

# Reframing Evil in Evolutionary and Game Theoretic Terms

*Theodore P. Seto*

Calling a person or act “evil” is an act of moral rejection. We thereby place the person or act beyond the bounds of the even arguably defensible. But what is “evil”? Is it merely a linguistic placeholder for our subjective moral response? Or can it be given a more concrete meaning? I am in the process of developing a theory of normative obligation based on evolutionary and game theory.<sup>1</sup> This chapter will preview relevant portions of my theory. It will then apply that theory to three problems. First, it will attempt to define evil. Second, it will address the problem of original sin – in secular terms, why it is that we sometimes find it hard to be good. Third, it will explore why evil is often associated with a desire to dominate. “Evil,” I conclude, can be defined both concretely and usefully.

## 1. An Evolutionary Theory of Motivation and Normative Obligation

The most serious problem with existing moral theory is that it fails adequately to answer the following question: Why are we motivated to be good? The moral theorist will argue passionately that goodness is X – adherence to a categorical imperative, for example, or maximization of total utility – but will offer no plausible reason why real human beings might actually care about X. Most simply assume that human beings are inherently motivated to care about goodness, however defined.

This is profoundly unsatisfactory. We know a lot about the origins of motivation, at least in broad outline. Evolutionary theory tells us that motivations, like other attributes, generally evolve because they are adaptive. This means that individuals with adaptive motivations are more likely to survive and reproduce than individuals without such motivations. And this, in turn, means that any definition of goodness not tied in some way to adaptivity is implausible.

A simple thought experiment demonstrates why this is so. Assume a population consisting of two types of individuals, good and bad – that is, people motivated to be “good” (however we define goodness), and people not so motivated. Assume further that good parents are more likely to produce good children and bad parents more likely to produce bad – in other words, that behaviours are to some extent transmissible. (For reasons beyond the scope of this chapter, the assumption of parent-child transmission is not necessary. It does, however, simplify my exposition.) Finally and most importantly, assume that being good is, in net effect, *not* adaptive at the individual level. In other words, assume that being good involves some net sacrifice to long-term individual prospects

for survival and reproduction. Over time, what will happen is that good people will survive and reproduce at a lower rate than bad people. They will not necessarily die out altogether, but they will become a smaller and smaller percentage of the population – eventually a percentage small enough that the average person will no longer care what goodness is or why good people are motivated to be good. And this is so simply because of the mathematics of probability. I am not invoking here an area of science that is controversial or uncertain. I am invoking basic math.

But we know that the average person does care about goodness. It follows that being good must be adaptive. Indeed, it follows that goodness should be amenable to definition as a particular kind of adaptive behaviour. This explains why we are motivated to be good: we are so motivated because good people tend, on average, to survive and reproduce more successfully than bad. To ignore this dynamic – to conjure moral theory out of philosophical ether and then simply assert that we should care about the result – is no more likely to produce enduring truth than fourteenth-century alchemy.

The next question, of course, is “Why is goodness adaptive?” The answer, I suggest, can be found in the theory of repeat games. This branch of mathematics has become quite elaborate in recent years. I will discuss it in only the simplest terms. In particular, I want to focus on a game known as the repeated Prisoner’s Dilemma.

Our game involves two players, A and B. Each can make only one decision: she can cooperate (C) or defect (D). The payoffs to each player depend on the decisions of both. The following is an example:

		Player B	
		C	D
Player A	C	3,3	1,4
	D	4,1	2,2

In the foregoing table, each pair of numbers represents the payoffs to the players, the first to A, the second to B. For example, if both cooperate (Row C and Column C), then each gets a payoff of 3. If A cooperates but B defects (Row C and Column D), then A only gets a payoff of 1, while B is rewarded with a payoff of 4. If they both defect (Row D and Column D), then each gets a payoff of 2.

Note A’s incentives. If she believes that B is going to cooperate, then A should defect, because then she will get a payoff of 4 rather than 3. If she believes that B is going to defect, then A should still defect, because then she will get a payoff of 2 rather than 1. In fact, in a single play of this game, it is always in A’s self-interest to defect. And because the table of payoffs is symmetric, the same is true of B’s incentives as well. Both

---

should defect. But if both defect, each will receive a payoff of 2, whereas if both cooperate, each will receive a payoff of 3. Hence the dilemma.

The dynamics of the game change if the players know that they are going to be playing more than once. Now if A defects, she knows that B may respond by himself defecting in the future, as a result of which each player will earn a progression of 2's rather than 3's.

What strategy for playing the repeated game is most successful? It turns out that the most successful strategies are variations of a strategy known, perhaps unfortunately, as "Tit for Tat."<sup>2</sup> Tit for Tat can be viewed as consisting of three parts: (1) begin by cooperating ("Do unto others as you would have them do unto you"), (2) if the other player defects, punish immediately ("An eye for an eye, a tooth for a tooth"), and (3) if the other player returns to cooperation, immediately return to cooperation yourself (in moral terms, forgive the other player). In other words, three of the world's most fundamental moral principles – the Golden Rule, punishment, and forgiveness – appear to be part of the most successful solutions to this purely mathematical game.

My moral theory is based on the assumption that this is not a coincidence. Individuals who appropriately cooperate, punish, and forgive tend to survive and reproduce more successfully than those who do not. As a result, the world has come to be dominated by individuals who, to some degree or another, have internalized these three principles. Morality, in short, responds to an underlying mathematics.

Life, of course, is far more complex than any mathematical game. Even the most sophisticated games under study today are but crude approximations of reality. We are not yet able adequately to specify either the relevant game or its optimal solution. Nevertheless, my theory assumes that such a game exists. I label its optimal solution the "principle of reciprocity", and I assume that Tit for Tat, the various formulations of the Golden Rule,<sup>3</sup> Kant's categorical imperative,<sup>4</sup> John Rawls' choice from behind the veil,<sup>5</sup> and the classic parental question "How would you feel if Mary did that to you?" are all approximations of this principle. Goodness is simply adherence to the principle of reciprocity – no more, no less.

Each culture implements this principle in a set of rules which members are obliged to observe vis-à-vis others of the same culture. I call such rules the culture's "ethos of reciprocity." The set of individuals thus bound and protected constitutes the "We" of that ethos. Those not so bound and protected are "They" or the "Others." History is in part the story of the expansion of our most general "We" – from the tribe to the city-state to the nation to humanity as a whole. This expansion is predicted by game theory; to the extent that moral actors are excluded from our "We," our relations with those actors are likely to be non-optimal because they will not be based on the principle of reciprocity. A Hobbesian international order is no more functional than a Hobbesian tribe. Over

time, those who internalize behaviours consistent with the principle of reciprocity vis-à-vis Others, bringing those Others within their “We,” are more likely to survive and reproduce than those who do not.

The fact that goodness is adaptive, of course, explains *why* we might be motivated to be good, but not *how*. Two common explanations are inadequate. If motivations evolve because they are adaptive, one might be tempted to hypothesize a genetic basis for goodness, what some have called an “altruism gene.” Much work remains to be done in this area; preliminarily, however, it seems unlikely that genes can explain all the complexities of moral decision-making. Goodness, we have seen, has at least three parts – being cooperative, punishing, and forgiving. To be fully effective, an altruism gene would have to code for all three. A gene that merely motivated us to be nice would *disable*, rather than enable, goodness when punishment is the optimal path. A subpopulation unable to fight back when taken advantage of would quickly be washed out of the gene pool. This is not to say that genes are irrelevant. We are genetically motivated to bond with others and to become angry when others treat us poorly; an entire region of our brain appears to be devoted to the specialized task of detecting cheating on the social contract.<sup>6</sup> I merely suggest that genes, by themselves, are not enough.

I also reject reason as a sufficient explanation for goodness. The adaptivity of many learned behaviours – sharing, for example – only becomes obvious with extensive experience, sometimes the experience of many generations. Were we motivated solely by reason and self-interest, individuals might never undertake such behaviours or do so only after painfully inventing the wheel again and again. Indeed, if reason were a sufficient explanation, good parenting would be irrelevant, and institutions such as religion, devoted to the transmission of codes of virtue, would be unnecessary. Each child, upon reaching the age of reason, would deduce the same moral code – an explanation wholly inconsistent with our experiences as parents and teachers. We have a name for individuals motivated solely by reason and self-interest: we call them “sociopaths.” But if not by genes or reason, how are we motivated to goodness?

My answer is that goodness should be viewed primarily as a set of internalized learned behaviours, transmitted from generation to generation. Although we do not fully understand the mechanics of such transmission, we do know that each generation comes to feel a compulsion to perform the transmitted behaviours and to feel discomfort – guilt or shame – when it does not. In my theory, learned behaviours evolve. That is to say, over time the population of learned behaviours becomes better adapted to the environmental conditions it faces. For the most part, this means that learned behaviours evolve to make individual humans better adapted to their environment as well. Most importantly, we learn more

---

each generation about what it means to be good, and we transmit those lessons, new and old, to our children, who hopefully internalize them and act on them regardless of reason and apparent self-interest.

The moral theory I propose is thus both consequentialist and deontological. It is consequentialist in its ultimate explanation. Behaviours are good because of their consequences – they help us to survive and reproduce. Operationally, however, my theory is almost completely deontological: it requires that we internalize rules and, for the most part, follow them regardless of their short-term expected consequences.

## 2. Defining “Evil”

With this introduction, I turn to the definition of “evil.” If goodness is adherence to the principle of reciprocity, one might be tempted to define evil simply as a failure to adhere to that principle. Without further elaboration, however, such a definition would gloss over at least three important problems.

First, the principle of reciprocity is not just the Golden Rule. It also requires punishment and forgiveness. Most of us find it intuitively easy to classify violations of the Golden Rule as “evil.” Perhaps less obviously, my theory suggests that a failure to forgive once another has returned to cooperation may be equally evil. Forgiveness of others, not merely personal adherence to the paths of righteousness, appears to be a necessary part of the most successful solutions to the repeated Prisoner’s Dilemma.

More difficult is the issue of punishment. Some have characterized punishment as itself an evil, albeit necessary. My theory suggests, to the contrary, that *failure to punish* may be just as evil as failure to be good oneself. Is this so?

Unfortunately, punishment is poorly modelled in current repeat game simulations. Real human beings do not merely cooperate or defect; we engage in an almost infinite range of social behaviours. Punishment may range from a mere withholding of approval to the most extreme fantasies of the vengeful. As a matter of game theory, all that is needed is an environment that decreases the likelihood that “bad” learned behaviours will survive and reproduce and increases the likelihood that “good” learned behaviours will do so. Articulating this goal is easy. Achieving it in the real world is not.

The problem is further complicated by the fact that modern society, for the most part, transfers the punitive role to the state. This transfer solves a problem inherent in repeat games – that cooperative solutions are often not the only evolutionarily stable ones. When an individual player assumes the punitive role, it may be ambiguous whether his nastiness is punishment (and therefore good) or mere defection (and therefore bad). This ambiguity may lead the other player to punish in turn.

If so, the result may not be long-term cooperation; it may instead be a stable cycle of mutual defection – in other words, a blood feud.

Law solves this problem by removing the punitive role to a neutral third party. Punishment by a third party pursuant to neutral rules is much less likely to be misconstrued. Obviously, to the extent that law is perceived as biased, it will be less effective at solving this problem. The most effective legal order, therefore, is one that treats all players as equal under neutral rules. The rule of law and equality under law are thus both solutions to a game theoretic problem.

Individually-administered punishment – vigilante justice – may actually undermine these solutions, at least with respect to infractions subject to legal correction. It may be helpful, perhaps even necessary, for individuals to disapprove of the defections of others. To permit such individuals to retaliate further, however, risks development of a state of blood feud. In a society of laws, it may be more adaptive for individuals to turn the other cheek. This is not because punishment is itself morally wrong; indeed, for the *state* to turn the other cheek – to fail to perform its punitive role – would be maladaptive and therefore wrong. It is rather because individual performance of a major punitive role often creates more problems than it solves. Failure to adhere to the Golden Rule and failure to forgive are both individual evils; in a society of laws, failure to punish (beyond the expression of disapproval) may not be.

A second problem with defining evil simply as a failure to adhere to the principle of reciprocity is that in my theory goodness is merely a subset of adaptivity. There may be situations in which adherence to the principle of reciprocity is itself maladaptive. If so, in such situations violation of the principle should not be condemned, and therefore should not be labelled “evil.” For example, although killing another is generally a violation of the principle of reciprocity, allowing oneself to be killed is not adaptive. Therefore, in both morality and law, we recognize self-defense as an exception to the wrong of killing another.

Perhaps an even more important example is the right to free expression of ideas. We might try to frame that right in terms of the principle of reciprocity: I allow you to say what you want because I want the right to say what I want. The problem is that the principle of reciprocity can also be used to justify censorship: I should not say things that will make you uncomfortable because I do not want you to say things that will make me uncomfortable. The right to free expression is more persuasively justified as necessary to facilitate the rapid evolution of learned behaviours. If we can criticize, propose, and debate, our learned behaviours will likely evolve more rapidly. The ability to adapt rapidly is itself extremely adaptive. Where free expression conflicts with the

---

principle of reciprocity, it is often adaptive – and therefore normatively right – to resolve that conflict in favour of free expression.

By suggesting that adaptivity sometimes overrides the principle of reciprocity, I do not mean to suggest that we are free to disregard received moral codes whenever it may appear adaptive to do so. To the contrary, those codes themselves represent society's accumulated wisdom about adaptivity; it is therefore essential that most actors behave as if bound most of the time. My position rather is that the codes themselves sometimes include deviations from an unalloyed application of the principle of reciprocity in situations where breach is more adaptive than adherence. This is particularly likely to be true where survival and reproduction are directly at issue. Hence perhaps the saying: "All's fair in love and war." A man who falls in love with and marries his best friend's wife may violate the principle of reciprocity; we do not, however, necessarily label his actions "evil." Similarly, many acts that would be reprehensible in other contexts are condoned in times of war.

It follows that the concept of evil must be limited to breaches of the principle of reciprocity that are themselves maladaptive. Unfortunately, adaptivity is extremely difficult to ascertain. If a society's moral codes represent its accumulated wisdom about adaptivity, this limitation might even be read to suggest that moral codes cannot themselves be evil. Consider, for example, the moral codes of the nineteenth-century American South, which treated people of African ancestry as subhuman. One might argue that such codes were adaptive to white Americans and therefore, under my definition, not evil.

To this line of reasoning I offer two responses. First, I view all existing learned behaviours as in the midst of evolution, not as evolution's final product. Just as we have not yet invented the perfect mousetrap, so we have not yet achieved the perfect moral code. Moral codes can embody and reflect states of stable defection; they do not automatically result in cooperation. More simply, the fact that a moral code exists does not necessarily make it right. Second, arguments of momentary advantage do not generally justify breaches of the principle of reciprocity – no more so for groups who create their own moral codes than for individuals. When adaptivity trumps reciprocity, it must do so pursuant to neutral rules applicable equally to all. Otherwise, it is likely to provoke defection in return.

A third problem with defining evil simply as a failure to adhere to the principle of reciprocity is that such a definition ignores the purpose of the concept itself. That purpose is to influence *intentional* behaviour. Labelling an act "evil" may deter the listener from performing the act. It may justify retribution when such an act occurs. It may communicate or reinforce values with respect to such acts. To be useful for any of these purposes, however, "evil" must be limited to acts that involve choice.

Labelling something evil will not deter if choice does not exist. In a sophisticated moral world, retribution is generally justified only if the wrong was not accidental. The communication of values, similarly, is intended to influence future deliberate action.

To make the concept of evil useful for these purposes, therefore, we must limit its scope to actions undertaken with some degree of intentionality – what Anglo-American criminal law would call *mens rea*. A mother who intentionally puts a knife into her child’s heart may be “evil”; a mother who is startled by an unexpected thunderclap and drops her child, thereby killing it, is not, even though the child is equally dead. The degree of intentionality required generally depends on the seriousness of the resulting breach of the principle of reciprocity. We may view a nasty comment as “evil” only if the resulting hurt was intended. If an act results in death, by contrast, we often label the act “evil” even if the actor is merely culpably inattentive to the possibility that death might result – as, for example, where a tobacco company persuades itself that its products are harmless. Defining with greater specificity the degree of intentionality required is beyond the scope of this chapter. For current purposes, it is enough to say that to perform its role in acts of moral rejection properly, “evil” must be limited to acts involving some degree of intentionality.

Finally, we normally reserve the term “evil” for serious breaches of our norms. Taking into account all of the foregoing caveats, we can therefore define “evil” as *a serious intentional maladaptive breach of the principle of reciprocity other than an individual failure to punish*. In short, “evil” is not merely a linguistic placeholder for a particular type of moral response; it is amenable to concrete definition. But can an evolutionary and game theoretic approach to defining evil aid in understanding the problem? The remaining two sections of this chapter assert that it can.

### 3. Original Sin (or Why It Is Sometimes Hard To Be Good)

At first blush, evolutionary theory might seem to lead to the conclusion that being good should be easy. If individuals who do not adhere to the principle of reciprocity are less likely to survive and reproduce, we might expect that humans would have evolved so as to be motivated unambiguously to adhere to that principle – that is, to be good. By this logic, it should be just as easy to be good when faced with moral choice as it is easy to drink water when faced with thirst. Humans not so motivated should be washed out of the gene pool. Temptation simply should not exist. But, of course, we know that it does. Why? How can evolutionary theory account for moral difficulty?

The answer is that the evolutionary pressure to adhere to the principle of reciprocity is not the only evolutionary pressure we face. To

---

explore this issue in all of the contexts in which it arises is beyond the scope of a single book chapter. Here, I will focus on one narrow but important part of human behaviour – mating and childrearing among sexually active heterosexuals. In most of its implementations, the principle of reciprocity requires marital fidelity: “I will be faithful to you and help raise any resulting children. In return, I expect you to be faithful to me and similarly participate in the support, rearing, and protection of our children.” Adherence to the principle is likely to result in more successful childrearing and therefore in more successful reproduction of the parents’ genes and behaviours. Yet childbirth is not limited to mated partners and even mated partners sometimes find adherence difficult. Why? Because the motivational structures that determine our mating and childrearing behaviours have evolved in response to competing evolutionary pressures – not merely in response to the principle of reciprocity. I call such competing evolutionary pressures “mandates.” Each species, in turn, resolves conflicts among the various mandates it faces in different ways.

The first and most fundamental is the *mandate of reproduction* – the direct evolutionary pressure to survive and reproduce. A species that fails to respond to this mandate will most assuredly become extinct. In humans, as in other species, we should therefore expect to find many both genetically-triggered and learned behaviours responsive to this mandate. And we do.

We are, for example, genetically motivated to seek pleasure and avoid pain. Sexual intercourse provides us pleasure; as a result we are motivated to engage in the acts necessary to produce offspring. But other, less obvious motivations are responsive to the mandate of reproduction as well. Human males, for example, produce effectively infinite amounts of sperm; they can increase the likelihood they will reproduce by spreading that sperm as widely as possible. Not surprisingly, we find that human males are often motivated to do so – a pattern of motivation and behaviour known as “polygyny” that conflicts directly with the obligation of marital fidelity. For the same reason, human males tend to be sexually stimulated by simple, fairly indiscriminate visual cues. Human females, by contrast, are biologically limited in their number of offspring. They can increase the likelihood of successful reproduction by being more selective in their mating. During the infertile portions of their cycle, they tend to be attracted to men who would make good husbands and childcare providers. Studies show that on days of fertility they tend, by contrast, to be motivated to mate with more virile men – again, often in direct conflict with the obligation of marital fidelity.

Many important attributes of multicellular species cannot be explained solely by reference to the mandate of reproduction. One of the most important such attributes is sex – reproduction through the mixing of genetic material. Perhaps counterintuitively, sex is not responsive to the

mandate of reproduction. Why? The most efficient way for an organism to reproduce is by cloning – the production of genetically identical offspring. Diploidal reproduction, in which each of two parents contributes half of a child’s genetic material, is only half as likely to reproduce a given parent’s genes. Indeed, diploidal reproduction ensures that none of us will ever actually perfectly reproduce; human children are always genetically different from both their parents. Thus sex represents a serious deviation from the mandate of reproduction.

Sex responds instead to a different mandate, which I call the *mandate of genetic diversity*. Evolution requires imperfect reproduction. In simple organisms with extremely large populations, such as bacteria, genetic mutation supplies the necessary imperfection. In species with more limited populations, including most multicellular organisms, mutation does not occur rapidly enough to permit evolution to operate at high enough speeds to allow species to adapt effectively to changing environmental conditions (in particular, to quickly evolving viruses, bacteria, and other parasites). Here sex – the production of offspring through the mixing of genetic material – comes to the rescue. Populations of creatures that reproduce sexually will be far more genetically diverse than populations of similar size that reproduce without such genetic mixing. When environmental conditions change, it is more likely that some portion of the sexually reproducing population will already carry the genes necessary to deal with that change. In other words, sex allows us to evolve to meet changing conditions more quickly.

If genetic diversity is adaptive, we ought to observe the mandate of genetic diversity operating in our choice of mates. And we do. Despite sex, we could reproduce more perfectly, and thereby respond more effectively to the mandate of reproduction, by mating with our closest genetic kin – in other words, through incest. The mandate of genetic diversity, however, predicts the evolution of inhibitions to incest; and, indeed, we all carry such inhibitions, both genetic and learned. The mandate of genetic diversity also predicts that our mating choices will be somewhat random; and, indeed, we are often motivated to mate with unexpected, sometimes even objectively unsuitable, partners – again often in conflict with our marital obligations. As Pascal observed, “Le coeur a ses raisons, que la raison ne connaît point.” (“The heart has its reasons, of which reason knows not.”)

The mandate of genetic diversity derives its power from the fact that the ability to evolve quickly is itself adaptive. But learned behaviours can adapt to changing conditions far more quickly than behaviours that are directly genetically triggered. It should therefore be an evolutionarily successful strategy to make major sacrifices to obtain the ability to carry and transmit learned behaviours. And we humans do. Unlike the young of

---

almost every other species, our young are born helpless. They cannot walk or feed themselves. Their heads are so large that they often kill their mothers in childbirth. They require over a decade to reach adulthood, in the meantime consuming extraordinary amounts of parental energy. From the perspective of simple survival and reproduction, human young make no sense whatever. Their characteristics respond instead to a third mandate: the *mandate of learned behaviour*, which rewards departures from the first two mandates to make learned behaviours possible.

This third mandate again has profound effects on our mating patterns, this time not necessarily in conflict with our marital obligations. It is no longer in the male's interest simply to impregnate and leave. Helpless children protected and raised by two parents are more likely to survive and reproduce. The male's participation in childrearing also allows him to transmit his learned behaviours, not merely his genes, to his offspring. It is also no longer in the female's interest simply to seek the sperm of the most genetically attractive male. She now needs a mate who will stick around to help raise her children; she may also be more concerned about the learned behaviours he will transmit to those children.

The evolutionary advantage of adhering to the principle of reciprocity is thus but one of many mandates, the *mandate of reciprocity*. In most cultures, reciprocity requires stable and monogamous matings. When we internalize our cultures' moral codes, we learn to feel discomfort if we breach the rules requiring such matings. The fact that we may ultimately respect those rules – that we give primacy to the fourth mandate – however, does not mean that doing so will necessarily be easy. Our genetic heritage retains the influence of the other three mandates. Men still feel the urge to spread their seed. We still become attracted to others when we shouldn't. No matter how effectively we internalize learned behaviours that implement the principle of reciprocity, we will always face temptation.

In my theory, original sin – the difficulty of being good – derives from such conflicts among the mandates. I have told the story of original sin in the marital context. Similar stories can be told in other contexts. We are motivated to evil because our motivational structures respond to more than just the principle of reciprocity. The very act of labelling something “evil” is an attempt to reinforce our internalized learned behaviours – responsive to the mandate of reciprocity – against conflicting motivations responsive to other mandates.

#### 4. **Evil and the Desire to Dominate**

One further such story explains why we often associate evil with a desire to dominate. Tolkien's Sauron<sup>7</sup> does not merely abhor the good; he insists on controlling the world. Satan is often similarly portrayed. As I have defined evil, of course, it should be possible to be thoroughly evil

---

without controlling anything; it should be equally possible to have great power without being evil. Indeed, religion commonly posits a God who is both omnipotent and good. Why the association between evil and power? And why, more specifically, is this association made in some circumstances but not others?

At least two factors, both responsive to the mandate of reproduction, contribute to the human desire to dominate. First, there is the simple desire to get one's way. If we can tell others what to do, we are more likely to be able to survive and reproduce. Not surprisingly, this desire is strong in humans, regardless of gender, from infancy on.

A second component of the human desire to dominate is more pronounced in males. I have already noted the human female tendency to be selective in mating. Such selectivity often takes the form of a preference for males of high status – a behavioural pattern known as “hypergamy.” The corresponding male preference for females of high status is substantially weaker; human males tend to be attracted to nubile females regardless of status (indeed, sometimes seemingly without regard to any other factor whatever). The result is that in non-monogamous societies, high status males will commonly be able to obtain multiple mates; lower status males will commonly be less reproductively successful. Even in modern ostensibly monogamous societies, studies show greater variability in the number of offspring of males than in the number of offspring of females. This suggests that even low status females are able to find males to impregnate them; low status males, by contrast, are often unable to reproduce at all. (Similar phenomena can be observed in many other species, especially among the larger mammals – deer, wolves, walruses, and gorillas, among others. Males compete for status. The winners mate; the losers do not.)

The net effect is to give males a powerful evolutionary reason to compete for dominance. Males not so motivated are less likely to reproduce. Significantly, when evil is associated with a desire to dominate, it is almost always portrayed as male. Female Sauron-equivalents (evil characters with a desire to take over the world) are not nearly as common.

Regardless of whether it derives from a desire to advance one's own survival and reproduction or from a male response to female hypergamy, the impulse to dominate almost inevitably conflicts with the principle of reciprocity. When we adhere to the principle of reciprocity, we typically do not get everything we want. This is why evil and the desire to dominate are often linked.

Motive, however, is critical. Those who seek power for altruistic purposes are viewed differently than those who seek power selfishly. History and literature are filled with images of true and just kingship, which necessarily implies rule for the benefit of all, not just for the benefit

of the ruler. Tolkien's Aragorn<sup>8</sup> is an obvious example. The purest version of true kingship, of course, is the monotheistic vision of God. God is not motivated to power by a need to survive and reproduce. He is free to respond solely to the mandate of reciprocity; alternatively, that mandate is his mandate, built into the mathematics of the universe and reflecting something essential about his character. It is only when a desire for power reflects selfishness, responsive solely to the mandate of reproduction, that we associate power with evil.

One further aspect of this relationship merits note. A party with disproportionate power is less likely to be punished. The theory of repeat games tells us that one who is never punished for his defections should become increasingly likely to defect. In the words of Lord Acton: "All power tends to corrupt; absolute power corrupts absolutely." Substantial power imbalances invite breaches of the principle of reciprocity, regardless of the initial good intentions of the powerful. One might posit that the United States and Israel, for example, approach their conflicts solely with good intentions. The power imbalance between either of these countries and its foes, however, creates great risk of increasing deviations from the principle of reciprocity over time. In the long run, such deviations are unlikely to be adaptive.

## 5. Conclusion

My theory of normative obligation is still under development. It is not my purpose here to justify or defend that theory in detail. Rather my purpose is to suggest that the theory may be able to do useful work – that it sheds light on some of the most important characteristics of evil. Evil can be defined. And it can be defined in a way that is rigorous, consistent with common usage, and illuminating.

## Notes

<sup>1</sup> Portions of the theory are outlined in Seto, 2001 and Seto, 2002.

<sup>2</sup> See generally Axelrod, 1985 and Axelrod, 1997.

<sup>3</sup> Consider the following quotations: In Christianity, "All things whatsoever ye would that men should do to you, do ye even so to them: for this is the Law and the Prophets" (Matthew 7:12); in Judaism, "What is hateful to you, do not to your fellowmen. That is the entire Law: All the rest is commentary" (*Talmud*, Shabbat 31a); in Islam, "Do to all men as you would wish to have done unto you, and reject for others what you would reject for yourselves" (Mohammed in the *Hadith*); in Confucianism, "Confucius was asked, 'Is there one word which may serve as a rule of practice for all one's life?' And he replied, 'Is "reciprocity" not such a

word? Do not to others what you do not want done to yourself” (*Analects of Confucius*); in Hinduism, “This is the sum of duty: do naught unto others which would cause you pain if done to you” (*Mahabharata* 5, 1517); in Buddhism, “Hurt not others in ways that you yourself would find hurtful” (*Udana-Varga* 5, 18); in Taoism, “Regard your neighbor’s gain as your own gain, and your neighbor’s loss as your own loss” (*T’ai Shang Ken Ying P’ien*); in Zoroastrianism, “That which is good for all and any one – For whomsoever – that is good for me. What I hold good for self, I should for all. Only Law Universal is true Law.”

<sup>4</sup> Kant, 1963a, 119-121 and Kant, 1963b, 18-22.

<sup>5</sup> Rawls, 1999, 118-123.

<sup>6</sup> Cosmides and Tooby, 1995, 163, 182, 205.

<sup>7</sup> Tolkien, 1965.

<sup>8</sup> Tolkien, 1965.

### References

- Axelrod, R. (1997), *The complexity of cooperation: agent-based models of competition and collaboration*. Princeton: Princeton University Press.
- (1985), *The evolution of cooperation*. New York: Basic Books.
- Cosmides, L. and J. Tooby. (1995), ‘Cognitive adaptations for social exchange’, in: J.H. Barkow *et al.* (eds.) *The adapted mind: evolutionary psychology and the generation of culture*. Oxford: Oxford University Press. 163-228.
- Kant, E. (1963), ‘Critical examination of practical reason’, in: *Kant’s ‘Critique of practical reason’ and other works on theory of ethics*. T.K. Abbott (trans.) London: Longmans, Green & Co. 87-262.
- (1963), ‘Fundamental principles of the metaphysic of morals’ in: *Kant’s ‘Critique of practical reason’ and other works on theory of ethics*. 1-84.
- Rawls, J. (1999), *A theory of justice*. Cambridge, MA: Belknap.
- Seto, T. (2001), ‘Intergenerational decision making: an evolutionary perspective’, *Loyola Los Angeles Law Review* 35: 235-76.
- (2002), ‘The morality of terrorism’, *Loyola of Los Angeles law review* 35: 1227-63.
- Tolkien, J.R.R. (1965), *Lord of the rings*. Boston: Houghton Mifflin.