we found, with vigilance for cheating increasing by ~20 percentage points when the potential for nepotistic benefits could be inferred.

Second, we measured an aspect of social causal cognition which we called "Ability:" the degree to which individuals are inferred to have access to information that would allow them to manipulate a cooperative situation in their own interest. We varied this factor independently of "benefit," by creating a scenario in which potential cheaters could be thwarted by the use of code numbers to identify children. Removal of ability to cheat reduced vigilance for cheating by ~20 percentage points.

Finally, we investigated a factor directly related to theory of mind: intent to cheat. Again, we varied this factor independently of the other two factors, by explicitly cueing subjects to the possibility of either intentional or accidental rule-breaking. While some accounts of cooperation might suggest that accidental and intentional rule-breaking are equally significant from a fitness point of view because they have equal effects in the world, an "assortment" view of cooperation suggests that what is most important is identifying the underlying motives, in order to sort cheaters from non cheaters. Consistent with this proposal, we found vigilance for cheating to be substantially higher in the intentional than the accidental conditions, again, by about ~20 percentage points. Performance was highest when all three factors were present.

These results provide support for the idea that the evolution of cooperation requires, or is facilitated by, the coevolution of mechanisms that make assortment between like-minded individuals possible. Moreover, they provide support for the proposal that assortment problems are solved not by a single cognitive mechanism, but many. In our study, inferences of intention, inferences of incentive structure and inferences of informational access all had independent effects on the capacity to detect cheating. In fact, we suspect that this is just the tip of the iceberg: a large number of cognitive capacities likely coevolved because of their synergistic effects in enabling human cooperation and sociality. We mean this not just in the narrow sense of tit-for-tat reciprocity, but in all the ways that humans can benefit through cooperative gains in trade, including the evolution of language, with its reliance on relationships between communicative intent and informational validity,[15] and the evolution of other social

learning and cultural transmission mechanisms, especially ones designed for systematic information sharing.[16-18]

We regard this synergistic, coevolutionary approach to the evolution of mental specializations as quite different than approaches that attempt to identify a single factor or capacity as the explanation for human uniqueness, as well as different from approaches to "modularity" or cognitive specialization that assume isolation of mechanisms, rather than interaction between them, to be the hallmark of specialization (reviewed in ref. 19). We suspect that the investigation of evolutionary synergies between cognitive specializations is likely to shed light on aspects of human cognitive evolution that could not be understood via the study of individual capacities alone.

## References


1. Axelrod R, Hamilton WD. The evolution of cooperation. Science 1981; 211:1390-6.
2. Boyd R, Richerson PJ. The evolution of reciprocity in sizeable groups. J Theor Biol 1988; 132:337-56.
3. Joshi NV. Evolution of cooperation by reciprocation within structured demes. J Genetics 1987; 66:69-84.
4. Hammerstein P. Why is reciprocity so rare in social animals? A Protestant appeal. In: Hammerstein P, Eds. Genetic and Cultural Evolution of Cooperation. Cambridge, MA: MIT 2003; 83-94.
5. Hamilton WD. The genetical evolution of social behaviour, I. J Theor Biol 1964; 7:1-16.
6. Lieberman D, Tooby J, Cosmides L. The architecture of human kin detection. Nature 2007; 445:727-31.
7. Humphrey N. The social function of intellect. In: Bateson PPG, Hinde RA, Eds. Growing Points in Ethology. Cambridge UK: Cambridge 1976; 303-21.
8. Krebs JR, Dawkins R. In: Krebs JR, Davies NB, Eds. Behavioural Ecology: An Evolutionary Approach. 2nd Ed. New York, NY: Sinauer 1984; 380-402.
9. Byrne RW, Whiten A. Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes and Humans. New York: Oxford 1988.
10. Baron-Cohen S. Mindblindness. Cambridge, MA: MIT 1995.
11. Ermer E, Guerin SA, Cosmides L, Tooby J, Miller MB. Theory of mind broad and narrow: reasoning about social exchange engages ToM areas, precautionary reasoning does not. Soc Neurosci 2006; 1:196-219.
12. McCabe KA, Smith VL, LePore M. Intentionality detection and "mindreading": Why does game form matter? Proc Natl Acad Sci USA 2000; 97:4404-9.
13. Cosmides L, Barrett HC, Tooby J. Adaptive specializations, social exchange and the evolution of human intelligence. Proc Natl Acad Sci USA 2010; 107:9007-14.
14. Cosmides L. The logic of social exchange. Cognition 1989; 31:187-276.
15. Sperber D, Wilson D. Relevance: Communication and Cognition. 2nd Ed. New York, NY: Blackwell 1995.
16. Boyd R, Richerson PJ. Culture and the Evolutionary Process. Chicago, IL: Chicago 1985.
17. Csibra G, Gergely G. Natural pedagogy. Trends Cog Sci 2009; 13:148-53.
18. Tomasello M, Carpenter M, Call J, Behne T, Moll H. Understanding and sharing intentions: The origins of cultural cognition. Behav Brain Sci 2005; 28:675-91.
19. Fodor J. The Modularity of Mind. Cambridge, MA: MIT 1983.