Over the past several decades, equity theory (Adams, 1963, 1965)—which predicts that members of organizations will strive to be rewarded equitably, rather than under- or overrewarded—has established itself as a foundational theory in organizational behavior and management. While equity theory has been highly successful in many respects, it has also been criticized for various reasons. The purpose of this chapter is to suggest that one of these criticisms—equity theory’s unreliability in situations of overreward—could be met more effectively if organizational scholars reformulated equity theory from an evolutionary psychological perspective. Equity theory would gain increased predictive power from this reformulation, because it would become better able to address the tendency of members to seek overreward—a tendency that has been recognized for decades in the social sciences as the “free rider problem” (Olson, 1965). An improvement in equity theory’s ability to address the free rider problem would significantly enhance the theory’s usefulness because the free rider problem is not just a trivial detail that equity theory overlooks. On the contrary, overcoming this problem is central to the health and success of any organization (Olson, 1965; Albanese & Van Fleet, 1985; Ostrom, Walker, & Gardner, 1992), and the reasons for this problem’s centrality are thrown into relief by the evolutionary psychological perspective (Price, Cosmides, & Tooby, 2002; Tooby, Cosmides, & Price, 2006). An evolutionarily updated equity theory would help to explain both why free rider problems commonly occur at all organizational levels (including managerial levels, as will be emphasized at the conclusion of this chapter), and also how solving these problems can be a key to improving organizational productivity.
Equity Theory: Summary and Predictions

The core elements of classic equity theory (Adams, 1963, 1965), hereafter CET, are as follows. Members engage in a process of social exchange with their organization in which they offer inputs for outcomes. Inputs are defined as the member's perceived contributions to the organization (e.g., work effort, skills), and outputs are things that the member receives from the organization (e.g., salary, social status). A member (referred to by Adams as “Person”) compares his or her own outcome-to-input ratio with those of a referent individual or group (called “Other”). Other is usually another individual; for example, a coworker who performs a similar role as Person, but Other may also be Person in another context such as a previous job. Consistent with the theory of cognitive dissonance (Festinger, 1957), members should become distressed when their outcome-to-input ratio is unequal to those of others, and should attempt to rectify this inequity. For example, imagine that Person perceives himself to have an outcome-to-input ratio of 1:1; for purposes of illustration, let's say that he receives one unit of payment in exchange for one full day's work. If Person perceives that Other's ratio is 2:1—that is, that Other receives two units of payment in exchange for one full day's work—then Person will perceive himself to be relatively underrewarded, and will attempt to reduce this discrepancy. One way he could accomplish this would be by acting to alter the values of his own inputs and outcomes and those of Other. That is, he could strive to increase his own outcome, decrease his own input, increase Other's input, or decrease Other's outcome. Other options would be to distort cognitively the values of any of these inputs or outcomes (for example, he could perceive his own inputs to be lower), change to a different referent Other whose outcome-to-input ratio was closer to his own, or leave the field entirely (for example, by quitting his job).

In situations in which Person perceives himself to be overrewarded rather than underrewarded compared to Other, CET's predictions remain fundamentally the same. If Person again perceives his own outcome-to-input ratio to be 1:1, but now perceives Other’s ratio to be 1:2, then he will perceive himself to be relatively overrewarded. He should again feel distress, and be motivated to relieve it by promoting equity via the same methods he would use in the case of underreward. One feature that Adams does suggest should differ between situations of underreward and overreward is the negative emotion that causes distress in Person and thus motivates him to promote equity: Person should experience anger when underrewarded and guilt when overrewarded.
Assessing the Predictiveness of Equity Theory

In evoking the concept of the relative outcome-to-input ratio, CET sketches the psychological computational processes that members must experience in order to judge the fairness of their exchange relationship with their organization. CET is remarkably parsimonious for a social science theory, and it has generally proved useful for predicting behavior. A relatively objective assessment of CET’s value is provided by Miner (2003), who had 71 organizational behavior scholars rate the importance of 73 organizational behavior theories on a 7-point scale. Ratings ranged from 2.71 to 5.97, and CET scored a 5.93, for third place overall. Miner also included CET in his list of the 25 organizational behavior theories with the highest scientific validity. The numerous research reviews that have been published on CET (e.g., Greenberg, 1990; Ambrose & Kulick, 1999; Colquitt, Conlon, Wesson, Porter, & Ng, 2001; Mowday & Colwell, 2003; Colquitt, Greenberg, & Zapata-Phelan, 2005; Bolino & Turnley, 2008) generally echo the sentiment that CET has been one of the more important and successful theories in organizational behavior. However, these reviews also acknowledge that CET has attracted a number of consistent criticisms. Bolino & Turley (2008) summarize these reviews in their list of the four most common criticisms of CET: First, CET is underspecified, particularly in terms of predicting reactions to inequity; second, it only considers equity rules, as opposed to other kinds of distribution rules (e.g., equality, need); third, it considers only distributive justice, to the exclusion of procedural and interactional justice; and fourth, it is significantly more accurate in predicting behavior in cases of underreward than in cases of overreward. While all of these criticisms are important, they each target a distinct problem that can be addressed individually, and the remainder of this chapter will focus on addressing the last one on this list.

CET’s reduced predictiveness in situations of overreward has always been a bit awkward for the theory. As noted above, CET is rooted in cognitive dissonance theory, and since situations of underreward are just as dissonant as those of overreward, there is no principled reason from this perspective why the former should be more objectionable than the latter. Nevertheless, Adams did state that situations of underreward should be less tolerable than those of overreward. This prediction “is derived from two assumptions: first ... that the threshold for the perception of inequity is higher when Person is overrewarded than when he is underrewarded; secondly ... that Person is motivated to minimize his costs and to maximize his gains” (Adams, 1965, p. 284). In these sentences, Adams establishes what Lakatos (1978) would consider to be an auxiliary or ad hoc hypothesis, as opposed to a core hypothesis, of CET. This auxiliary hypothesis helps CET accommodate the relative tolerability of overreward,
but it is by definition difficult to integrate an auxiliary hypothesis into a central theory, and so some unresolved questions remain. For example, is there any principled rationale behind the first assumption above, that “the threshold for the perception of inequity is higher when Person is overrewarded”? Perhaps it could be explained in terms of the second assumption (“Person is motivated to minimize his costs and to maximize his gains”); however, if members are motivated to minimize costs and maximize gains, why does CET theory predict in the first place that they should be motivated to avoid (rather than seek) overreward situations?

In practice, when tests of CET have shown members to be more tolerant of overreward than underreward, Adams’s auxiliary hypothesis has been implicitly or explicitly invoked, often expressed in terms of egocentric bias (e.g., Greenberg, 1981, 1983; Loewenstein, Thompson, & Bazerman, 1989; Lerner, Somers, Reid, Chiriboga, & Tierney, 1991; Thompson & Loewenstein, 1992; Diekmann, Samuels, Ross, & Bazerman, 1997; Leung, Tong, & Ho, 2004). As a result, CET plus egocentric bias (in cases of overreward) has been established as a useful theory of how members perceive their exchange relationships with organizations. While CET loses some symmetry and simplicity from this bolting-on of egocentric bias, it gains a significant amount of predictive power. This seems like a reasonable trade-off, because prediction is, after all, the scientific bottom line. To paraphrase Einstein (1934), a theory should be as simple as it can be, without sacrificing any predictive power.

However, although the egocentric bias auxiliary hypothesis does help CET navigate through situations of overreward, it may not go far enough. CET may have an underlying flaw in its foundation, of which its struggles with overreward are merely symptomatic. Despite CET being supplemented with egocentric bias, one of its fundamental assumptions remains intact: members prefer equity and are averse to inequity. They may be more averse to underreward than to overreward, but nevertheless, they prefer equity over both. Although this core hypothesis does contain some important elements of truth about human behavior in organizations, it also contains a serious error (Olson, 1965; Ostrom et al., 1992), an error that is avoided by taking an evolutionary psychological perspective on cooperation in groups (Price et al., 2002; Tooby et al., 2006).

**Evolutionary Equity Theory (EET)**

Organizations are cooperative groups, often large and complex enough to contain many smaller nested cooperative groups (departments, teams, committees, etc.). Evolutionary psychology makes a set of general
predictions about behavior in cooperative groups that are in some ways consistent with the predictions of CET, and in other ways quite different. These differences may help resolve some of the issues that have challenged CET. Because the simple theory that will be sketched here has some fundamental aspects in common with CET, it is perhaps best regarded as an evolutionarily informed update of CET, as opposed to a radical new theory. For that reason, the updated theory will be referred to as evolutionary equity theory (hereafter EET). Before presenting EET, it would help to review briefly some of the basic assumptions of evolutionary psychology.

From an evolutionary psychological perspective (Tooby & Cosmides, 1992, 2005), human minds are composed of information-processing mechanisms (adaptations) that were designed by natural selection primarily to promote the survival and reproduction of the individual in ancestral environments. Psychological adaptations can be thought of as “if, then” devices, which process informational inputs from the environment and respond by producing the psychological and behavioral outputs that would have been, on average, the adaptive response to such information in the ancestral past (e.g., IF you see a snake, THEN experience fear and back off). Because behavior is governed by these psychological adaptations, people behave in ways that, in ancestral environments, allowed them to acquire personal adaptive advantages and overcome personal adaptive problems. Contributing to a group cooperative effort would have provided ancestral humans with valuable opportunities to engage in contribution-for-benefit exchange (analogous to CET’s input-for-outcome exchange). However, it would also have presented them with a unique set of risky adaptive problems and tempting opportunities for adaptive advantage.

The Free Rider Advantage, the Exploitation Problem, and the Consequences Problem

Ethnographic and archaeological evidence suggests that in ancestral environments, humans formed cooperative groups for the purpose of, for example, foraging for food, waging war, building structures, defending against predators, and developing political alliances (Lee & DeVore, 1968; Kelly, 1995). EET is based on consideration of the kinds of adaptive advantages and problems that would have been encountered by individual members of these cooperative groups.

In typical group cooperative interactions, members who reap the highest personal advantages, at least in the short term, are those whose benefit-to-contribution ratios are the highest. This principle has been studied for decades in social science, usually in the context of collective actions for the production of public goods (Ölson, 1965; Hardin 1968). When the benefit produced is a public good (i.e., a resource to which all members have equal access), all members benefit equally, so those who contribute the
least to production (free riders) have the highest benefit-to-contribution ratios. CET is usually applied to situations involving private goods (e.g., salaries, which can differ across individuals), but private goods situations can involve free rider problems as well, because benefit-to-contribution ratios can vary in private goods situations just as they can in public goods situations (an assumption on which CET is, of course, founded). Because natural selection is influenced primarily by the survival and reproduction of individuals as opposed to groups (Darwin, 1859; Williams, 1966; Tooby & Cosmides, 1992), it would have initially rewarded group members who preferred to have relatively high benefit-to-contribution ratios (Price, 2006a; Tooby et al., 2006), an opportunity which can be referred to as the free rider advantage.

In pursuing their advantage, free riders would have presented an immediate adaptive problem for more cooperative co-members (i.e., members with lower benefit-to-contribution ratios): in order to contribute adaptively, members would have had to ensure that their own benefit-to-contribution ratios did not disadvantage them relative to free riders. A standard finding of formal models of the evolution of cooperation is that free riders, if unchecked, will eventually exploit cooperators to extinction (e.g., Boyd & Richerson, 1988; Henrich, 2004). Therefore, selection would have disfavored members who were apathetic about being exploited by free riders, and would have favored those who were motivated to avoid such exploitation (Price, 2006b; Price et al., 2002). The danger that free riders pose to cooperators can be labeled the exploitation problem.

In order to solve the exploitation problem, evolution designed people to be averse to having benefit-to-contribution ratios that are lower than those of co-members. When members perceived that they were being exploited, the anger they experienced would have motivated them to try and reduce their disadvantages relative to free riders. They could have done so by striving to increase their own benefits and/or decrease their own contributions. However, they could also have directed this anger toward the free riders themselves, and tried to lower free riders’ benefits and/or increase free riders’ contributions. For example, by directing punishment strategically at free riders, they might have been able to lower free riders’ net benefits while at the same time inducing them to contribute more (Price et al., 2002; Tooby et al., 2006). The anger of exploited members would also have led them to ostracize free riders from future cooperative interactions; the more a member acquired a reputation for free riding, the fewer exchange opportunities he would have had. In summary, because free riders presented an adaptive problem for co-members, they risked punishment and ostracism; and the more they acquired a reputation for free riding, the greater these costs would have been. Free riders thus faced a potential consequences problem.
According to EET, the evolutionary cat-and-mouse dynamic between cooperators and free riders would have designed the set of psychological mechanisms that governs the behavior of the average member of a modern human group. Because of the exploitation problem, members object when they perceive co-members’ benefit-to-contribution ratios to be lower than their own. At the same time, because of the free rider advantage, members are more likely to seek relatively high benefit-to-contribution ratios when they perceive the social risks of free riding to be lower. Finally, when these social risks are perceived to be higher, members are more averse to having relatively high benefit-to-contribution ratios, because they strive to avoid the consequences problem.

Comparing the Predictions of CET and EET

CET and EET agree that in order for a member to assess the extent of his advantage or disadvantage relative to others, he needs to compare his own benefit-to-contribution ratio to those of others. Further, the two theories overlap to a large extent in their predictions about how that member should react if he perceives himself to be relatively underrewarded. If Member A perceives his own benefit-to-contribution ratio to be 1:1, and Member B’s to be 2:1, then both theories predict that A will experience anger and be motivated to equalize the ratios. The main difference in the theories is that A’s motivation to achieve equity is explained by CET in terms of aversion to cognitive dissonance and by EET in terms of aversion to the exploitation problem.

Diverging Predictions in the Case of Overreward (Free Riding)

In the case of overreward, the predictions of CET and EET begin to diverge. If Member A perceives that his own benefit-to-contribution ratio is 1:1, and that B’s ratio is 1:2, then CET predicts that A should be motivated by guilt to equalize the ratios; however, A’s motivation threshold for seeking equity in this situation will be higher than it is in the case of underreward. CET explains A’s motivation to achieve equity again in terms of A’s desire to avoid cognitive dissonance, and explains A’s higher threshold for pursuing equity in terms of egocentric bias. CET does not predict that A will be motivated to pursue or maintain situations of overreward for himself, or that A’s appetite for overreward will be influenced by his perception of the extent to which free riding entails social risks. EET, in contrast, makes both of these predictions.
According to EET, the free rider advantage will make being overrewarded seem appealing to some extent, but this extent will be limited by the degree to which free riding makes one vulnerable to the consequences problem. Thus in order for Member A to decide how to react to being overrewarded, he will need to assess the likelihood that his free riding will be monitored and discovered by others, and evaluate others’ willingness and ability to ostracize or punish him in retaliation for free riding. For example, others’ willingness to punish may depend on their anger levels, and others’ ability to punish may depend on the probability that they will interact in the future with A, or on their formidability relative to A. When detection and punishment are more likely, the consequences problem will be more salient, and A’s preference for equity over overreward will increase. When the consequences problem is less relevant, A will more likely strive to maintain or even increase the extent of his free riding.

**The Collapse of Cooperation When Free Riding Goes Unpunished**

Because CET and EET make different predictions about how members react to situations of overreward, they also make different assessments about the extent to which free riding threatens the health of organizations. To a greater extent than EET, CET predicts that inequity in organizations is an automatically, efficiently self-correcting phenomenon. For example, imagine an organization in which unfairness is rampant: half the members perceive themselves to be underrewarded, and half perceive themselves to be overrewarded. Both halves should be motivated to seek equity: anger should motivate the underrewarded half to (e.g.,) work less hard, and guilt should motivate the overrewarded half to (e.g.,) work harder, until equity emerges as if from organic self-organization. With CET’s egocentric bias auxiliary hypothesis, the picture loses some of its idealized symmetry, because the overrewarded half will need to perceive a relatively high level of overreward before they are motivated to strive for equity. Nevertheless, CET assumes a relatively high level of equitable self-correction in organizations, because it proposes that equity-oriented influences (the emotions of anger and guilt) are located entirely within the psychologies of members themselves, and are experienced automatically upon the perception of inequity.

EET, on the other hand, emphasizes that organizations will often not spontaneously and efficiently self-correct towards greater equity. In the half-underrewarded, half-overrewarded organization, EET predicts that the underrewarded half will indeed be motivated to increase equity; they will respond to the exploitation problem by (e.g.,) working less hard. However, the motivation of the overrewarded to increase equity will depend on how they answer questions related to the consequences problem: Has anyone noticed my free riding and will they notice it in the future? If so, will they have any desire or power to punish or ostracize me? The threat posed by the free rider problem to organizations has been recognized in the management literature for years, particularly in prominent reviews by Albanese & Van Fleet (1985) and Kidwell & Bennett (1993). Outside of the field of management, there is a huge literature on the free rider problem, going back at least as far as the classic works by Olson (1965) and Hardin (1968). Some more recent documenters of free riding’s negative relationship with organizational productivity have been the political scientist Elinor Ostrom (e.g., Ostrom, 1990, 2000, 2009; Ostrom et al., 1992) and the numerous experimental economists who are cited below. Much of this research on the free rider problem has focused on the variable explanations of others’ willingness to monitor and punish free riding, such as the probability that others will detect A’s free riding, and others’ ability to ostracize or punish him in retaliation for free riding. For example, others’ willingness to punish may depend on their anger levels, and others’ ability to punish may depend on the probability that they will interact in the future with A, or on their formidability relative to A. When detection and punishment are more likely, the consequences problem will be more salient, and A’s preference for equity over overreward will increase. When the consequences problem is less relevant, A will more likely strive to maintain or even increase the extent of his free riding.
future? If so, will they have any desire or power to punish or ostracize me? Do I perceive these potential costs to be lower than the advantage of free riding? When the answers to such questions are more likely to be “no,” free riding is more likely to persist or increase, and the motivation of the underrewarded will therefore continue to decay.

Unchecked free riding may ultimately cause cooperation in the organization to unravel completely, as members compete with one another in pursuit of the free rider advantage. The worse free riding gets, the more one has to free ride in order to maintain a free rider advantage (i.e., in order to be advantaged over someone with a very high benefit-to-contribution ratio, your ratio must be higher still). The result can be a negative feedback loop that is disastrous for the organization as a whole: as free riders become increasingly exploitative, their co-members increasingly lose motivation, and overall productivity continuously declines. EET suggests that the way to avoid this collapse is to increase the salience of the consequences problem for would-be free riders. Unless free riding is perceived as being sufficiently socially risky, a significant fraction of members will seek overreward. Their sense of guilt will often not in itself be sufficient to eliminate this temptation, contrary to the implications of CET.

Existing Evidence for Testing EET Against CET

As noted above, the predictions of CET and EET tend to converge in situations of underreward, and to diverge in situations of overreward. The emphasis here will be on evidence pertaining to the latter type of situation, because this kind of evidence permits an assessment of which theory is more predictive. In general, this evidence suggests that in many different types of groups and in many different cultures, members are often motivated to seek benefit-to-contribution ratios that endow them with a free rider advantage, unless their pursuit of this advantage is thwarted by the consequences problem.

The threat posed by the free rider problem to organizations has been recognized in the management literature for years, particularly in prominent reviews by Albanese & Van Fleet (1985) and Kidwell & Bennett (1993). Outside of the field of management, there is a huge literature on the free rider problem, going back at least as far as the classic works by Olson (1965) and Hardin (1968). Some more recent documenters of free riding’s negative relationship with organizational productivity have been the political scientist Elinor Ostrom (e.g., Ostrom, 1990, 2000, 2009; Ostrom et al., 1992) and the numerous experimental economists who are cited below. Much of this research on the free rider problem has focused on
public goods situations, in which a group cooperative effort produces a resource that is equally shared by each group member. As noted above, because public goods are distributed equally among members regardless of contribution, all members end up with equal outcomes. Public goods distribution systems are thus, in an important sense, different from the private goods systems envisioned by CET. Nevertheless, CET should be as capable of making predictions in public goods systems as is it in equity-based systems; it simply has to consider that in public goods systems, all group members have equal outcomes. In public goods situations, then, CET predicts that members will strive to equalize their inputs with those of co-members, in order to match their equal outcomes. In contrast, EET predicts that in public goods situations, the less members perceive that their free riding will entail negative social consequences, the more they will strive for inputs that are lower than those of co-members.

**Free Riders in Public Goods Games**

The public goods game, a standard paradigm from experimental economics, has produced evidence that is especially illuminating and relevant for testing CET versus EET. A public goods game is essentially an experimental collective action, in which members of small (usually four-person) groups are given an endowment of real money and have the opportunity to contribute toward the production of a public good. At the end of the game, whatever public good has been produced by the group is redistributed equally among all members, regardless of the extent to which a member contributed toward its production. Thus, members who contribute the least reap the highest net benefits. The game is designed to create a free rider advantage, and so it presents an ideal context for testing the CET prediction that people will shun rewards that are proportionally greater than those received by others: if players are in fact often motivated to seek relative overreward, this game gives them the opportunity to do so.

Results from public goods games consistently show that many players are indeed motivated to seek overreward, and that this free riding is a chronic problem in the maintenance of group productivity. The standard pattern of contributing is as follows. Contributing is highest at the outset of the game; players apparently realize that cooperation is potentially productive, and begin by contributing on average about 40 to 60% of their endowment. However, some players contribute less than others, and continue to undercontribute as the game progresses, even after they have received feedback about the relative meagerness of their contributions. In response to this free riding, the higher contributors lower their own contributions, in order to mitigate their own exploitation. Thus, as the result of persistent free riding, average contributions dwindle as the game progresses, and group productivity decays gradually over time (e.g., Andreoni, 1988;
Free Riders as a Blind Spot of Equity Theory


However, even if free riding is a persistent problem in these games, that pattern does not mean that the majority of players are motivated to free ride. It could just be a few bad apples who ruin the production effort for everyone else, by behaving exploitatively and thus provoking lowered contributions from annoyed co-players. What do the data suggest about the percentage of players who free ride? The answer to this question is actually a bit more complicated than it may first appear. Researchers usually define a free rider as a player who consistently contributes below the mean co-player contribution, while those who contribute at or above this mean are considered cooperators. Studies that have attempted to classify players into different categories have sometimes concluded that free riders are in the minority (e.g., Fischbacher et al., 2001; Kurzban & Houser, 2005; Gunnthorsdottir, Houser, & McCabe, 2007). However, how far below the mean co-player contribution a player can go before he or she is considered a free rider rather than a cooperator, is necessarily arbitrary to some degree (Gunnthorsdottir et al., 2007). Further, the percentage of players who free ride does not remain constant across all public goods game conditions. Indeed, a major goal of public goods game research is to discover the conditions that influence the extent of free riding. So in addressing the issue of the prevalence of free riding, for now we will just note that it is common enough to damage group productivity significantly, and that its prevalence depends in large part on the conditions under which the game is played. The most important of these conditions is reviewed below.

Punishment as a Deterrent to Free Riding

Punishment of free riders is the most basic and effective solution to the free rider problem that experimental economists have investigated. That is not to say that punishment is the only possible solution; other solutions that have been proposed and explored include, for example, allowing greater verbal communication among players (e.g., Ostrom et al., 1992; Bochet, Page, & Putterman, 2006), manipulating the extent to which players can develop stable reputations (e.g., Andreoni, 1988; Croson, 1996), and presenting players with cues suggesting that they are being monitored (Haley & Fessler, 2005; Burnham & Hare, 2007). Punishment is not even the only kind of social sanction that may help prevent free riding; positive sanctions (rewards) may be effective as well, although studies of this effectiveness have produced mixed conclusions (Andreoni, Harbaugh, & Vesterlund, 2003; Sefton, Shupp, & Walker, 2007). However, punishment’s power to deter free riding has been studied relatively extensively and documented relatively convincingly, so in consideration of space constraints,
it will be the focus here. It should be noted that much of this research on punishment has been motivated explicitly by evolutionary considerations (e.g., Barclay, 2006; Boyd, Gintis, Bowles, & Richerson, 2003; Price et al., 2002; Price, 2005). This is true because an evolutionary approach, with its emphasis on the free rider problem and the consequences problem, suggests that a consequence like punishment, when severe enough to negate the advantage of free riding, will be an effective solution to the free rider problem. The idea of punishment as a solution to the free rider problem is thus highly compatible with EET.

Through punishment, public goods game players are given the opportunity to retaliate against undercontributing co-players by imposing some sort of cost on them, often in the form of a monetary fine. Fining interactions typically involve above-mean contributors imposing endowment-reducing fines on below-mean contributors, so these fines can appropriately be considered punishment of free riders (Ostrom et al., 1992; Fehr & Gächter, 2000, 2002; Nikiforakis, 2008; however, fining of above-mean contributors also occurs, often by individuals who have themselves been fined previously, and who presumably want revenge on their perceived punishers). The imposition of these fines is usually made to be costly, and players who choose to punish other players must use their own funds to pay for it. Despite this costliness, many players willingly pay to punish free riders, and punishment of free riders generally turns out to be a highly effective way of keeping contributions high, and so of staving off the decay of cooperation that usually characterizes public good games (e.g., Fehr & Gächter, 2000, 2002; Masclet, Noussair, Tucker, & Villeval, 2003; Sefton et al., 2007; Yamagishi, 1986; Ostrom et al., 1992; Nikiforakis, 2008). Because punishment prevents the cooperative interaction from unraveling due to free riding, the threat of being punished seems to motivate higher contributions from players who would otherwise free ride. Punishment is thus group-beneficial: contributions are higher when punishment is possible, so each group member’s share is larger. Because punishment is group-beneficial and costly to impose, it is sometimes referred to as altruistic punishment (e.g., Fehr & Gächter, 2002; Boyd et al., 2003; Barclay, 2006), and seen as a kind of second-order contribution to the public good (Yamagishi, 1986).

Besides monetary sanctions, nonmonetary social sanctions also seem to be fairly effective ways of eliciting higher contributions from would-be free riders in public goods games. For example, Masclet et al. (2003) allowed standard monetary sanctions in some rounds and nonmonetary sanctions in others. In the nonmonetary rounds, players could sanction free riders by assigning points, which indicated “disapproval” of the free riding but had no monetary consequences. These nonmonetary sanctions affected contributions positively, although less positively than did the monetary sanctions. Ostracism is another kind of social sanction that can help prevent free riding. A public goods game study by Cinyabuguma,
Page, & Putterman (2005), which included a baseline treatment with no sanctioning and an expulsion treatment in which players could vote to expel free riders from their group, found that contributions were significantly higher in the expulsion treatment. Further, in public goods games where players are given access to information about the past behavior of their potential co-players, and then allowed to choose their co-players, free riders tend to be left out in the cold. Higher contributors tend to choose one another and end up forming the most cooperative and productive groups, while free riders are left behind to form less-productive groups (Ehrhart & Keser, 1999; Sheldon, Sheldon, & Osbaldiston, 2000; Page et al., 2005; Barclay & Willer, 2007). Thus, partner choice helps solve the free rider problem because it allows cooperators to ostracize free riders, and to thus put themselves in situations where they can contribute without fear of exploitation.

The evidence suggests that the free rider problem, as well as punishment’s effectiveness as a solution to this problem, are both cross-culturally universal. This universality has been demonstrated not only in the above-cited public goods games with punishment played throughout North America, Europe, and Japan, but also via methods besides those of experimental economics. For example, comparative studies conducted among Californians and indigenous Shuar in the Ecuadorian Amazon suggest that in both cultures, punitive sentiment toward free riders is highest among those who are themselves the most willing to contribute to collective actions (Price et al., 2002; Price, 2005). And Ostrom, who has probably conducted more cross-cultural research on collective action than anyone else, notes that “in all known self-organized resource governance regimes that have survived for multiple generations, participants invest resources in monitoring and sanctioning the actions of each other so as to reduce the probability of free riding” (Ostrom, 2000, p. 138). The universality of free riding, and of free rider punishment, is consistent with the idea that these behaviors are the products of species-typical psychological mechanisms, and thus consistent with EET.

Finally, a crucial caveat about punishment must be stated: if administered unnecessarily or for the wrong reasons, punishment can backfire and actually have a negative effect on productivity in collective actions. There is a risk that external incentives like punishments will crowd out voluntary cooperativeness among participants (Titmuss, 1970; Vollan, 2008). Those who are already disposed toward contributing, and who do not need an external incentive to induce them to do so, may regard punishment as unwarranted coercion and contribute less under the threat of punishment than they would have in the absence of this threat. Punishment can also be detrimental if it is administered in a way that seems unfair or illegitimate; for example, in a way that suggests overly selfish motives on the part of the punisher (Fehr & Rockenbach, 2002).
Is the Above Evidence More Consistent with CET or EET?

One general behavior pattern, present in the cross-cultural evidence from the public goods game and other studies, is consistent with both CET and EET: group members are averse to being underrewarded. Relatively high contributors to public goods games, for example, will voluntarily act to punish and avoid co-players who have higher benefit-to-contribution ratios (free riders), and will reduce their own contributions if they become trapped in a group with such co-players. However, another major behavior pattern observed in the above studies seems compatible with EET but not with CET: the less of a risk there is that free riding will be punished in some way, the more likely group members will be to actively seek relative over-reward. This pattern is consistent with EET, which sees pursuit of the free rider advantage as a predictable outcome in the absence of the consequences problem. However, active free riding seems to undermine CET since CET predicts aversion to overreward, even if this aversion should be reduced relative to the extent to which it occurs in underreward situations.

As noted above, social scientists have long been aware of the theory that free riding poses a problem for group cooperative interactions. Indeed, the pioneering theoretical statements about this problem (Olson, 1965; Hardin, 1968) are only slightly younger than CET itself. Literature on the free riding problem has also been reviewed and discussed in mainstream management publications for years (e.g., Albanese & Van Fleet, 1985; Kidwell & Bennett, 1993). However despite the visibility of the free rider problem to management and organizational behavior audiences, little effort has been made to address the incompatibility of the free rider problem with CET. This oversight could be corrected fairly easily, by replacing CET’s traditional foundation, rooted in cognitive dissonance theory, with a foundation that is based in evolutionary theory and that identifies the adaptive advantages and problems that members of cooperative groups would have faced in ancestral environments.

Applying EET in Organizations

As mentioned above, there have been notable efforts to raise the awareness of management scholars about the relevance of the free riding problem to organizational productivity (Albanese & Van Fleet, 1985; Kidwell & Bennett, 1993). Nevertheless, this problem has received relatively little attention compared to the large amount of research that has been inspired by, and has often been interpreted as being largely supportive of, CET.
Consequently, many researchers may be biased in favor of CET’s view that inequity in organizations is to a large extent self-correcting. However, if EET is correct, then this CET perspective overestimates the extent to which overreward situations will self-correct, and underestimates the potential of the free riding tendency to infect an organization. Consequently, advice to managers that is rooted in CET, without due regard for the kinds of corrections offered by EET, could fail to diagnose one of the most significant threats to sustained productivity in cooperative groups.

Managers should be made aware of the universal pervasiveness of the free riding problem and of the corrosive effects it can potentially have at all organizational levels. But if free riding is not a problem that will just solve itself, then what can managers do to solve it? As noted above, various solutions to the problem have been proposed and explored, but the effectiveness of punishment has been especially well studied and well documented, so it will be the focus here. More specifically, the focus will be on what punishments for free riding should be explicitly designed to accomplish, and what they should be designed to avoid.

The ultimate goal of free rider punishment should be not to harm the guilty, but rather to maintain the health of the organization as a whole, by creating an environment in which people can contribute without fear of exploitation. If some members of an organization perceive that they are being exploited by others, it may be necessary for management to step in and demonstrate that taking advantage of contributors will not be tolerated. This approach to solving the exploitation problem for high-contributing workers does not mean that managers must adopt an overly cynical philosophy of human nature. It would be inaccurate, for example, for managers to assume that all of their employees are striving to exploit one another; research suggests that there is considerable variation in the extent to which individuals are motivated to free ride, and that substantial numbers of people may lack this motivation (Fischbacher et al., 2001; Kurzban & Houser, 2005; Gunnthorsdottir et al., 2007). Nevertheless, managers must expect that free riding strategies may well be pursued commonly and vigorously in organizations in which they are permitted to flourish. To expect otherwise could lead to results that are demoralizing for high-contributing employees and thus dangerous for the organization.

However, although managers should not underestimate the potential of free riding to become a problem, it is equally important that they avoid overestimating this potential, and that they strive to ensure that punishments for free riding are necessary and fair. Free riding may sometimes be perceived to be a larger problem than it actually is, and as noted above, punishments that are perceived as unnecessary or as overkill can crowd out genuine desires to contribute voluntarily. For example, for noncontribution to be true free riding, it must be intentional, and not be due to (e.g.,) illness or injury (Price, 2006a). Punishments directed at unintentional
noncontribution will generally be perceived as unfair and demoralizing. It should be clear to everyone in the organization that punishments are directed only toward individuals who have intentionally pursued above-mean benefit-for-contribution ratios, and that the purpose of the punishment is to solve the exploitation problem for higher contributors and to thus promote the health of the organization as a whole. Punishments will be more likely to backfire if they are perceived as having additional, more selfish motives (e.g., if they are seen as a way for management to extract an amount of effort out of workers that is disproportionately large, given the extent of the workers' free riding).

What About When the Worst Free Riders Are the Managers Themselves?

The fact that free riding is a very real problem in organizations (Albanese & Van Fleet, 1985; Kidwell & Bennett, 1993; Ostrom, 2009), even at the very highest levels of organizational structure, has been illustrated repeatedly and vividly over the past several months of the global financial crisis, as leaders of various failed institutions have been accused of happily accepting “reward for failure.” A dramatic recent example is Sir Fred Goodwin, who led the Royal Bank of Scotland to the largest annual corporate loss in U.K. history (£24.1 billion), before stepping down to receive an annual pension of about £700,000 (Treanor, 2009). From the perspective of the typical observer, it seems as though Sir Fred has receiving a very large benefit in exchange for a massively negative contribution. This situation has inspired outrage among most British citizens, who regard Sir Fred as a kind of ultimate free rider, and has also prompted a desperate search by Gordon Brown’s government to find some way of legitimately punishing Sir Fred, who apparently did not break any laws. In the absence of any legitimate threat of punishment, Sir Fred has appeared quite willing to seek and accept what most regard as a perversely unfair benefit-to-contribution ratio. While his is an extreme case, his behavior nevertheless seems to contradict CET and be consistent with EET. (The fact that EET sees this free riding as “natural” does not excuse it, of course; after all, it is just as “natural” to want to punish Sir Fred as it is for him to want to free ride.)

The phenomenon of top-level free riding underlines the fact that managers who are interested in punishing free riding fairly will often find themselves in the position of having to go after those at the very highest levels of an organization. It is important to emphasize this fact because it is often underappreciated. In the management literature, reviews of the free rider problem (e.g., Albanese & Van Fleet, 1985; Kidwell & Bennett, 1993) have tended to focus on what managers can do to prevent free riding by employees. But although the term free rider may conjure up an image in the minds of some of shirking workers at lower organizational levels,
this assumption is likely false. Indeed, in a culture such as the United States in the early twenty-first century, when it is not unusual for the salaries of top management to be hundreds of times higher than those of the average worker, managers would seem to be in a relatively good position to acquire disproportionately high benefit-to-contribution ratios, if their contributions are not actually hundreds of times higher than those of the workers. It should be noted that that is a big “if.” As Novak (1991) points out, the decisions of a talented manager can enrich a company to an extent that easily justifies, at least in financial terms, a salary that seems astronomical when compared to the average worker. Therefore, because the talent of managers is often an extremely valuable input, their extremely high salaries should not necessarily seem objectionable from the perspective of classic or evolutionary equity theory. On the other hand, higher salaries put management under greater pressure to perform well, because there is more opportunity for their inputs to fall short of their outcomes. The perception that this shortfall is occurring should become more likely when general economic conditions are bad, because in these times it becomes harder for management to make company-enriching decisions. That is probably why anger about high executive pay has become more common during the financial crisis of recent months. This idea—that managers are more likely to be perceived as free riders when times are bad—is quite interesting and probably underresearched (“What’s Behind the Rise in CEO Pay?”, 2005).

Sir Fred Goodwin was mentioned above primarily because his case provides a colorful and timely illustration of top-level free riding. One could question the appropriateness of using his case as an example, however, because much of the most visible outrage he has inspired (for example, from the British public) has been extraorganizational. Perhaps it would be more appropriate to regard the “organization” in this illustration not as RBS but as the larger British society, with Sir Fred being seen by many British citizens as having been lavishly rewarded for damaging British society (that is not to suggest, of course, that Sir Fred’s behavior could not also have inspired anger within RBS). Regardless, evidence does suggest that the perception of managerial free riding probably is a common and potentially damaging within-organization phenomenon. When top management salaries are perceived to exceed their inputs, it can create distributive justice concerns among middle managers and lower-level workers (Cowherd & Levine, 1992; Wilhelm, 1993), which may produce cynicism (Andersson & Bateman, 1997), lack of commitment, and lower product quality (Cowherd & Levine, 1992). Although management researchers do not usually use the term free riders to describe managers in this context, the term is probably apt as a generic descriptor of members whose benefit-to-contribution ratios are perceived to be inequitably high.
As noted above, while free riding can occur at any organizational level, the management literature has tended to focus on lower-level rather than top-level free riding. The focus here on top-level free riding has been an effort to correct that, and there is also one further reason why the problem of top-level free riding merits special attention: top-level free riding will usually be more difficult to police and control than lower-level free riding. With lower-level free riding, management can be expected to step in and solve the problem because they have incentives to do so: by ending free riding they can benefit themselves as well as their whole organization. However, when it is the managers themselves who have been seduced by the free rider advantage, their interests (at least in the short term) will lie in allowing their own free riding to persist. Because free riders cannot be relied upon to police and punish themselves, the problem of top-level free riding is better solved by shareholders whose interests lie in promoting success for the whole organization (and not just for a particular individual or group within the organization), and who understand that free rider problems at any level will likely impede this success.

Endnote

1. A more detailed description of the public goods game follows. The game typically involves a group of participants in a lab who are subdivided into several game-playing groups; for example, 16 participants may compose 4 separate 4-player public goods games. Games are played via the lab’s computer network for at least 10 rounds, under essentially anonymous conditions; players are known to co-players only by handles such as Player One, Player Two, etc. Each round, players are granted an endowment of tokens that are exchanged for real cash at the end of the game. Each player can choose whether to contribute none, some, or all of his tokens toward the production of a public good. The game’s parameters are set to create a situation in which individuals gain an advantage from free riding. More formally, the rules of this game are as follows: in a game involving a group of $n$ members, a player’s individual contribution $c$ is multiplied by $m$ to create a public good $mc$, which is shared equally by all players. The value of $m$ is always greater than 1.0, so that contributions actually do produce resources for the group. However, $m$ is also always less than $n$, which means that one can profit more individually if one fails to contribute, and instead free rides on the contribution efforts of co-members. (If $m > n$, in contrast—i.e., if contributions are multiplied in an extremely productive way—then individuals gain more from contributing than from free riding). At the end of each round, whatever public good has been produced by all players’ contributions is redistributed equally among all group members regardless of whether they contributed or
not. Following this redistribution, players are informed of the contribution and share sizes of co-players, and so are able to adjust their behavior based on the behavior of others.

References


