Journal of Personality and Social Psychology
2012, Vol. 102, No. 6, 1252–1270

0022-3514/12/$12.00 DOI: 10.1037/a0027026

The Psychosemantics of Free Riding: Dissecting the Architecture of a Moral Concept

Andrew W. Delton, Leda Cosmides, Marvin Guemo, Theresa E. Robertson, and John Tooby
University of California, Santa Barbara

For collective action to evolve and be maintained by selection, the mind must be equipped with mechanisms designed to identify free riders—individuals who do not contribute to a collective project but still benefit from it. Once identified, free riders must be either punished or excluded from future collective actions. But what criteria does the mind use to categorize someone as a free rider? An evolutionary analysis suggests that failure to contribute is not sufficient. Failure to contribute can occur by intention or accident, but the adaptive threat is posed by those who are motivated to benefit themselves at the expense of cooperators. In 6 experiments, we show that only individuals with exploitive intentions were categorized as free riders, even when holding their actual level of contribution constant (Studies 1 and 2). In contrast to an evolutionary model, rational choice and reinforcement theory suggest that different contribution levels (leading to different payoffs for their cooperative partners) should be key. When intentions were held constant, however, differences in contribution level were not used to categorize individuals as free riders, although some categorization occurred along a competence dimension (Study 3). Free rider categorization was not due to general tendencies to categorize (Study 4) or to mechanisms that track a broader class of intentional moral violations (Studies 5A and 5B). The results reveal the operation of an evolved concept with features tailored for solving the collective action problems faced by ancestral hunter-gatherers.

Keywords: evolutionary psychology, cooperation, free riding, concepts, moral psychology

Supplemental materials: http://dx.doi.org/10.1037/a0027026.supp

From prehistory to the present, human survival has depended on productive labor, much of which was carried out by groups of people coordinating their actions to reach a common goal and then sharing the resulting benefits. This style of cooperation—often called collective action—is seen across human societies. Indeed, ancestral humans cooperatively hunted big game on African, Asian, and American grasslands; they built communal fish traps, attacked mammoths, and drove bison over cliffs. In the modern world, Amish farmers have community barn raisings, Amazonians cooperatively clear jungle for crops, and soccer fans cheer as their favorite teams use collective action to compete for points and glory. Military units train to provide group defense, teams of factory workers assemble cars, scientists collaborate on research projects, and international volunteers create Wikipedia. Human societies, if they existed at all, would be unrecognizably different if there were no collective action.

Yet this form of cooperation is rare in other species. The major exception is the social insects, such as ants, bees, and termites, in which its evolution and stability are explained by close genetic relatedness (Hamilton, 1964; E. O. Wilson, 1974). Although exploiting opportunities for cooperation would be widely advantageous, very few species actually exhibit collective action among individuals who are not close genetic relatives. Aside from humans, chimpanzees are the most uncontroversial and well-studied example, although the circumstances in which they display this behavior are far more limited than in humans (Boesch, 2002; M. L. Wilson & Wrangham, 2003). The fact that humans routinely engage in collective action raises the hypothesis that our psychological architecture contains evolved specializations that allow us to solve the formidable problems that prevent its evolution in other species (Tooby, Cosmides, & Price, 2006). Here, we report empirical investigations of one proposed component of this zoologically unusual architecture: a mechanism for identifying free riders.

The presence of individuals with a disposition to free ride—that is, to take the benefits of group cooperation without contributing to the cooperative project—can jeopardize the evolution of collective action. Evolutionary game theory shows that adaptations for contributing to collective actions that are open to free riding cannot evolve and be stably maintained in a species unless cooperators can find a way to exclude free riders, create incentives for them to become contributors, or otherwise reduce their welfare so that it is

This article was published Online First January 23, 2012.

Andrew W. Delton, Leda Cosmides, Marvin Guemo, and Theresa E. Robertson, Center for Evolutionary Psychology, Department of Psychological and Brain Sciences, University of California, Santa Barbara; John Tooby, Center for Evolutionary Psychology, Department of Anthropology, University of California, Santa Barbara.

We thank Toshio Yamagishi and the members of the Center for Evolutionary Psychology for constructive feedback on earlier versions of this article. This work was supported by a National Institutes of Health Director's Pioneer Award to Leda Cosmides and National Science Foundation Grant 0951597 to Leda Cosmides and John Tooby.

Correspondence concerning this article should be addressed to Andrew W. Delton, Department of Psychological and Brain Sciences, University of California, Santa Barbara, CA 93106-9660. E-mail: andy.delton@gmail.com

lower than cooperators (e.g., Boyd & Richerson, 1988, 1992; Hauert, De Monte, Hofbauer, & Sigmund, 2002a, 2002b; Panchanathan & Boyd, 2004). Consistent with these game theoretic analyses, empirical studies of human cooperation show that free riding often elicits anger from contributors, many of whom respond by punishing the free rider or by down-regulating their own contributions to the group effort (e.g., Fehr & Gachter, 2000; Kerr et al., 2009; Kiyonari & Barclay, 2008; Kurzban, McCabe, Smith, & Wilson, 2001; Masclet, Noussair, Tucker, & Villeval, 2003; Price, 2005, 2006b; Price, Cosmides, & Tooby, 2002; Taggar & Neubert, 2008; Yamagishi, 1986). Note, however, that these responses to free riders entail a logically prior question: What counts as free riding?

In this article, we put the free rider concept under the microscope to see what criteria the mind uses to categorize someone as a free rider. In dissecting the architecture of this moral concept, we employ a standard social psychological method in a novel way. For over 30 years, the memory confusion protocol has been used to study when people categorize others by their sex, race, or other characteristics, where the experimenter has decided which individuals belong to which category (Taylor, Fiske, Etcoff, & Ruderman, 1978; see also Klauer & Wegener, 1998). This protocol uses patterns in recall errors to assess social categorization. As in previous research, we used this method to determine whether free riders are seen as a category at all. However, we also used it to determine which information the mind uses to assign an individual to the category FREE RIDER (see Sherman, Castelli, & Hamilton, 2002, for a similar extension of this method). Are people categorized as free riders based only on their objective contributions to the collective action? Do their intentions matter? If we find that free riders are categorized, does the category have content specific to free riding, or does the mind fold free riders into a single category, such as moral violators? By parametrically varying the input subjects receive, we tested competing hypotheses against one another. Our goal was to uncover the semantics of the concept FREE RIDER. In so doing, we take a step toward addressing the question, What is the "stuff of thought" that makes collective action possible? (On conceptual semantics, see, e.g., Barrett, 2005; Jackendoff, 2006; Pinker, 2007; Talmy, 2000; Tooby, Cosmides, & Barrett, 2005; consistent with common practice in this literature, we use small caps to highlight a proposed element of conceptual structure.)

The Puzzle of Collective Action

The fact that people cooperate in groups does not seem puzzling at first. After all, collective action provides large benefits not otherwise available. However, the incentive structure faced by individual participants constitutes a major barrier to mobilizing collective action (Olson, 1965). The joint efforts of a collective action often (though not always) produce public or common goods[1]: resources whose use cannot be limited to those who have contributed to producing them (Dawes, 1980; Olson, 1965; Ostrom, 1990; Van Lange, Liebrand, Messick, & Wilke, 1992). Incentives to free ride exist when those who pay the cost of contributing get the same reward as those who contribute nothing. This creates a social dilemma: Everyone would be better off if the good were produced, but each individual is made better off if they free ride rather than expend effort to produce it. Defense is a classic example of a public good: Any given citizen is just as safe if she avoids paying taxes, but if everyone free rides, then there is nothing to stop aggressive invasion. This logic applies generally to any cooperative activity in which rewards are decoupled from effort. Because this incentive structure seems to preclude many actual collective actions, the existence of collective action has been an enduring puzzle in political science and economics.

One might doubt the kind of rational calculations and short-term self-interest presupposed by an economic analysis of incentives under rational choice. But evolutionary biologists find that analogous problems arise on evolutionary timescales when payoffs are in the currency of fitness. Evolutionary analyses require no assumptions about rationality; indeed, they can be usefully applied to cooperation among bacteria (Axelrod & Hamilton, 1981). This approach models interactions among agents endowed with well-defined decision rules that produce situationally contingent behavior. These decision rules are sometimes called *strategies* by evolutionary biologists—but no conscious deliberation by bacteria (or humans) is implied (or ruled out) by this term. In evolutionary game theory, a decision rule or strategy that garners higher payoffs (i.e., food or other resources) leaves more "offspring"—more copies of itself—in the next generation than alternatives that garner lower payoffs (Maynard Smith, 1982). By analyzing the relative payoffs (and therefore reproductive consequences) of alternative decision rules, evolutionary biologists can determine which strategies natural selection is likely to favor and which are likely to be selected out.

Using this framework, imagine a population of agents who have opportunities to contribute to collective actions. Each agent is equipped with one of two possible decision rules. One causes unconditional cooperation: Agents with this rule always contribute to collective actions. The other causes its agent to never contribute: These agents always free ride. The benefits of collective action will be reaped by both types of agent, but the costs of contributing to it will be borne only by the unconditional cooperators. This means that free riders will experience higher net payoffs than unconditional cooperators. Because agents with higher net payoffs produce more offspring, the free rider design will leave more copies of itself in the next generation than the design that always contributes. As this reproductive advantage continues over generations, the population will be increasingly composed of agents endowed with the *always-free-ride* design. Collective action will eventually disappear as the proportion of cooperators becomes vanishingly small.

How then can psychological adaptations for participating in collective action evolve? To evolve and be stably maintained by selection, designs that cause cooperation need to accrue a higher average payoff than designs that cause free riding. When there are repeated interactions, strategies that cooperate conditionally can outperform exploitive strategies by channeling their cooperative efforts toward other cooperators and away from free riders. To

[1] Both public and common goods share the property of nonexcludability (i.e., noncontributors can also benefit from them). However, common goods are rivalrous (i.e., one person's consumption reduces others' potential consumption, such as parking spaces on public streets), whereas public goods are nonrivalrous (e.g., use of an idea). For our analysis, only the feature of nonexcludability is relevant.

accomplish this, contributors need to exclude free riders from cooperative interactions, create incentives for them to contribute, or otherwise cause their welfare to be less than cooperators'. This can be achieved by avoiding free riders, refusing to cooperate when they are present, or punishing them when they fail to contribute (Boyd & Richerson, 1992; Hauert et al., 2002a, 2002b; Panchanathan & Boyd, 2004; Price et al., 2002; Tooby & Cosmides, 1988; Tooby et al., 2006). For the brain to implement a strategy for conditional cooperation, it requires a mechanism that identifies free riders and distinguishes them from conditional cooperators.

When we say that conditional cooperators must identify free riders, we mean that designs that cause conditional cooperation will not be selected for and maintained in a population unless they identify those with a disposition to free ride. By disposition to free ride, we mean those with a greater tendency to free ride than others, whether because of ontogenetic calibration, heritable genetic variation, or the nature of the current situation (Tooby & Cosmides, 1990). With this consideration in mind, an evolutionary game theoretic analysis predicts that free riders must be identified as those whose mechanisms are calibrated to take benefits without contributing.

Who Should Count as a Free Rider? Evolutionary Constraints on the Concept

Strategies for conditional cooperation are designed to cooperate with other cooperators and withdraw from or punish free riders. To do so, these strategies require a system that categorizes individuals as free riders versus cooperators based on cues and criteria. Which cues and criteria it evolves to use should be determined, in part, by the costs and benefits of hits, misses, false alarms, and correct rejections in repeated interactions (Delton, Krasnow, Cosmides, & Tooby, 2011; Haselton & Nettle, 2006; Kiyonari, Tanida, & Yamagishi, 2000; Yamagishi, Terai, Kiyonari, Mifune, & Kanazawa, 2007).

Hits—correctly identifying free riders—trigger the withdrawal of cooperation, avoidance, or punishment of free riders; their effect is to limit the extent to which free riders benefit at the expense of conditional cooperators. When the categorization system generates a miss—mistaking a free rider for a cooperator—the cooperator contributes, thereby suffering an initial loss that benefits the free rider. This error is likely to be corrected in later collective actions, however, when the free rider continues to undercontribute.

A free rider concept that produces many hits and few misses will limit the losses that conditional cooperators suffer when interacting with free riders. But limiting these losses is not sufficient to give cooperators a selective advantage over free riders. To outreproduce designs that free ride, cooperators need to reap the benefits of repeated mutual cooperation. When the free rider detection system produces a correct rejection—that is, when it correctly identifies its partner as a conditional cooperator—this triggers cycles of reciprocated cooperation and, therefore, strings of benefits.

The benefits of repeated mutual cooperation fail to materialize, however, when there are false alarms—that is, when a conditional cooperator is misclassified as a free rider. False alarms trigger cycles of mutual noncooperation between conditional cooperators. If Jack incorrectly categorizes Jill as a free rider, he will punish or withdraw cooperation from Jill. Jill is likely to respond by withdrawing cooperation from Jack. Because of this initial false alarm, Jack and Jill thereby miss out on a string of benefits that each could have harvested by cooperating in collective actions with the other. Indeed, agent-based simulations show that this error is such a costly mistake that selection creates agents biased to avoid it even at the price of tolerating many misses (Delton, Krasnow, et al., 2011).

Given the differential costs of these errors, what criteria should a conditional cooperator use to identify another individual as a free rider in collective actions, to be punished or excluded? One possibility is a classification rule based on a simple behavioral cue: "Categorize everyone who fails to contribute as a free rider." This rule will catch most free riders by generating a high proportion of hits and virtually no misses, but it will do so at the price of generating false alarms: Conditional cooperators will be misclassified as free riders. This is because events in which an individual fails to contribute to a collective action can occur without revealing the presence of an individual calibrated to free ride.

At some point, virtually all conditional cooperators encounter a situation in which they will not or cannot contribute to a collective action. A system designed to minimize false alarms should use these situations as exclusion criteria: When they are present, an individual who undercontributes to a collective action should not be classified as a free rider. Conditional cooperators can fail to contribute to collective actions because they are in the presence of free riders (Panchanathan & Boyd, 2003), by mistake (e.g., lapses in memory or attention; Axelrod, 1984), or—importantly—due to bad luck (accidents, injuries, illness, or failure to procure the necessary resources despite trying). In the ecological conditions faced by our hunter-gatherer ancestors, rates of injury and disease are so high that virtually everyone in a band is unable to contribute to collective efforts for substantial periods of time (weeks or months; Sugiyama, 2004). High variance in foraging success is also typical even for individuals who are expending a great deal of effort. Among the Ache of Paraguay, for example, men return empty handed on four out of 10 hunts (Kaplan, Hill, & Hurtado, 1990). When there is high variance in individual foraging success due to luck, rather than effort, hunter-gatherers typically buffer this risk by turning the provisioning of that resource into a collective action (Cashdan, 1992; Cosmides & Tooby, 1992; Gurven, 2004; Kameda, Takezawa, & Hastie, 2003; Kameda, Takezawa, Tindale, & Smith, 2002; Kaplan & Hill, 1985; Kaplan, Hill, Lancaster, & Hurtado, 2000). When reversals in fortune are frequent, temporary,[2] and easy to detect, this can be a winning strategy: Everyone in the risk pool is able to contribute on some occasions, but not others.

Punishing conditional cooperators who are temporarily unable to contribute due to illness, injury, accidents, and other forms of bad luck will not increase their level of contribution, making this a waste of energy. Withdrawing cooperation from them may be less costly than punishment in the short run, but it will trigger cycles of mutual withdrawal, leaving one with an ever-shrinking pool of cooperative partners. Everyone eventually experiences misfortune.

[2] The selection pressures are different if the disability is permanent (Kurzban & Leary, 2001).

In sum, the evolutionary function of a free rider concept is to defend conditional cooperators against exploitation without creating too many false alarms. Doing so requires a system that represents and tracks another person's behavior so that it can (when warranted) correctly connect an attributed disposition (to free ride) with that particular person (who thereby becomes categorized as a FREE RIDER). Events in which an individual fails to contribute to a collective action may be sufficient to activate a free rider detection system, but that system should resist tagging that person as a free rider when situational cues provide evidence that this failure was caused by bad luck, innocent mistakes, or because the focal individual was responding to the presence of others calibrated to free ride.

Possible Computational Theories of Free Rider Categorization

We consider four rules that could be used to categorize individuals as free riders on a collective action. One—the free rider strategy rule—satisfies the evolutionary constraints outlined above. The other three do not; they represent alternative hypotheses suggested by rational choice theory, reinforcement learning, moral psychology, attribution theory, and domain-general approaches to categorization.

The Free Rider Strategy Rule

The evolutionary analysis calls for a free rider categorization rule that is activated by events in which a person undercontributes but uses specific behavioral and situational cues to distinguish exploiters (hits) from conditional cooperators (false alarms).

When an exploitive design is activated in individuals, their undercontributions will be motivated by the payoffs of exploitation: They may consume resources instead of contributing them, or they may decide to not to expend the effort needed to achieve the cooperative goal (despite being able to do so). In folk terms, there will be evidence of an intent to undercontribute, motivated by a goal to benefit from the uncontributed resource or by the goal of avoiding the cost of making the expected contribution.

When individuals calibrated for conditional cooperation undercontribute, these payoff clues will usually be absent, and others will be present instead. They will try to procure the resource (but fail due to accidents or bad luck), suffering an energetic expense, or they will not try because they are suffering from an incapacitating illness or injury. Conditional cooperators may intentionally withhold contributions when too many free riders are present. But they will not benefit themselves by consuming a resource they had agreed to provide when other contributors are present. These considerations suggest the following categorization rule:

> If a member of a collective action intentionally fails to contribute even in the presence of other contributors, then categorize the member as a free rider. (free rider strategy rule)

Here, *intentional* is shorthand for actions organized to meet one of two ends: benefiting from the uncontributed resource or avoiding the costs of producing it.

Implementing the free rider strategy rule requires a categorization system that uses inferences produced by the theory of mind system (Baron-Cohen, 1995; Callaghan et al., 2005; Harris, 1990; Leslie, 1994; Malle & Knobe, 1997). This network operates from infancy onward (Onishi & Baillargeon, 2005; Surian, Caldi, & Sperber, 2007); it includes features designed to infer what goals are implied by an individual's actions and whether the outcome produced occurred by accident or by the systematic operation of decision rules.

The free rider strategy rule is broadly consistent with past research in social psychology on the role of intentions in cooperation. For instance, studies have shown that decisions to contribute in social dilemmas are regulated in part by expectations of whether others intend to contribute as well (e.g., Dawes, Mctavish, & Shaklee, 1977; Messe & Sivacek, 1979). Moreover, when the environment is noisy—and hence partially obscures true intent—decision rules that are better at revealing cooperative intentions are more successful in sustaining mutual cooperation (Klapwijk & Van Lange, 2009; Tazelaar, Van Lange, & Ouwerkerk, 2004).

The hypothesis that the mind embodies this free rider strategy rule extends prior work in several ways. First, it goes beyond an interpretation of behavior as volitional versus accidental. It specifies what payoffs are used as criteria for inferring whether the goal of a volitional action was exploitive or cooperative and places situational constraints on which intentional undercontributions are relevant (e.g., responses to free riders do not count). Second, it allows us to discriminate between different kinds of intentional moral violations in a fine-grained and principled way: Tests reported herein reveal a specialized FREE RIDER concept that dissociates from other moral violations. Third, it explains why these particular intentions, goals, payoffs, and situational constraints matter, addressing an ultimate (evolutionary) level of causation: Discriminating free rider designs from other sources of undercontribution is necessary for conditional cooperation in collective actions to evolve and be maintained by selection.

The Return Rate Rules

The free rider strategy rule is based on an evolutionary game theoretic approach, but other approaches suggest alternative rules for defining who counts as a free rider. For instance, reinforcement learning and rational choice theory,[3] with their focus on the proximate rewards or payoffs experienced by the organism, would predict that what matters to the categorizer is the bottom line: how much is contributed by a member of a collective action to the group. If free rider categorization is based on how much labor or resources are actually contributed by each person—by their return rates—then who counts as a free rider should be determined by a rule more like the following:

> If a member of a collective action fails to contribute, then categorize the member as a free rider. (return rate rule)

One could also envision a variant of this rule that compares relative return rates:

> If a member of a collective action contributes at a rate less than other members of the collective action, then categorize the member as a free rider. (relative return rate rule)

[3] Here, we are referring specifically to rational choice theory with a short-term, selfish, profit-maximizing utility function. Although there are an infinite number of possible utility functions, this is the one most commonly assumed.

These rules are essentially restatements of one component of a recent social neuroscience theory of collective action (Seymour, Singer, & Dolan, 2007). This theory attempts to explain the punishment of free riders and the maintenance of collective action in terms of reinforcement and observational learning. A variant of the relative return rate rule was explicitly tested by Masclet and colleagues (2003). Although their experiment was not designed as a critical test between the free rider strategy rule and the relative return rate rule, they showed that high contributors punished as a function of (a) the degree to which others contributed less than the high contributor did and (b) the degree to which others contributed less than the group average.

The return rate rules do not satisfy the evolutionary constraints discussed above. Because people with an activated free rider strategy undercontribute, these rules would generate many hits, but at a steep price in false alarms. In the high-variance environments of ancestral humans, misfortune sometimes prevented conditional cooperators from contributing to collective actions. Return rate rules would misclassify these individuals as free riders because they assess delivered contributions alone, without considering illness, injury, accidents, effort expended, or other exclusion criteria.

The Moral Violator Rule

According to the moral violator rule, there is nothing special about free riding; it is just one moral violation out of many. In this view, the mind contains a computational system for detecting moral violations, which sorts all moral violators into the same mental category.

Like the free rider strategy rule, the moral violator rule would categorize people based, in part, on their intentions—psychologists at least since Piaget have recognized that intentionality plays an important role in generating moral intuitions. For example, intentionality judgments figure prominently in Mikhail's (2007) computational theory of universal moral grammar (see also, e.g., Cushman, 2008), attribution theorists have shown that intentions play a major role in attributions of morality (Weiner, 1993; Weiner, Perry, & Magnusson, 1988), and morality plays a central role in impression formation (Cuddy, Fiske, & Glick, 2008; De Bruin & Van Lange, 1999; Skowronski & Carlston, 1987; Wojciszke, Brycz, & Borkenau, 1993).

Intentionally violating an agreement is usually considered a moral violation, whether or not the agreement is to contribute to a collective action. So, is there any reason to think that there is a special category of FREE RIDER with its attendant categorization rule? Maybe all we need is the following:

> If a person commits an intentional moral violation, then categorize that person as immoral. (moral violator rule)

Given the strong intuition that there is a general category of morally relevant situations and actions and that there are moral versus immoral people, it is not unreasonable to hypothesize that the mind has conceptual primitives that capture the moral–immoral distinction and rules for categorizing people using these primitives. But is this enough?

Because they focus on ancestrally recurrent content domains, evolutionary perspectives on social psychology can provide guidance about what kinds of content—including moral content—the mind should distinguish and privilege in its processing (Bugental, 2000; Kenrick, Li, & Butner, 2003; Kenrick, Maner, & Li, 2005). For instance, given that different types of moral violations require different downstream responses, an evolutionary perspective strongly predicts that the mind should have multiple, domain-specific moral concepts. Consider, for example, what count as adaptive responses to the moral violations of free riding and sexual infidelity. A person's sexual infidelity may be grounds to avoid choosing them as a romantic partner, yet that person may be a good cooperative partner for a collective action. Likewise, it might be wise to avoid choosing a person who free rides on group efforts as a partner for a collective action, yet that person might be a good mate—indeed, in some cases, they may be desirable, as long as they do not free ride on you (Price, 2006a).

The moral violator rule creates a category too coarse to support adaptive behavior. Punishing or excluding everyone this rule classifies as immoral would prevent one from reaping the benefits of collective action with good conditional cooperators who have committed moral violations unrelated to free riding.

The Arbitrary Categorization Rule

Last, we consider a very general, content-free possibility. Many theorists believe that humans have an extremely powerful and domain-general categorization system and/or statistical inference system. Such systems will spontaneously categorize along any distinction they can discern. This implies the following rule:

> If two sets differ along any dimension, then categorize them as separate types. (arbitrary categorization rule)

Indeed, the research that introduced the memory confusion protocol assumed as a working hypothesis that social categorization was accomplished by mechanisms that also operated over nonsocial stimuli (Taylor et al., 1978). There is still no general consensus on whether social categorization is accomplished by general categorization mechanisms that are also used for nonsocial stimuli (Schneider, 2004, p. 79) or by a series of mechanisms specialized for categorizing the social world. Thus, the arbitrary categorization rule is a viable alternative. If it is experimentally supported, then finding that people categorize free riders may not tell us anything specific about the concept of free riding. For this reason, experiments eliminating this rule would strengthen the case for a concept based on the free rider strategy rule.

The Present Research

On the basis of an evolutionary analysis, we propose that the human mind contains specialized computational mechanisms for engaging in collective action. These should include a procedure designed to identify free riders that uses the free rider strategy rule. This rule categorizes as free riders those individuals who intentionally fail to contribute, even in the presence of other contributors. This contrasts with other reasonable categorization rules: the return rate rules, the moral violator rule, and the arbitrary categorization rule.

To test these rules against one another, the methods we used integrated the realism of narratives and the precision of experimental games. Specifically, we presented subjects with a vivid scenario of a group working together for survival. They read about a group of people who were stranded together on an island, some of whom were injured. The uninjured individuals, eight men, were described as agreeing to work together to find food and to bring whatever they found back to camp to share with the whole group. After this, subjects viewed a series of captioned photos describing what each man did on five foraging days. Using these descriptions, we precisely equated or manipulated the contributions and intentions of the foragers. This allowed us to test which behavioral and situational cues elicited the formation of distinct categories and which did not. To measure categorization, we used the memory confusion protocol created by Taylor and colleagues (1978).

In addition to this measure of categorization, we also employed a set of free rider criterion measures; these allowed us to test whether the categories formed by subjects reflected the criteria of the free rider strategy rule, as opposed to a category with some other content. On the basis of various proposals for how evolution has shaped adaptations for collective action, free riders can be expected to reliably elicit a cluster of responses from cooperators and be associated with a cluster of personality traits, as follows. Criterion 1: There should be a motivation to punish free riders (Boyd & Richerson, 1992; Gintis, 2000; Henrich & Boyd, 2001; Price et al., 2002). Empirical work has shown that many individuals are willing to punish free riders and that punishment sustains cooperation in collective actions (e.g., Fehr & Gachter, 2000; Masclet et al., 2003; Miles & Greenberg, 1993; Yamagishi, 1986). Criterion 2: There may be motivation to reward contributors (Kiyonari & Barclay, 2008). Criterion 3: Free riders should be viewed as untrustworthy, selfish, and unlikable when evaluated in the context of a collective action, and these traits should be seen as dispositional rather than situational (Alexander, 1987; Dunbar, 2004; Panchanathan & Boyd, 2004). Criterion 4: People should be reluctant to choose free riders as fellow participants in future collective actions (Alexander, 1987; Kurzban & Leary, 2001). (Although the predictions for other types of interactions are less clear, we also tested whether individuals do not want to interact with free riders in dyadic settings.) Criteria 3 and 4 are based on theories of selective partner choice, known variously as assortative interaction, direct reciprocity, indirect reciprocity, and gossip-based strategies (Alexander, 1987; Boyd & Richerson, 1988; Dunbar, 2004; Nowak & Sigmund, 2005; Panchanathan & Boyd, 2004; Price, 2006b).

Taken in isolation, each of these four responses could be generated in response to individuals other than free riders. However, we are looking for a pattern, the signature of a free rider. At this point in theory development, it is unclear which set of measurable responses would be perfectly unique to free riders. Nevertheless, individuals who are categorized as free riders should be easily distinguishable from individuals categorized as, for example, incompetent. The latter category should elicit a very different cluster of responses, which does not include punitive sentiments (Cuddy et al., 2008; Neuberg, Smith, & Asher, 2000).

All experiments used essentially the same procedure, which is described in Study 1. Table 1 summarizes the manipulations used in each study, the categorization rule being tested, the prediction that follows from each rule, and a short conclusion drawn from each study. Studies 1 and 2 tested for the operation of the free rider strategy rule by comparing intentional and accidental undercontributors, while holding constant the actual amount each contributes. Study 3 tested the alternative hypothesis that the mind con-

Table 1
Methodological and Theoretical Overview of the Categorization Studies

Study	Rule tested	Manipulation	Contrasting categories		Prediction	Prediction confirmed?	Conclusion
1	Free rider strategy rule	Intentionality of undercontribution	Found food, ate it	Found food, accidentally lost it	Categorization as FREE RIDERS and COOPERATORS	Yes	Free rider strategy rule confirmed
2	Free rider strategy rule	Intentionality of undercontribution	Did not look for food	Tried but never found food	Categorization as FREE RIDERS and COOPERATORS	Yes	Free rider strategy rule confirmed
3	Return rate rules for free rider categorization	Amount of contribution	Found food, accidentally lost it	Always contributed, lost personal items	Categorization only along a dimension of competence	Yes	Return rate rules falsified
4	Arbitrary categorization rule	Reason for unintentional undercontribution	Found food, accidentally lost it	Tried but never found food	No categorization	Yes	Arbitrary categorization rule falsified
5A/5B	Moral violator rule	Type of immoral action	Found food, ate it	Unprovoked battery (5A)/theft (5B)	Categorization as FREE RIDERS and other immoral type	Yes	Moral violator rule may exist, but cannot explain all results

Note. The prediction column does not give the prediction from each rule. Instead, it gives the prediction from the larger theoretical perspective adopted here.

tains one of the return rate rules by comparing individuals who contribute less versus more to the group, while holding constant their intentions to contribute. Study 4 tested another alternative hypothesis—that the mind contains an arbitrary categorization rule—by comparing individuals who differ along a dimension that is semantically identifiable but not relevant on an evolutionary account. Studies 5A and 5B tested whether the moral violator rule is sufficient to account for the results of Studies 1–4, by examining whether free riders are categorized separately from two other types of moral violator.

Study 1: Keeping the Resource Instead of Contributing

Study 1 tested for the operation of the free rider strategy rule. Each forager fails to contribute on two of five days, but for different reasons: Some eat the food they find, whereas others lose it by accident. This equates return rates but varies cues to exploitive intent. The return rate rules predict that these targets will not be sorted into two distinct categories based on these intent cues. The free rider strategy rule predicts that they will—and that members of these two categories will be represented as free riders versus cooperators, as revealed by how subjects rate their impressions of and responses to each target.

Method

Subjects. Seventy-four undergraduates (37 female) enrolled in introductory psychology and introductory physical anthropology classes participated in exchange for partial course credit. Each worked independently at semiprivate computer workstations.

Procedure. Subjects learned about a group of men stranded on a deserted island after their plane's engines were damaged by a storm. Some were severely injured. The uninjured ones agreed to search for food to bring back and share with the group, including those too injured to forage. The foragers searched individually. Each forager was represented by a photograph along with a caption describing his actions and outcomes as he searched for food. Subjects were asked to form impressions of these men (the targets).

Subjects first completed a learning phase in which they learned what happened on five days of foraging. For each day, the activities of the same eight men were shown, one man at a time, each for 10 seconds. Forty unique sentences were used as captions; these were chosen by the computer in random order without replacement (within the constraints discussed below). To emphasize their interdependence, a short vignette about the targets working as a group (e.g., making shelter) appeared after each day; these did not refer to the targets individually.

Each target was paired with two diagnostic sentences—ones that implemented the experimental manipulation—and three nondiagnostic sentences (ones depicting uneventful—but successful—cooperation). On the first day of foraging, every target was paired with a nondiagnostic sentence. Thereafter, the computer randomly determined whether a target was paired on a given day with a nondiagnostic sentence or a diagnostic one, with the constraint that no target received a diagnostic sentence two days in a row. After this learning phase, there was a brief filler task (to eliminate short-term memory effects), followed by a surprise memory task. With all of the eight photos displayed on the screen, each of the 40 sentences appeared one at a time (in random order); for each, the subject was asked to click on the person "who did this." Finally, subjects completed a series of items assessing their reactions to each target.

Materials.

Photographs. Eight facial photographs of young white men were used as stimuli. On the basis of preratings by a separate group of 16 subjects, these faces were chosen from a larger pool so as to be approximately matched on apparent trustworthiness and competence. To further avoid potential confounds, for each subject the computer randomly selected which men would appear as intentional versus accidental undercontributors.

Sentences for the memory confusion protocol. Every target was paired with nondiagnostic sentences on three foraging days; these depicted the target as having found food that he is bringing back to share with the group (e.g., "He watched a flock of birds fly overhead and then took the coconuts he'd found back to camp"). Pairings of photos and nondiagnostic sentences were randomized for each subject, eliminating any potential confounds between person-types and nondiagnostic sentence content.

On the other two foraging days, targets were paired with diagnostic sentences. In Study 1, four targets were paired with diagnostic sentences that depicted intentional failures to contribute (e.g., "He looked around to make sure no one was watching and ate the lobster he had caught"). The other four targets were paired with diagnostic sentences that depicted accidental failures to contribute (e.g., "He slipped as a loose river rock gave way and lost the peaches he'd found to the churning water"). Prior to constructing the sentences, 63 additional subjects rated how willing they would be to eat various food items. We used these ratings to equate the desirability of the food items in the two categories of diagnostic sentences, to rule out potential confounds.

Dependent measures.

Categorization measure. Our categorization method was modeled after that developed by Taylor and colleagues (1978). In Taylor and colleagues' memory confusion protocol, the pattern of errors made by subjects in the surprise memory test is used to infer social categorization. To do this, we compared within-category confusions to between-category confusions. A within-category confusion occurs when a subject misattributes an act by, for example, an intentional undercontributor to a different intentional undercontributor. A between-category confusion occurs when a subject misattributes an act by, for example, an intentional undercontributor to an accidental undercontributor. If subjects make more within- than between-category confusions, this indicates that they are categorizing by the intentionality of the targets. (Correct responses are not analyzed because one cannot know if they are due to accurate memory, a within-category confusion, or random guessing.)

There are three ways to make a within-category confusion (a correct attribution to the original target is not an error), but four ways to make a between-category confusion. To correct for this difference in base rates, we multiplied the total number of between-category confusions by 3/4. (Without such a correction, random responding would appear as systematic attribution to the other category.) Using these corrected values, we created a categorization score for each subject: the number of within-category confusions minus the number of between-category confusions.[4] If subjects sorted individuals into two categories based on intentionality, then these difference scores should be significantly larger than zero.

Reactions to the targets. After the memory confusion protocol, each target's photo was displayed (one at a time) while subjects made a series of ratings about that target. Responses were made on 7-point scales, with anchors on 1 and 7. Two questions tapped punitive and reward sentiments by asking to what extent each target deserved to be punished/rewarded for his actions on the island (anchored at *Not At All* and *Very Much*). Two tapped willingness to collaborate in the future by asking, for each target, how willing the subject would be to work with that individual or have him as a member of their team (anchored at *Not At All* and *Very Much*). Five questions asked subjects to what extent a given adjective characterized a target: *trustworthy, competent, selfish, likeable,* and *aggressive,* with the first three being directly related to cooperation. For these adjectives, the scales were anchored at, for example, *Not At All Trustworthy* and *Very Trustworthy.* Two questions tapped dispositional versus situational attributions about cases in which the target failed to bring back food by asking to what extent this outcome was influenced by the target's "true personality" and to what extent by the situation (anchored at *Not At All* and *Very Much*). To check whether our attempt to manipulate perceptions of intentionality was successful, subjects were asked to what extent the target was behaving "on purpose" on those occasions when he did not bring back the food he had found (anchors at *Not At All* and *Very Much*).

Data analysis strategy. Because diagnostic and nondiagnostic sentences were likely to show different patterns of effects, we disaggregated by sentence diagnosticity. (Nondiagnostic sentences describe successful contribution and might be systematically (mis)attributed to people who accidentally failed to contribute.) Given that the diagnostic sentences carried our manipulation, we focus on them. (All analyses revealed categorization effects when calculated over nondiagnostic and diagnostic sentences together, but an examination of the data showed that these effects were driven entirely by the diagnostic sentences.)

For the diagnostic sentences, we also disaggregated the overall categorization score into its two subcomponents—one based on diagnostic sentences describing intentional failures to contribute and one based on diagnostic sentences describing accidental failures to contribute. By determining whether these separate categorization scores were both significantly greater than zero (and whether they differed from each other), we can determine whether categorization of one or the other person-type was driving the overall categorization score.

We used two-tailed p values and the Pearson correlation coefficient, r, as a measure of effect size (Rosenthal et al., 2000). Readers can calculate t values as Sqrt($r^2 \times df/[1 - r^2]$). Despite no prior predictions, across all studies we tested every categorization difference score and every rating measure for sex differences. Given the number of unplanned comparisons, a threshold p value of .001 is more liberal than a Bonferroni correction. Nonetheless, only one comparison had $p < .001$. We conclude that sex differences do not qualify any of our results.

It should be emphasized that at no point in the experiment was there any mention that some individuals might be cheating, stealing, or free riding; that some individuals might fail to contribute or that this failure might be intentional or unintentional; that there were two types of targets; or that subjects should try to categorize the targets into separate groups. Any categorization revealed by patterns of errors was spontaneously generated by the subjects.

Results

The results, including descriptive statistics and p values, are summarized in Tables 2, 3, and 4.

Manipulation check. The intentionality manipulation was successful: When targets failed to bring back food, the actions of the individuals who ate the food were perceived as more intentional than those of the individuals who lost the food ($r = .64$, $p < .001$; see Table 2).

Does categorization occur based on exploitive intent? Yes. The free rider strategy rule predicts that subjects would have categorization scores greater than zero, at least for the diagnostic sentences. As shown in Table 4, categorization did occur for the diagnostic sentences, with a very large effect size ($r = .70$, $p < .001$). Moreover, this effect was due to categorization of both types of people: those who failed by accident and those who failed intentionally. The strength of categorization did not differ for these two person-types. The nondiagnostic sentences revealed no evidence of categorization ($r = .00$).

Were targets perceived as free riders and cooperators? Although subjects' minds sorted the targets into distinct categories, does the category content correspond to the concepts FREE RIDER and COOPERATOR? If they do, then subjects will rate individuals who intentionally failed higher on the measures of punishment, selfishness, and personality as cause, relative to the individuals who accidentally failed. Additionally, subjects will rate the individuals who intentionally failed lower on the measures of reward, likeability, trustworthiness, willingness to have on their team, willingness to work with, and situation as cause. We made no predictions regarding the measures of competence and aggressiveness.

All predictions were supported: Compared to those who failed to contribute by accident, targets who failed intentionally reliably elicited the cluster of responses that would be expected of free riders (rs $\geq .35$, ps $< .01$; see Table 2). Moreover, the results revealed that targets who intentionally failed to contribute were viewed as more aggressive than targets who accidentally failed.

[4] Many studies using the memory confusion protocol treat within- and between-category errors (after correcting the latter for differing base rates) as separate levels within a repeated-measures analysis (e.g., Stangor, Lynch, Duan, & Glass, 1992; Taylor et al., 1978). Our approach using difference scores is formally identical and leads to the same statistical conclusions. This follows because our computation of difference scores is identical to computing a two-level within-subjects contrast, and such within-subjects contrasts are statistically identical to a repeated measures test (Rosenthal, Rosnow, & Rubin, 2000). Our approach, used previously by Kurzban, Tooby, and Cosmides (2001), has the added benefit of focusing attention on the theoretically important difference between the two types of errors, rather than the absolute numbers of errors, which is less informative.

Table 2
Means (Standard Deviations) for Reactions to the Target Types: Study 1

Reaction	Intentional failures	Unintentional failures	r
Punishment	4.06 (1.22)	3.12 (1.07)	$.58^{***}$
Reward	3.27 (1.16)	4.26 (1.29)	$.58^{***}$
Work with	3.37 (1.07)	4.36 (1.26)	$.57^{***}$
Have on team	3.25 (1.17)	4.34 (1.25)	$.60^{***}$
Trustworthy	3.37 (0.89)	4.43 (1.11)	$.61^{***}$
Selfish	4.47 (1.07)	3.35 (1.03)	$.62^{***}$
Likeable	3.51 (0.98)	4.38 (1.00)	$.59^{***}$
Aggressive	3.84 (1.20)	3.46 (1.00)	$.35^{**}$
Competence	4.69 (1.05)	4.79 (1.01)	.10
Personality as cause	4.65 (0.97)	3.57 (1.04)	$.62^{***}$
Situation as cause	3.97 (1.15)	4.82 (0.94)	$.48^{***}$
Intentionality (manipulation check)	4.85 (0.92)	3.71 (1.10)	$.64^{***}$

Note. Greater means indicate that a target is perceived as deserving more of a given outcome, being more desirable as a particular type of cooperation partner, or having more of a given trait or that a given cause is more applicable to the target. Response options ranged from 1 to 7. All comparisons had $df = 73$.
$^{**} p < .01$. $^{***} p < .001$.

There was no difference in the perceived competence of the target types.

Are perceptions of intentionality driving categorization? The two target types behaved in ways that conveyed differences in their exploitive intent; perceptions of these differences should be positively correlated with how strongly subjects categorized the targets. To test this, intentionality difference scores were computed for each subject (intentionality ratings: intentional minus accidental undercontributors). A greater difference score indicates that the behavior of those who ate the food was seen as more intentional than the behavior of those who lost food. We then correlated the intentionality difference scores with the overall diagnostic categorization scores. If perceptions of intentionality are related to categorization, then this correlation should be positive. This was the case: $r(72) = .36$, $p < .001$. This is consistent with the hypothesis that perceptions of intentionality (at least partially) drive free rider categorization.

Discussion

These results provide strong support for the free rider strategy rule. Results for the diagnostic sentences showed that both the intentional and accidental failures generated strong and significant categorization scores. Additionally, the cluster of responses to the targets is consistent with the hypothesis that the two target types were represented as free riders and cooperators, respectively. For instance, intentional undercontributors were viewed as deserving more punishment, being less trustworthy, and being less desired as a cooperative partner—reactions and attributions that, based on theory, should be directed toward free riders. Because targets with equal return rates were sorted into different categories, the results undermine the central prediction of the return rate rules for free rider categorization.

Study 2: Not Contributing Through Withholding Effort

Study 1 produced the pattern of results predicted by the free rider strategy rule. Was this because people were tracking cues of exploitive intent (as predicted), or was it an artifact of the specific manipulation we used? In Study 1, the intentional undercontributors disproportionately benefited—they consumed the food they found instead of sharing it with others—which is a cue of exploitive intent. But what if the mind is tracking benefits received rather than inferring an intention to free ride? To test against this possibility, the targets in Study 2 did not differ in the amount of food they consumed. Instead, we provided cues to exploitive versus cooperative intent by varying how much effort the targets expended while searching for food. Not trying is a cue to exploitive intent: It suggests one is motivated to avoid the cost of procuring resources for the group. Trying is a cue to cooperative intent; costs are incurred even if one fails. Not surprisingly, effort is routinely identified as important to judgments of cooperativeness (Gurven, 2004; Hill, 2002; Kerr, 1983; Miles & Klein, 2002).

Method

The design of this experiment was identical to Study 1, except as noted. There were 60 subjects (37 female), all undergraduates recruited from the same subject pools. Sixteen new diagnostic sentences were created. For the exploitive intent sentences, targets did not expend any effort and so did not find food (e.g., "He decided to take a nap instead of searching for food to bring back"). For the cooperative intent sentences, the targets tried hard to procure food but did not succeed (e.g., "He searched all over the island but found no food to bring back"). Thus, effort (and, therefore, cues to exploitive intent) was varied while amount of food provided was held constant. Importantly, nothing suggested

Table 3
Means and Standard Deviations of the Categorization Scores

Study	Overall categorization	Categorization separating by target type		Nondiagnostic sentences
1	3.40 (3.48)	Free riders: 1.88 (2.37)	Cooperators: 1.52 (2.33)	0.00 (3.47)
2	3.65 (3.90)	Free riders: 2.45 (2.60)	Cooperators: 1.20 (2.61)	−0.29 (3.48)
3	2.55 (3.03)	Lost food: 0.79 (2.65)	Always contributed: 1.75 (2.15)	0.39 (3.46)
4	0.56 (3.07)	Lost food: 0.41 (2.37)	Could not find food: 0.15 (1.88)	0.36 (3.80)
5A	2.02 (3.63)	Free riders: 0.50 (2.34)	Batterers: 1.51 (2.46)	0.25 (3.77)
5B	1.62 (2.84)	Free riders: 0.68 (2.07)	Thieves: 0.94 (2.20)	−0.33 (3.47)

Note. Except for the Nondiagnostic sentences column, all columns are based on data for diagnostic sentences (see main text).

Table 4
Categorization Effect Sizes (as Pearson Correlation Coefficients, r)

Study	Overall categorization	Categorization separating by target type		Difference in categorization by type	Nondiagnostic sentences
1	.70***	Free riders: .62***	Cooperators: .55***	.11	.00
2	.69***	Free riders: .69***	Cooperators: .42***	.34**	.08
3	.65***	Lost food: .29*	Always contributed: .64***	.25†	.11
4	.18	Lost food: .17	Could not find food: .08	.09	.10
5A	.49***	Free riders: .21†	Batterers: .53***	.31*	.07
5B	.50***	Free riders: .31**	Thieves: .40***	.08	.10

Note. Except for the Nondiagnostic sentences column, all columns are based on data for diagnostic sentences (see main text).
† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

that the exploitive targets got more to eat than anyone else. The measure assessing subjects' willingness to work with the targets was modified to read *work with one-on-one*; this was to more directly assess interest in engaging in dyadic cooperation with the target (a potentially distinct construct from willingness to engage in collective cooperation). An additional manipulation check assessed how much effort each target displayed as he searched for food (scale anchored at 1 [*None At All*] and 7 [*A Lot*]).

Results

The results, including descriptive statistics and p values, are summarized in Tables 3–5. They replicated those of Study 1, again providing strong evidence for the free rider strategy rule.

Manipulation checks. The effort manipulation was successful (see Table 5). Targets who were described as having expended effort were perceived as having expended more effort than targets who were described as not expending any effort ($r = .66$, $p < .001$). Moreover, when targets failed to provide food, this outcome was perceived as more intentional when it was caused by lack of effort rather than having tried but failed ($r = .62$, $p < .001$).

Does categorization occur based on intention to contribute? In Study 2, as in Study 1, responses to diagnostic sentences revealed that subjects categorized targets based on intention to contribute ($r = .69$, $p < .001$; see Tables 3 and 4). This effect size was essentially identical to Study 1's ($r = .70$), which argues that categorization in Study 1 was driven by cues of exploitive intent and was not boosted by greater caloric gains. Categorization scores for diagnostic sentences were significantly stronger for intentional failures (by hypothesis, free riders; $r = .69$, $p < .001$) compared to accidental failures ($r = .42$, $p < .001$; p of the comparison $<$ .01). The nondiagnostic sentences revealed no evidence of categorization ($r = .08$).

Table 5
Means (Standard Deviations) for Reactions to the Target Types: Study 2

Reaction	Intentional failures	Unintentional failures	r
Punishment	3.78 (1.16)	2.67 (1.05)	.64***
Reward	3.66 (1.08)	4.86 (0.86)	.69***
Work with	3.46 (1.02)	4.64 (0.93)	.67***
Have on team	3.46 (1.01)	4.74 (0.93)	.71***
Trustworthy	3.65 (0.99)	4.73 (0.98)	.64***
Selfish	4.18 (1.11)	3.10 (0.95)	.62***
Likeable	3.61 (0.99)	4.60 (1.01)	.62***
Aggressive	3.44 (0.93)	3.60 (1.15)	.14
Competence	4.09 (0.88)	4.74 (0.88)	.51***
Personality as cause	4.58 (0.91)	3.43 (0.92)	.65***
Situation as cause	3.75 (1.07)	4.74 (0.94)	.61***
Effort (manipulation check)	3.85 (0.85)	4.92 (0.80)	.66***
Intentionality (manipulation check)	4.05 (0.80)	3.08 (0.88)	.62***

Note. Greater means indicate a higher rating on the measure. Response options ranged from 1 to 7. All comparisons had $df = 59$.
*** $p < .001$.

Were targets perceived as free riders and cooperators? Results from the reaction and impression measures were largely consistent with those of Study 1: Targets who intentionally failed to contribute generated responses consistent with categorization as free riders (rs $> .60$, ps $< .001$). For instance, they were seen as less trustworthy, less desirable as a work partner or team member, and more deserving of punishment (see Table 5). For the two measures without predictions, the results were opposite from Study 1: Intentional failures were viewed as less competent, but there was no difference in perceived aggressiveness.

Do perceptions of intention to contribute correlate with categorization? Replicating the effect of Study 1, Study 2 showed that greater differences in perceptions of intentionality between the target types were associated with stronger categorization, $r(58) = .38$, $p < .01$. But could this relationship be driven by the difference in competence, not intentionality per se? No: Subjects' perceptions of differences in competence were not correlated with how strongly they categorized the targets, $r(58) = .18$, $p = .17$. Moreover, the relationship between intentionality and categorization remained unchanged when differences in competence were partialed out, partial $r(57) = .36$, $p < .01$. This supports the hypothesis that perceptions of intention to contribute—but not competence as a forager—drove categorization.

Discussion

Replicating Study 1, Study 2's results are consistent with the free rider strategy rule but not easily explained by the return rate rules. Despite having identical return rates, targets were spontaneously sorted into different categories depending on whether their failure to contribute was intentional. Compared to those who tried but failed, targets who failed because they did not try elicited a cluster of responses appropriate for free riders. Because all targets contributed equally in Studies 1 and 2, the results cannot be explained by return rate rules, which can only sort by differences in contributions. In contrast, the results of both studies are consistent with the free rider strategy rule.

Study 3: Is a Return Rate Rule Used to Categorize Free Riders When Return Rates Differ?

The return rate rules cannot explain the results of Studies 1 and 2. But it is possible that a return rate categorization rule is activated only when targets do vary in their return rates. Study 3 tested this by presenting targets differing in their return rates. All of them intend to contribute, all demonstrate some incompetence by virtue of losing something. Some targets always provision the group but sometimes lose a personal item (e.g., a camera). Other targets never lose a personal item but sometimes fail to provision the group because they accidentally lose the food they were trying to bring back.

There were two possibilities for Study 3. The first was that one of the return rate rules might categorize people as free riders, although perhaps it would operate more weakly than the free rider strategy rule (after all, an apparently honest failure might actually be an act of deception). The second possibility was that contribution level is used as cue to a different category altogether. Multi-individual cooperation requires much more than the ability to detect free riders (e.g., Neuberg et al., 2000). All else equal, people should prefer to selectively interact with the most competent individuals who are willing to work with them. Ratings from a preliminary study (n = 34) revealed that losing the group resource of food is viewed as less competent than losing a personal item in an otherwise identical situation (r = .57, p = .001). Given those ratings, losing the group resource of food (vs. a personal item) might be used to categorize people along a dimension of competence (Cuddy et al., 2008). Thus, although people may categorize based on contribution level, this category may reflect distinctions in competence, as opposed to a distinction between free riders and cooperators. This can be revealed by rating patterns and their correlation with categorization.

Method

All methods were identical to Study 1, except as noted. Fifty-nine undergraduates (38 female) participated. Sixteen new diagnostic sentences were created. Eight depicted targets as losing personal items by accident while retaining a collected food (e.g., "He wanted to check the time while taking strawberries back, but noticed his watch had fallen off"). The other eight depicted targets as losing food items (e.g., "Almost at camp, he noticed his bag had torn and the oranges he'd been carrying had been lost somewhere"). Thus, we varied contribution level while holding cooperative intent constant. Importantly, all targets lost an item. Questions that read "When this person did not bring food back . . . ." were changed in Study 3 to read "When this person lost or dropped something . . . ." The "work with" question was as in Study 2 ("work with one-on-one").

Results

The results are summarized in Tables 3, 4, and 6.

Manipulation check. As expected, there was no difference in how intentional the outcome was perceived to be as a function of whether the target lost food or personal items (r = .14; see Table 6).

Table 6
Means (Standard Deviations) for Reactions to the Target Types: Study 3

Reaction	Lost food	Always contributed	r
Punishment	2.57 (1.07)	2.56 (1.24)	.01
Reward	4.57 (1.26)	4.61 (1.28)	.06
Work with	4.46 (0.97)	4.48 (1.10)	.03
Have on team	4.58 (1.03)	4.67 (1.07)	.11
Trustworthy	4.53 (0.97)	4.59 (1.01)	.08
Selfish	2.85 (0.94)	2.81 (0.98)	.05
Likeable	4.51 (0.92)	4.51 (1.09)	.00
Aggressive	3.18 (1.14)	3.23 (1.18)	.04
Competence	4.57 (1.05)	4.74 (1.11)	$.22^{\dagger}$
Personality as cause	3.51 (1.06)	3.44 (1.07)	.08
Situation as cause	4.70 (0.94)	4.67 (0.97)	.03
Intentionality (manipulation check)	2.48 (1.02)	2.60 (1.07)	.14

Note. Greater means indicate a higher rating on the measure. Response options ranged from 1 to 7. All comparisons had df = 58.
$^{\dagger}p < .10$.

Did categorization occur based on contribution level? On the basis of the diagnostic sentences, subjects categorized by contribution level (r = .65, p < .001; see Table 3). Target individuals who always contributed (but lost personal items) were categorized more strongly than targets who sometimes lost food and thereby undercontributed (see Tables 3 and 4). The nondiagnostic sentences revealed no effects of categorization (r = .11).

What is the content of the categories formed? Consistent with the preliminary study, targets who contributed food but lost personal items were viewed as more competent than those who lost food (r = .22). This effect was marginal (p = .09, two-tailed), but significant if we take into account the prior prediction based on the preliminary results (p = .045, one-tailed).

If free riders are defined by a return rate rule, then discovering that someone has contributed less should activate anti-free rider responses—even when their undercontribution was accidental. These putative free riders should be seen, for example, as less trustworthy, less desirable for group cooperation, and more deserving of punishment. Yet no differences on these or any other items expected for free riders approached significance (all rs ≤ .10, all ps > .4; see Table 6). This suggests that lower return rates are not being used to categorize people as free riders but may be used to categorize them as incompetent.

Is competence correlated with categorization when cooperative intentions are held constant? Yes. There was a positive correlation (marginally significant) between competence difference scores and categorization, $r(57)$ = .23, p = .074. (Competence difference scores were calculated as average competence for targets losing personal items minus average competence for targets losing food.) Controlling for intentions did not alter this correlation, partial $r(56)$ = .24, p = .072. (Taking into account prior predictions, both correlations are significant with a one-tailed test at p < .05.) In Study 3, cues of cooperative intent were held constant; only amount of contribution delivered varied. Accordingly, there was no relationship between intention difference scores and categorization, $r(57)$ = .01, p = .94. This relationship was not improved by controlling for the competence difference score, partial $r(56) = -.05$, $p = .70$. In this experiment, perceptions of competence, but not perceptions of intention to contribute, were correlated with stronger categorization.

Discussion

Study 3 provides no evidence for either return rate rule for free rider categorization—despite a 40% difference in return rate between high and low contributors. All targets intended to contribute to the group, but some undercontributed by accidentally losing food. This difference in return rates was used to categorize targets in Study 3. But the content of the categories subjects formed was related to competence, not free riding: There were no differences in perceived trustworthiness, selfishness, punishment-worthiness, or any other free rider criterion measure. Instead, the only impression or reaction measure that revealed a difference between low and high return rate targets was the competence rating—and competence, but not intentionality, correlated with the strength of categorization.

Study 4: Domain-Specific Decision Rule or General Tendency?

We propose that the results of Studies 1 and 2 are due to a domain-specific decision rule for categorizing free riders. But could they instead be caused by a mechanism designed to categorize sets of entities whenever there is any systematic difference between them? If so, then this arbitrary categorization rule could account for the high levels of categorization observed above. On this hypothesis, people categorize targets on any available dimension—certainly on any differences that can be linguistically marked—and do so spontaneously. Previous work with the memory confusion protocol has shown that meaningless, but otherwise salient, visual cues do not elicit categorization (Stangor et al., 1992, Experiment 5), providing suggestive evidence that the arbitrary categorization rule is not operative. To conduct a critical test of the arbitrary categorization rule using our methodology, however, we used stimuli from Studies 1 and 2.

In Study 4, the diagnostic sentences all described situations in which the target failed to provision the group despite trying, but the reasons for this outcome differed. Four targets found a food item then lost it (the exact stimuli from Study 1), and four tried to find food but failed (the exact stimuli from Study 2). None should be viewed as free riders because all of them demonstrated cooperative intent. If failures to contribute activate a system for distinguishing free riders from cooperators, that system should categorize all of these targets as COOPERATORS; if so, they will not be sifted into separate categories.

In contrast, the arbitrary categorization rule would sort them into two distinct categories, for several reasons. First, they differed in foraging success—some found food, some did not—providing a difference for the mechanism to operate on. Second, this particular dimension of difference is easy to mark linguistically: Some targets found food (and lost it), others failed to find food. Third, the results of Studies 1 and 2 demonstrate that social categories can indeed be formed in response to the sentences used in Study 4. If categorization in Study 4 is as strong as in Studies 1 and 2, then the hypothesis that the prior results were caused by an arbitrary categorization rule, rather than the free rider strategy rule, cannot be eliminated.

Method and Results

Study 4's methods were identical to Study 2's, with one exception: The sentences from Study 1 that described targets losing food replaced the sentences from Study 2 that described targets who did not try to find food. Sixty-seven undergraduates (42 female) participated. The primary results of Study 4 are summarized in Tables 3 and 4 and in Supplementary Table 1, included in the online supplementary materials.

Did categorization occur when the targets had different foraging experiences but their intentions and contributions were equated? No. Neither the diagnostic sentences ($r = .18$, *ns*) nor the nondiagnostic sentences ($r = .10$, *ns*) revealed evidence of categorization (see Tables 3 and 4). Given this lack of categorization, it is not surprising that all questions assessing subjects' responses to the targets failed to show any significant differences (see Supplementary Table 1 in the online supplementary materials). Given that no significant effects emerged, we did not conduct correlational analyses.

Were subjects simply not paying attention? If subjects were not, for whatever reason, paying attention to the stimuli, then these null results would be uninformative. As a measure of subjects' attention, we can examine the overall number of errors subjects make in their attributions. At the limit, if subjects paid no attention to the face–sentence pairs, responding on the recall task should have been random. It was not: In this study and the others, the number of errors was always below the chance level of 35 errors (all ps $< .001$; see Table 7). Moreover, if subjects were paying less attention in this study than the others, error rates should have been significantly greater. They were not: Error rates did not vary across the six categorization studies, based on a one-way analysis of variance (ANOVA), $F(5, 384) = 0.87$, $p = .50$, and post hoc least significant difference (LSD) tests revealed no significant pairwise differences (all ps $> .09$). In fact, Study 4 actually had a descriptively smaller error rate compared to four of the five other categorization studies. Thus, a lack of attention to the materials (for whatever reason) cannot explain the fact that subjects did not sort the targets into two categories in Study 4.

Discussion

An arbitrary categorization rule, if it exists, should have sorted targets in Study 4 into two different categories. Instead, no signif-

Table 7
Mean Number of Errors and Error Rates as a Function of Study

Study	Mean number of errors (*SD*)	Errors as percentage of total attributions
1	29.99 (4.27)	75%
2	28.93 (4.27)	72%
3	29.66 (5.10)	74%
4	29.10 (4.55)	73%
5A	30.30 (4.23)	76%
5B	29.79 (3.94)	74%

Note. Each subject made 40 attributions.

icant categorization effects occurred in Study 4, where all the targets demonstrated cooperative intent. Although these cooperators differed in their food-finding success and they failed to contribute to the group for different reasons, these differences were not used for social categorization. Moreover, the absence of categorization was not due to a lack of attention by subjects. Nor was it a function of the sentences used: All elicited high levels of categorization in Studies 1 or 2. In Studies 1 and 2, however, they appeared alongside a theoretically meaningful contrast: free riders. This suggests that a free rider strategy rule is activated when people fail to contribute to a collective action and that it privileges categories such as COOPERATOR and FREE RIDER over other types of distinctions.

One could argue—and we would agree—that whatever categorization processes the mind uses, their operation should be contingent on the situation. Thus, categorization along Dimension X might happen in Situations A, B, and C, but not J, K, and L. Could that be the case here? If the distinction between Study 4 targets was not relevant to the situation, then the null results would have little bearing on whether a general, but contingent, categorization process exists. Although possible, this seems unlikely: The scenario described targets who differed in the amount of food they were able to acquire during a life and death situation where food was scarce. The dimension of food acquisition certainly seems relevant given the situation; it is therefore difficult to see how a general but contingent categorization ability would not use it. The lack of categorization is explicable, however, on the present theory, which predicts that individuals who try to contribute should be folded into the category COOPERATOR.

Studies 5A and 5B: Free Riders or Intentional Moral Violators?

As a final alternative, we consider whether the categorization results from Studies 1–4 could have been produced by the moral violator rule rather than the free rider strategy rule. According to this hypothesis, the mind is designed to place moral violators in a separate category from other individuals but does not make finer distinctions among the moral violators. If this were true, the same pattern of results might have been found even in the absence of a system specialized for categorizing free riders. To test against this possibility, Studies 5A and 5B compared free riders (stimuli from Study 1) to two other classes of moral violators.

In Study 5A, the comparison class was individuals who were physically violent in an unprovoked manner. Unprovoked, intentional battery was chosen because it should be defined as immoral by any theory of moral psychology (Mikhail, 2007). Even if Study 5A is successful, however, it leaves open a slightly more textured alternative hypothesis. The mind might distinguish some types of moral violations from others, but the criteria might be coarse grained, distinguishing intentional battery (a violent crime) from violations of agreements, contracts, and property rights. On this view, free riders would be distinguished from batterers. But because the free riders have violated their agreement to share the resources they find with the group, they will not be distinguished from those who violate other agreements, contracts, or property rights. To test against this possibility, in Study 5B the comparison class was individuals who stole a resource that was communally owned by the group. Not only does this provide a closer contrast to free riders than batterers but it also holds constant the identity of the wronged party—free riders free ride on the group and thieves steal from the group.

If the results of the prior experiments were generated by a moral violator rule, in the absence of a free rider strategy rule, then all the targets in Studies 5A and 5B will be sorted into the same category: MORAL VIOLATOR. There will be no evidence that the targets are sorted into two distinct categories—especially ones involving the fine-grained distinction tested in Study 5B. After all, Study 4 demonstrated that the presence of an obvious difference between two sets of targets is not sufficient to elicit categorization effects. So if subjects sort moral violators into two distinct categories in Study 5, this cannot be easily attributed to the mere fact that free riders behave differently from thieves and batterers.

By contrast, there should be strong categorization effects if the mind contains a free rider strategy rule alongside rules for categorizing other kinds of moral violations. Different kinds of moral violation require different responses, so the mind should be designed to make fine-grained distinctions among them. The free rider strategy rule should be just one of many such rules.

Method

Except as noted, the methods were identical to Study 2. There were 64 undergraduate subjects in Study 5A (34 female) and 66 (33 female) in Study 5B. For the diagnostic sentences, four targets were paired with the free rider sentences from Study 1, which described the target as finding food but eating it himself; this was true in Studies 5A and 5B. In Study 5A, the other four targets were paired with eight new diagnostic sentences. These depicted the targets as providing food and being physically violent toward others on the island (e.g., "After bringing peaches to camp, he picked a fight and beat someone up"). These sentences described aggression without provocation (because provoked aggression might be seen by subjects as justifiable and not immoral). Study 5A therefore contrasted targets who free ride with targets who commit battery. In Study 5B, the other four targets were paired with eight new diagnostic sentences that depicted the targets as stealing a resource owned by the group (e.g., "After returning to camp, he took fuel normally used for communal cooking, so he could warm himself at night"). Study 5B therefore contrasted targets who free ride with targets who steal from the group. In both Studies 5A and 5B, impression and reaction questions that previously read "When this person did not bring food back . . . ." were changed to read "When this person did something wrong or inappropriate . . . ."

Results

The results of Studies 5A and 5B are summarized in Tables 3 and 4 and in Supplementary Tables 2 and 3 available online.

In Study 5A, does categorization occur when the targets are free riders and violent people? On the basis of the diagnostic sentences, free riders and physically violent people were categorized as separate types ($r = .49$, $p < .001$; see Table 4). Interestingly, physically violent individuals generated stronger categorization scores (see Table 4). As usual, there was no evidence of categorization based on the nondiagnostic sentences ($r = .07$).

In Study 5B, does categorization occur when the targets are free riders and thieves who stole from the group? On the basis of the diagnostic sentences, free riders and thieves were categorized as separate types ($r = .50$, $p < .001$), and their levels of categorization were not statistically distinguishable (see Table 4). There was no evidence of categorization based on the nondiagnostic sentences ($r = -.10$).

Reactions to the targets. Supplementary Tables 2 and 3 available online present the full data for reactions to the targets in Studies 5A and 5B; here, we summarize these results. In Study 5A, as expected, physically violent individuals were seen as more aggressive than free riders ($r = .46$). Moreover, the physically violent individuals were seen as deserving more punishment and (marginally) as deserving less reward. Subjects also indicated that they would rather work one-on-one with a free rider than a physically violent individual. There was no difference in how desirable the two targets were as members of one's team. Physically violent people were seen as behaving more intentionally, and they were seen as less likeable. No other effects were significant (see Supplementary Table 2, available online). We note that these negative reactions were found to violent targets even though they contributed more food than free riders.

In Study 5B, free riders, relative to thieves, were seen as less desirable as members of one's team, less desirable to work with one-on-one, less trustworthy, more selfish, and (marginally) deserving of less reward (see Supplementary Table 3, available online). Interestingly, although free riders were viewed less negatively than batterers in Study 5A, they were perceived more negatively than thieves in Study 5B—even though thieves were stealing resources from the group, whereas free riders were failing to provide promised resources to the group (both benefited from these actions).

We propose that these effects are due in part to a specialized FREE RIDER concept. However, it is known that the mind can form context-sensitive categorical distinctions based on an underlying continuous dimension (e.g., Hampton, 2007; Jackendoff, 1983). As an alternative account of Studies 5A and 5B, perhaps subjects were simply distinguishing along a continuous dimension of, for example, moral severity. As we detail in the supplemental online information, this is unlikely. If this alternative is correct, then a perceived difference on at least one of the impression items should predict the strength of categorization (assuming that at least one of these taps the dimension being used for categorization). None do (for both studies, all *r*s $< .17$ in absolute magnitude, all *p*s $> .18$). Even though there were differences in reactions to free riders and other moral violators, none of these differences appear to have driven categorization.

Discussion

Studies 1–4 presented results that are well explained by the operation of the free rider strategy rule but are also consistent with the moral violator rule. Operating by itself, the moral violator rule should not produce two distinct categories in Study 5A or in Study 5B because all the targets were moral violators. But this is not what happened: Both Studies 5A and 5B revealed strong categorization effects when contrasting free riders with batterers or thieves. Combining this with Study 4's finding that arbitrary distinctions are not categorized, these results eliminate the hypothesis that the categorization of free riders found in Studies 1 and 2 was caused by a coarse-grained mechanism that only discriminates moral violators from nonviolators.

Our data are therefore consistent with the free rider strategy rule. However, these data are also consistent with the possibility that a superordinate level of categorization exists where free riders, thieves, and batterers are all marked as immoral (such as the warmth/morality dimension in the stereotype content model; Cuddy et al., 2008). Nonetheless, a more specific rule distinguishing free riders from batterers and thieves is either preempting its operation here or carving more fine-grained distinctions within the superordinate category.

Although predicted by evolutionary considerations, the fact that thieves who steal from the group and those who free ride on the group are distinguished from each other is surprising from most theoretical perspectives (e.g., attribution theories). Both types of violation are broadly related to agreements, contracts, and property rights, so, in sorting individuals into these two categories, the mind is making a very fine-grained distinction between these two types of nonviolent moral violators.

It remains an open question, however, whether free riders would be distinguished from an even more similar category, that of cheaters—individuals who intentionally defect on an agreement to engage in social exchange (Cosmides, Barrett, & Tooby, 2010). Social exchange involves cooperation for mutual benefit between two agents. Multiple lines of evidence show that the mind contains mechanisms specialized for social exchange, including a concept of CHEATER (Cosmides & Tooby, 2005; Ermer, Guerin, Cosmides, Tooby, & Miller, 2006; Fiddick, Cosmides, & Tooby, 2000; Stone, Cosmides, Tooby, Kroll, & Knight, 2002; Sugiyama, Tooby, & Cosmides, 2002). Indeed, the human mind has mechanisms that are good at detecting violations of social exchanges when these reveal the presence of a cheater—an intentional violator—but not when the violations are due to innocent mistakes (Cosmides et al., 2010; Fiddick, 2004). Moreover, adaptations for collective action will likely rely on adaptations for social exchange (Tooby et al., 2006). Thus, the concepts of FREE RIDER and CHEATER may be intimately related. Even if future work reveals the concepts to be essentially identical—perhaps a superordinate category of DEFECTOR—the present research nonetheless adds to the literature on cheater detection in three important ways. First, no previous research has shown that defectors are confusable with one another, a hallmark of entities all assigned to the same category. Second, previous work has usually examined a single individual contracting with another agent (which may be another individual or a group); this is the first research to show that the DEFECTOR concept applies to individuals all mutually contracting with a single collective agent that is composed of the same individuals (here, the group on the island). Third, although past research has contrasted cheaters with violators of other types of social rules relating to obligations, requirements, and entitlements, this is the first research to show that defection is distinguished from other, clearly moral violations.

General Discussion

Free riders pose a barrier to the evolution and persistence of collective action. For such cooperation to evolve and to persist, the mind must be able to solve the adaptive problem posed by free riders. With this research, we have attempted to identify one component of the psychological architecture that evolved to enable collective action: the criteria the mind uses to identify free riders. Considered together, our results support the existence of a psychological adaptation for categorizing free riders: the free rider strategy rule.

To be an adaptation, a hypothesized mechanism must show efficiency, economy, selectivity, specificity, and precision—in other words, it must show evidence of special design for solving an adaptive problem (Williams, 1966; see also Dawkins, 1986; Tooby & Cosmides, 1992). Our results show that free rider categorization was a precise process: Only those people who failed to contribute by virtue of consuming the benefit or avoiding the cost of producing it were categorized as free riders (Studies 1 and 2); moreover, this was not due to the operation of an arbitrary categorization system (Study 4). Second, the process was selective: Free riders were distinguished from other moral violators such as batterers and thieves—even though the thieves were taking from the group, a harm comparable to that inflicted by free riders (Studies 5A and 5B). Third, responses to free riders were quite specific: Free riders reliably elicited a cluster of predicted responses (e.g., less trust, more punishment), and this cluster was not elicited by individuals who accidentally failed to contribute (Studies 1–3). Fourth, categorization happened efficiently and spontaneously: No explicit instructions were given to look for free riders, cheaters, or thieves, nor was there a suggestion to the subjects that two types of people existed on the island to whom they should attend. Instead, the experimental instructions asked subjects to form impressions of the targets. If anything, this instructional set should have led to a decrease in category-based impression formation and an increase in individuation of the targets (Fiske & Neuberg, 1990).

In the process of testing for the free rider strategy rule, we ruled out several alternative hypotheses of more general scope: the arbitrary categorization rule, the moral violator rule, and the return rate rules for free rider categorization. But are there other ways to explain the results? Here, we consider two additional alternatives, attribution theory and rational choice theory.

Study 5, which tested the moral violator rule against the free rider strategy rule, separated the theory elaborated here from attribution theory. Attribution theory provides a framework for understanding how lay perceivers categorize behavior as caused by situational or dispositional forces (for reviews, see Gilbert, 1995; Moskowitz, 2005). One potential cue for a dispositional attribution might be inferring that a behavior was intentionally caused. But attribution theory provides no basis for predicting that the mind will spontaneously distinguish among intentional moral violators, sorting free riders into a different category from batterers and thieves. Yet free riders were distinguished from these other intentional moral violators in Studies 5A and 5B even though free riding, battery, and theft are all due to internal causes that are controllable (Weiner, 1993; Weiner et al., 1988). Moreover, attribution theory provides no basis for predicting the specific pattern of responses to free riders. In contrast, these responses were a first-order prediction of an evolutionary analysis of the adaptive problem posed by free riders.

Similarly, rational choice theory cannot explain the full pattern of results. Rational choice models with standard self-interested utility functions require decision rules that base choices on differences in objective payoffs. The less individuals contribute to a collective action, the less they should be preferred as collective action partners and the more punitive sentiment they should attract. Payoffs to other collective action members are lower whether a fellow member failed to contribute by accident or intention, so this distinction should not matter to a rational utility maximizer. Obviously, this first-order prediction of an economic account is contradicted by the results of Studies 1 and 2, where intentional undercontributors were not only sorted into a different mental category from accidental ones but also attracted more punitive sentiment and were less preferred as future cooperative partners.

To rescue an economic perspective, one could tweak rational choice theory by positing that when an individual undercontributes by intention, others use this as a probabilistic cue to future undercontribution. On this view, intentions are used as a proxy to predict the real variable of interest: future return rates from various targets. This twist on rational choice theory cannot, however, explain the results of Study 3. In that study, intentions to contribute were held constant, but targets differed in their relative return rates. When all targets intend to contribute, it is not possible to use differences in their intentions as a cue to predict differences in their future return rates. However, subjects in this study had access to a cue that should be an even better predictor of future return rates: the targets' present return rates. A cue-driven economic model would have to predict that targets who contribute at a lower rate will be less preferred as collective action partners and seen as more worthy of punishment. In contrast to this straightforward prediction, targets with low return rates were just as preferred for future collective actions as those with high return rates; they were also seen as equally trustworthy, likeable, and unworthy of punishment.

Limitations

Like many experimental lab-based studies, this research is limited by the use of an undergraduate population, the use of a hypothetical scenario, the fact that subjects were not invested in the group of people they learned about, the static and artificial nature of the photos and sentences, and so forth. However, this research provides experimental evidence that converges with more ecologically valid observational work demonstrating that exploitive intentions are an important component of free rider identification in ethnographically diverse populations (Gurven, 2004; Hill, 2002; Price, 2005). For example, Price (2006b) measured contributions to a *minga*—a collective action—among the Shuar (a hunter-horticulturalist group in Ecuador). In this *minga,* individuals labored to clear a field to grow sugarcane, with the goal of splitting the profits after it was sold. These Shuar made a sharp distinction between unintentional and intentional absences from the *minga* and were very accurate in keeping track of when an individual was intentionally absent. Unintentional absences (such as illness) were excused, but intentional absences were not. Finally, the more days that individuals in the *minga* were intentionally absent, the more punitive sentiment they received.

Implications and Future Directions

Our goal was to create a high-resolution map of a small but important piece of conceptual machinery: the FREE RIDER concept. In our view, this is but one piece of a large and interconnected set of concepts related to collective action and coalitional cooperation. In other work, for instance, we have mapped some of the concepts involved in coalitional affiliation (Cosmides, Tooby, & Kurzban, 2003; Kurzban, Tooby, & Cosmides, 2001) and in integrating newcomers into coalitions (Cimino & Delton, 2010; Delton & Cimino, 2010). The work on newcomers is particularly instructive in light of the current article. When newcomers receive benefits merely by joining the group, they pose similar adaptive problems as standard free riders. But a detailed functional analysis reveals important differences as well. For example, newcomers should not activate exclusion sentiment in the same way as free riders—a prediction that has been empirically verified and is relevant to understanding hazing (Cimino & Delton, 2010; Delton & Cimino, 2010). This illustrates the value of using a functional lens to further fractionate the underlying conceptual machinery of social thought.

One important direction for future research would be to refine the conceptual semantics of the free rider concept. So far, the results suggest that the mind defines FREE RIDER in terms of other pieces of conceptual structure, something along the lines of: Given a COLLECTIVE ACTION, *if* an AGENT is a PARTICIPANT and a BENEFICIARY and INTENTIONALLY fails to CONTRIBUTE through an EXPLOITIVE MOTIVATION, *then* mark the AGENT as a FREE RIDER. All of these pieces of conceptual structure are themselves open to empirical investigation (the AGENT and INTENTION concepts have already received some; Johnson, 2000). For example, what defines a PARTICIPANT? Does one have to explicitly or implicitly agree to become a member of a collective action, or is simply being an (able-bodied) BENEFICIARY enough? Preliminary research suggests that the mind does contain an inference such that being a beneficiary—even if one has not agreed to participate—nonetheless causes a person to be viewed as obligated to contribute (Delton, Nemirow, Robertson, Cimino, & Cosmides, 2011). However, the logic of this piece of conceptual structure and others remains to be fully articulated.

The Architecture of FREE RIDER and Other Moral Concepts

The goal of this research was to dissect the architecture of a moral concept: FREE RIDER. The human mind does not equate free riders with undercontribution, nor does it lump free riders into a general category that contains all moral violators. Instead, as predicted by an adaptationist approach, the mind classifies individuals as free riders only when their behavior indicates they have a psychological design or calibration that causes them to consume benefits while withholding contributions. This fits with predictions from evolutionary game theory: An evolved adaptation for detecting free riders should use criteria that identify only those individuals whose behavior poses a threat to the stability of collective action. We do not think this is an isolated case. Just as it allowed the empirical illumination of part of the architecture of this moral concept, sustained use of adaptationist reasoning should be able to shed light on other features of our moral and social psychology that are hidden from view.

References

Alexander, R. D. (1987). *The biology of moral systems.* New York, NY: Aldine de Gruyter.

Axelrod, R. (1984). *The evolution of cooperation.* New York, NY: Basic Books.

Axelrod, R., & Hamilton, W. D. (1981, March 27). The evolution of cooperation. *Science, 211,* 1390–1396. doi:10.1126/science.7466396

Baron-Cohen, S. (1995). *Mindblindness.* Cambridge, MA: MIT Press.

Barrett, H. C. (2005). Enzymatic computation and cognitive modularity. *Mind and Language, 20,* 259–287. doi:10.1111/j.0268-1064.2005.00285.x

Boesch, C. (2002). Cooperative hunting roles among Tai chimpanzees. *Human Nature, 13,* 27–46. doi:10.1007/s12110-002-1013-6

Boyd, R., & Richerson, P. J. (1988). The evolution of reciprocity in sizable groups. *Journal of Theoretical Biology, 132,* 337–356. doi:10.1016/S0022-5193(88)80219-4

Boyd, R., & Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology, 13,* 171–195. doi:10.1016/0162-3095(92)90032-Y

Bugental, D. B. (2000). Acquisition of the algorithms of social life: A domain-based approach. *Psychological Bulletin, 126,* 187–219. doi:10.1037/0033-2909.126.2.187

Callaghan, T., Rochat, P., Lillard, A., Claux, M. L., Odden, H., Itakura, S., . . . Singh, S. (2005). Synchrony in the onset of mental-state reasoning: Evidence from five cultures. *Psychological Science, 16,* 378–384. doi:10.1111/j.0956-7976.2005.01544.x

Cashdan, E. (1992). Spatial organization and habitat use. In E. A. Smith & B. Winterhalder (Eds.), *Evolutionary ecology and human behavior* (pp. 237–266). New York, NY: Aldine de Gruyter.

Cimino, A., & Delton, A. W. (2010). On the perception of newcomers: Toward an evolved psychology of intergenerational coalitions. *Human Nature, 21,* 186–202. doi:10.1007/s12110-010-9088-y

Cosmides, L., Barrett, H. C., & Tooby, J. (2010). Adaptive specializations, social exchange, and the evolution of human intelligence. *Proceedings of the National Academy of Sciences of the United States of America, 107,* 9007–9014. doi:10.1073/pnas.0914623107

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163–228). New York, NY: Oxford University Press.

Cosmides, L., & Tooby, J. (2005). Neurocognitive adaptations designed for social exchange. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 584–627). Hoboken, NJ: Wiley.

Cosmides, L., Tooby, J., & Kurzban, R. (2003). Perceptions of race. *Trends in Cognitive Sciences, 7,* 173–179. doi:10.1016/S1364-6613(03)00057-3

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in Experimental Social Psychology, 40,* 61–149. doi:10.1016/S0065-2601(07)00002-0

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition, 108,* 353–380. doi:10.1016/j.cognition.2008.03.006

Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology, 31,* 169–193. doi:10.1146/annurev.ps.31.020180.001125

Dawes, R. M., Mctavish, J., & Shaklee, H. (1977). Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation. *Journal of Personality and Social Psychology, 35,* 1–11. doi:10.1037/0022-3514.35.1.1

Dawkins, R. (1986). *The blind watchmaker.* New York, NY: Norton.

De Bruin, E. N. M., & Van Lange, P. A. M. (1999). Impression formation and cooperative behavior. *European Journal of Social Psychology, 29,* 305–328. doi:10.1002/(SICI)1099-0992(199903/05)29:2/3<305::AID-EJSP929>3.0.CO;2-R

Delton, A. W., & Cimino, A. (2010). Exploring the evolved concept of newcomer: Experimental tests of a cognitive model. *Evolutionary Psychology, 8,* 317–335.

Delton, A. W., Krasnow, M. M., Cosmides, L., & Tooby, J. (2011). The evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters. *Proceedings of the National Academy of Sciences of the United States of America, 108,* 13335–13340. doi:10.1073/pnas.1102131108

Delton, A. W., Nemirow, J., Robertson, T. E., Cimino, A., & Cosmides, L. (2011, June–July). *Obligated to contribute? The effects of excludability on moralization in collective action.* Paper presented at the 23rd Annual Human Behavior and Evolution Society Conference, Montpellier, France.

Dunbar, R. I. M. (2004). Gossip in evolutionary perspective. *Review of General Psychology, 8,* 100–110. doi:10.1037/1089-2680.8.2.100

Ermer, E., Guerin, S., Cosmides, L., Tooby, J., & Miller, M. (2006). Theory of mind broad and narrow: Reasoning about social exchange engages ToM areas, precautionary reasoning does not. *Social Neuroscience, 1,* 196–219. doi:10.1080/17470910600989771

Fehr, E., & Gachter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review, 90,* 980–994. doi:10.1257/aer.90.4.980

Fiddick, L. (2004). Domains of deontic reasoning: Resolving the discrepancy between the cognitive and moral reasoning literatures. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 57*(A), 447–474.

Fiddick, L., Cosmides, L., & Tooby, J. (2000). No interpretation without representation: The role of domain-specific representations and inferences in the Wason selection task. *Cognition, 77,* 1–79. doi:10.1016/S0010-0277(00)00085-8

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression-formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology, 23,* 1–74. doi:10.1016/S0065-2601(08)60317-2

Gilbert, D. T. (1995). Attribution and interpersonal perception. In A. Tesser (Ed.), *Advanced social psychology* (pp. 99–147). New York, NY: McGraw-Hill.

Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of Theoretical Biology, 206,* 169–179. doi:10.1006/jtbi.2000.2111

Gurven, M. (2004). To give and to give not: The behavioral ecology of human food transfers. *Behavioral and Brain Sciences, 27,* 543–583. doi:10.1017/S0140525X04000123

Hamilton, W. D. (1964). The genetical evolution of social behaviour. *Journal of Theoretical Biology, 7,* 1–52. doi:10.1016/0022-5193(64)90038-4

Hampton, J. A. (2007). Typicality, graded membership, and vagueness. *Cognitive Science, 31,* 355–384. doi:10.1080/15326900701326402

Harris, P. (1990). The child's theory of mind and its cultural context. In G. Butterworth & P. Bryant (Eds.), *Causes of development* (pp. 215–237). Hillsdale, NJ: Erlbaum.

Haselton, M. G., & Nettle, D. (2006). The paranoid optimist: An integrative evolutionary model of cognitive biases. *Personality and Social Psychology Review, 10,* 47–66. doi:10.1207/s15327957pspr1001_3

Hauert, C., De Monte, S., Hofbauer, J., & Sigmund, K. (2002a). Replicator dynamics for optional public good games. *Journal of Theoretical Biology, 218,* 187–194. doi:10.1006/jtbi.2002.3067

Hauert, C., De Monte, S., Hofbauer, J., & Sigmund, K. (2002b, May 10). Volunteering as Red Queen mechanism for cooperation in public goods games. *Science, 296,* 1129–1132. doi:10.1126/science.1070582

Henrich, J., & Boyd, R. (2001). Why people punish defectors: Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal of Theoretical Biology, 208,* 79–89. doi:10.1006/jtbi.2000.2202

Hill, K. (2002). Altruistic cooperation during foraging by the Ache, and the evolved human predisposition to cooperate. *Human Nature, 13,* 105–128. doi:10.1007/s12110-002-1016-3

Jackendoff, R. (1983). *Semantics and cognition.* Cambridge, MA: MIT Press.

Jackendoff, R. (2006). *Language, consciousness, culture: Essays on mental structure.* Cambridge, MA: MIT Press.

Johnson, S. C. (2000). The recognition of mentalistic agents in infancy. *Trends in Cognitive Sciences, 4,* 22–28. doi:10.1016/S1364-6613(99)01414-X

Kameda, T., Takezawa, M., & Hastie, R. (2003). The logic of social sharing: An evolutionary game analysis of adaptive norm development. *Personality and Social Psychology Review, 7,* 2–19. doi:10.1207/S15327957PSPR0701_1

Kameda, T., Takezawa, M., Tindale, R. S., & Smith, C. M. (2002). Social sharing and risk reduction: Exploring a computational algorithm for the psychology of windfall gains. *Evolution and Human Behavior, 23,* 11–33. doi:10.1016/S1090-5138(01)00086-1

Kaplan, H., & Hill, K. (1985). Food sharing among Ache foragers: Tests of explanatory hypotheses. *Current Anthropology, 26,* 223–246. doi:10.1086/203251

Kaplan, H., Hill, K., & Hurtado, A. (1990). Risk, foraging and food sharing among the Ache. In E. Cashdan (Ed.), *Risk and uncertainty in tribal and peasant economies* (pp. 107–144). Boulder, CO: Westview Press.

Kaplan, H., Hill, K., Lancaster, J., & Hurtado, A. M. (2000). A theory of human life history evolution: Diet, intelligence, and longevity. *Evolutionary Anthropology, 9,* 156–185. doi:10.1002/1520-6505(2000)9:4<156::AID-EVAN5>3.0.CO;2-7

Kenrick, D. T., Li, N. P., & Butner, J. (2003). Dynamical evolutionary psychology: Individual decision rules and emergent social norms. *Psychological Review, 110,* 3–28. doi:10.1037/0033-295X.110.1.3

Kenrick, D. T., Maner, J. K., & Li, N. P. (2005). Evolutionary social psychology. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 803–827). Hoboken, NJ: Wiley.

Kerr, N. L. (1983). Motivation losses in small groups: A social dilemma analysis. *Journal of Personality and Social Psychology, 45,* 819–828. doi:10.1037/0022-3514.45.4.819

Kerr, N. L., Rumble, A. C., Park, E. S., Ouwerkerk, J. W., Parks, C. D., Gallucci, M., & van Lange, P. A. M. (2009). "How many bad apples does it take to spoil the whole barrel?": Social exclusion and toleration for bad apples. *Journal of Experimental Social Psychology, 45,* 603–613. doi:10.1016/j.jesp.2009.02.017

Kiyonari, T., & Barclay, P. (2008). Cooperation in social dilemmas: Free riding may be thwarted by second-order reward rather than by punishment. *Journal of Personality and Social Psychology, 95,* 826–842. doi:10.1037/a0011381

Kiyonari, T., Tanida, S., & Yamagishi, T. (2000). Social exchange and reciprocity: Confusion or a heuristic? *Evolution and Human Behavior, 21,* 411–427. doi:10.1016/S1090-5138(00)00055-6

Klapwijk, A., & Van Lange, P. A. M. (2009). Promoting cooperation and trust in "noisy" situations: The power of generosity. *Journal of Personality and Social Psychology, 96,* 83–103. doi:10.1037/a0012823

Klauer, K. C., & Wegener, I. (1998). Unraveling social categorization in the "Who said what?" paradigm. *Journal of Personality and Social Psychology, 75,* 1155–1178. doi:10.1037/0022-3514.75.5.1155

Kurzban, R., & Leary, M. R. (2001). Evolutionary origins of stigmatization: The functions of social exclusion. *Psychological Bulletin, 127,* 187–208. doi:10.1037/0033-2909.127.2.187

Kurzban, R., McCabe, K., Smith, V. L., & Wilson, B. J. (2001). Incremental commitment and reciprocity in a real-time public goods game. *Personality and Social Psychology Bulletin, 27,* 1662–1673. doi:10.1177/01461672012712009

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the*

National Academy of Sciences of the United States of America, 98, 15387–15392. doi:10.1073/pnas.251541498

Leslie, A. M. (1994). ToMM, ToBy, and agency: Core architecture and domain specificity. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 119–148). Cambridge, England: Cambridge University Press. doi:10.1017/CBO9780511752902.006

Malle, B. F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology, 33,* 101–121. doi:10.1006/jesp.1996.1314

Masclet, D., Noussair, C., Tucker, S., & Villeval, M. C. (2003). Monetary and nonmonetary punishment in the voluntary contributions mechanism. *American Economic Review, 93,* 366–380. doi:10.1257/000282803321455359

Maynard Smith, J. (1982). *Evolution and the theory of games.* Cambridge, England: Cambridge University Press.

Messe, L. A., & Sivacek, J. M. (1979). Predictions of others responses in a mixed-motive game: Self-justification or false consensus? *Journal of Personality and Social Psychology, 37,* 602–607. doi:10.1037/0022-3514.37.4.602

Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences, 11,* 143–152. doi:10.1016/j.tics.2006.12.007

Miles, J. A., & Greenberg, J. (1993). Using punishment threats to attenuate social loafing effects among swimmers. *Organizational Behavior and Human Decision Processes, 56,* 246–265. doi:10.1006/obhd.1993.1054

Miles, J. A., & Klein, H. J. (2002). Perception in consequences of free riding. *Psychological Reports, 90,* 215–225.

Moskowitz, G. B. (2005). *Social cognition.* New York, NY: Guilford Press.

Neuberg, S. L., Smith, D. M., & Asher, T. (2000). Why people stigmatize: Toward a sociofunctional framework. In T. F. Heatherton, R. E. Kleck, M. R. Hebl, & J. G. Hull (Eds.), *The social psychology of stigma* (pp. 31–61). New York, NY: Guilford Press.

Nowak, M. A., & Sigmund, K. (2005, October 27). Evolution of indirect reciprocity. *Nature, 437,* 1291–1298. doi:10.1038/nature04131

Olson, M. (1965). *The logic of collective action: Public goods and the theory of groups.* Cambridge, MA: Harvard University Press.

Onishi, K. H., & Baillargeon, R. (2005, April 8). Do 15-month-old infants understand false beliefs? *Science, 308,* 255–258. doi:10.1126/science.1107621

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action.* Cambridge, England: Cambridge University Press.

Panchanathan, K., & Boyd, R. (2003). A tale of two defectors: The importance of standing for evolution of indirect reciprocity. *Journal of Theoretical Biology, 224,* 115–126. doi:10.1016/S0022-5193(03)00154-1

Panchanathan, K., & Boyd, R. (2004, November 25). Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature, 432,* 499–502. doi:10.1038/nature02978

Pinker, S. (2007). *The stuff of thought: Language as a window into human nature.* New York, NY: Viking Press.

Price, M. E. (2005). Punitive sentiment among the Shuar and in industrialized societies: Cross-cultural similarities. *Evolution and Human Behavior, 26,* 279–287. doi:10.1016/j.evolhumbehav.2004.08.009

Price, M. E. (2006a). Judgments about cooperators and free riders on a Shuar work team: An evolutionary psychological perspective. *Organizational Behavior and Human Decision Processes, 101,* 20–35. doi:10.1016/j.obhdp.2006.06.001

Price, M. E. (2006b). Monitoring, reputation, and "greenbeard" reciprocity in a Shuar work team. *Journal of Organizational Behavior, 27,* 201–219. doi:10.1002/job.347

Price, M. E., Cosmides, L., & Tooby, J. (2002). Punitive sentiment as an anti-free rider psychological device. *Evolution and Human Behavior, 23,* 203–231. doi:10.1016/S1090-5138(01)00093-9

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). *Contrasts and effect sizes in behavioral research: A correlational approach.* Cambridge, England: Cambridge University Press.

Schneider, D. J. (2004). *The psychology of stereotyping.* New York, NY: Guilford Press.

Seymour, B., Singer, T., & Dolan, R. (2007). The neurobiology of punishment. *Nature Reviews Neuroscience, 8,* 300–311. doi:10.1038/nrn2119

Sherman, S. J., Castelli, L., & Hamilton, D. L. (2002). The spontaneous use of a group typology as an organizing principle in memory. *Journal of Personality and Social Psychology, 82,* 328–342. doi:10.1037/0022-3514.82.3.328

Skowronski, J. J., & Carlston, D. E. (1987). Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases. *Journal of Personality and Social Psychology, 52,* 689–699. doi:10.1037/0022-3514.52.4.689

Stangor, C., Lynch, L., Duan, C. M., & Glass, B. (1992). Categorization of individuals on the basis of multiple social features. *Journal of Personality and Social Psychology, 62,* 207–218. doi:10.1037/0022-3514.62.2.207

Stone, V. E., Cosmides, L., Tooby, J., Kroll, N., & Knight, R. T. (2002). Selective impairment of reasoning about social exchange in a patient with bilateral limbic system damage. *Proceedings of the National Academy of Sciences of the United States of America, 99,* 11531–11536. doi:10.1073/pnas.122352699

Sugiyama, L. S. (2004). Illness, injury, and disability among Shiwiar forager-horticulturists: Implications of health-risk buffering for the evolution of human life history. *American Journal of Physical Anthropology, 123,* 371–389. doi:10.1002/ajpa.10325

Sugiyama, L. S., Tooby, J., & Cosmides, L. (2002). Cross-cultural evidence of cognitive adaptations for social exchange among the Shiwiar of Ecuadorian Amazonia. *Proceedings of the National Academy of Sciences of the United States of America, 99,* 11537–11542. doi:10.1073/pnas.122352999

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science, 18,* 580–586. doi:10.1111/j.1467-9280.2007.01943.x

Taggar, S., & Neubert, M. J. (2008). A cognitive (attributions)-emotion model of observer reactions to free-riding poor performers. *Journal of Business and Psychology, 22,* 167–177. doi:10.1007/s10869-008-9058-0

Talmy, L. (2000). *Toward a cognitive semantics: Vol. 1. Concept structuring systems.* Cambridge, MA: MIT Press.

Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology, 36,* 778–793. doi:10.1037/0022-3514.36.7.778

Tazelaar, M. J. A., Van Lange, P. A. M., & Ouwerkerk, J. W. (2004). How to cope with "noise" in social dilemmas: The benefits of communication. *Journal of Personality and Social Psychology, 87,* 845–859. doi:10.1037/0022-3514.87.6.845

Tooby, J., & Cosmides, L. (1988, April). *The evolution of war and its cognitive foundations* (Institute for Evolutionary Studies Technical Report No. 88–1). Paper presented at the Evolution and Human Behavior Meetings, Ann Arbor, MI.

Tooby, J., & Cosmides, L. (1990). On the universality of human nature and the uniqueness of the individual: The role of genetics and adaptation. *Journal of Personality, 58,* 17–67. doi:10.1111/j.1467-6494.1990.tb00907.x

Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J. H. Barkow, L., Cosmides & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 19–136). New York, NY: Oxford University Press.

Tooby, J., Cosmides, L., & Barrett, H. C. (2005). Resolving the debate on innate ideas: Learnability constraints and the evolved interpenetration of motivational and conceptual functions. In P. Carruthers, S. Laurence, &

S. Stitch (Eds.), *The innate mind: Structure and content* (pp. 305–337). New York, NY: Oxford University Press.

Tooby, J., Cosmides, L., & Price, M. E. (2006). Cognitive adaptations for *n*-person exchange: The evolutionary roots of organizational behavior. *Managerial and Decision Economics, 27,* 103–129. doi:10.1002/mde.1287

Van Lange, P. A. M., Liebrand, W. B. G., Messick, D. M., & Wilke, H. A. M. (1992). Introduction and literature review. In W. B. G. Liebrand, D. M. Messick, & H. A. M. Wilke (Eds.), *Social dilemmas* (pp. 3–28). Oxford, England: Pergamon Press.

Weiner, B. (1993). On sin versus sickness: A theory of perceived responsibility and social motivation. *American Psychologist, 48,* 957–965. doi:10.1037/0003-066X.48.9.957

Weiner, B., Perry, R. P., & Magnusson, J. (1988). An attributional analysis of reactions to stigmas. *Journal of Personality and Social Psychology, 55,* 738–748. doi:10.1037/0022-3514.55.5.738

Williams, G. C. (1966). *Adaptation and natural selection.* Princeton, NJ: Princeton University Press.

Wilson, E. O. (1974). *The insect societies.* Cambridge, MA: Harvard University Press.

Wilson, M. L., & Wrangham, R. W. (2003). Intergroup relations in chimpanzees. *Annual Review of Anthropology, 32,* 363–392. doi:10.1146/annurev.anthro.32.061002.120046

Wojciszke, B., Brycz, H., & Borkenau, P. (1993). Effects of information content and evaluative extremity on positivity and negativity biases. *Journal of Personality and Social Psychology, 64,* 327–335. doi:10.1037/0022-3514.64.3.327

Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology, 51,* 110–116. doi:10.1037/0022-3514.51.1.110

Yamagishi, T., Terai, S., Kiyonari, T., Mifune, N., & Kanazawa, S. (2007). The social exchange heuristic: Managing errors in social exchange. *Rationality and Society, 19,* 259–291. doi:10.1177/1043463107080449

Received January 16, 2011
Revision received December 13, 2011
Accepted December 19, 2011 ■