MICHAEL E. MCCULLOUGH

BEYOND REVENGE

THE EVOLUTION OF THE FORGIVENESS INSTINCT

CHAPTER FIVE

FAMILY, FRIENDSHIP, AND THE FUNCTIONS OF FORGIVENESS

The desire for revenge isn't a disease or a defect in human nature. It's authentically human, and it has always been a part of us. This fact might suggest that a dismal forecast for our species' future is in order. We're not going to get rid of humanity's desire for revenge with better social programs, better pharmaceuticals, or better psychotherapy. But there's some good news, too. Evolutionary science leads us squarely to the conclusion that forgiveness is also a built-in feature of human nature. As we'll see in the next few chapters, the capacity to forgive, like the desire for revenge, is a standard-issue human social instinct. Every neurologically intact person comes into this world outfitted with the capacity to forgive under certain circumstances.

Under certain circumstances. Those are three important words. If we want to learn how to make the world a less vengeful, more forgiving place, it's important that we figure out what those "certain circumstances" are. A good way to start is by taking a close look at the two functions for which the capacity to forgive was naturally

selected—its ability to help ancestral humans get along with their genetic relatives, and its ability to help ancestral humans establish and maintain cooperative relationships with nonrelatives.

GETTING ALONG WITH OUR RELATIVES

Many animals—humans certainly *not* excepted—are known to retaliate against their blood relatives, on occasion quite harshly.[1] Yet it's relatively rare for human beings to use blood revenge against their own kin. When someone kills another person, the victim is most likely a stranger, acquaintance, or romantic partner. Only very rarely do people kill their own children, their own parents, or their own siblings.[2] True, history and literature are replete with stories of people who did kill their brothers, sisters, parents, or children in pursuit of power, wealth, or love. You can also find plenty of evidence that people have engaged in infanticide (usually in extremely dire straits) for thousands of years. But in general, people are and, as far as we can tell, always have been hesitant to seek blood revenge (and, it's probably safe to assume, other forms of revenge) against their genetic relatives unless they have an awful lot to gain if they do, or unless they have an awful lot to lose if they don't.[3]

What is it exactly that makes revenge against genetic relatives so rare? Filial bonds, gratitude, mutual dependency, brotherly love—sure, all of these help to restrain our vengeance against loved ones, but from the perspective of evolution, there's a more fundamental cause of our restraint: harsh revenge against a blood relative, insofar as it actually reduces the relative's fitness, reduces the avenger's fitness as well. Like it or not, your sister Tracy and your cousin Tommy are carrying around some of your genes (siblings share 50 percent of their genes in common; for first cousins it's 12.5 percent), so if you remove their genes from the gene pool, you're removing some of yours as well. The idea that my fitness isn't dependent simply on how many offspring *I* have (that is, how many of my genes I get into future generations through my own reproductive success), but also on how many offspring my genetic relatives have, is called *inclusive fitness.* Inclusive fitness is the real measure of how well we're actually doing in the game of achieving evolutionary immortality.

The biologist J.B.S. Haldane once quipped that although he wouldn't lay down his life for a brother, he would do so for two brothers or eight cousins. Saving the lives of two brothers is equivalent to saving half of your genes, twice. Harshly retaliating against a genetic relative, therefore, damages your inclusive fitness. Biologists often put it more subtly, perhaps asserting that the payoffs that an individual receives from retaliation are "modified by kinship."[4] But the bottom line is that if retaliating against your sister costs her two "fitness units," it's going to cost you one fitness unit right off the bat because you're one of your sister's major genetic shareholders. Therefore, a proneness to forgive our blood relatives probably evolved because people who had such a trait were able to avoid shooting themselves in the foot (or the genes?) by reducing the fitness of their blood relatives.

COOPERATING WITH NONRELATIVES

Forgiveness probably also served a second function during human evolution: encouraging cooperation among nonrelatives. As we've already seen, revenge has helped to solve that problem, but forgiveness is also part of the solution. Indeed, a big part of why we're inclined to forgive our friends, neighbors, and associates today is probably because forgiveness enabled our ancestors to develop and maintain the cooperative alliances that they needed to thrive in large groups. Rather than thinking of the relationship between revenge and forgiveness as one of disease and cure, or poison and antidote, we'd do better to think of revenge and forgiveness as a team of midwives that helped give birth to human beings' ultra-cooperativeness. It's worth taking some time to savor this idea, and we'll do so in the remainder of this chapter by breathing the rarefied air of game theory.

GAMES GUPPIES PLAY

Recall our predator-inspecting guppies from Chapter Four. Let's consider the costs and benefits of predator inspection, starting with the costs. Predator inspection requires time and energy that could otherwise be used for foraging. There's also the risk

that a predator-inspecting guppy will become a sunfish's next meal. Therefore, it would be in Bob the guppy's best interests to shirk his predator-inspection duties—if only he could get somebody else do the dirty work.

However, predator inspection produces valuable information: it tells the fish whether it's safe to continue looking for food or whether they ought to clear the area to avoid getting gobbled up. Eventually, somebody's going to get eaten on the job, but the risk to any single guppy on any single mission (if it has a willing inspection partner) is fairly low. Therefore, the average payoff, in fitness terms, is higher for the individual guppies if *somebody* inspects predators than if *nobody* inspects predators. However, there is one type of fish that fares very poorly: the poor sucker who volunteers for a predator-inspection mission with a partner who plans to abandon him once they're in the thick of things. Recall that a fish is much more likely to get eaten during predator inspection when working solo than when working with a partner.

We can rank the possible outcomes in terms of their net benefits to a predator-inspecting guppy. The best outcome goes to an individual who betrays his partner once they're out on the job (high benefit in the form of information, no cost in the form of mortality risk). The second-best outcome goes to the individual who, when his turn comes, volunteers for a predator-inspection mission with a partner that will hold up his own end of the deal (high benefit in the form of information, small cost in the form of mortality risk). The third-best outcome goes to the fish if *nobody* volunteers for predator-inspection missions (no benefit, no cost), and the very worst outcome goes to the hapless fish who goes out on a mission along with a partner who shirks his responsibilities in the middle of the job (high benefit in the form of information, but very high cost in the form of mortality risk).

There's a predicament here: the self-interested course of action for a guppy on a predator-inspection mission can't be determined in a vacuum, but rather, depends on what its partner does. Figure 5.1 shows how the effect of Guppy A's strategy choices on its fitness can only be predicted if we also know what Guppy B is going to do.

FIGURE 5.1. THE PREDATOR-INSPECTING GUPPY'S DILEMMA

		Guppy B's Strategy	
		Advance	*Stall or Retreat*
Guppy A's Strategy	*Advance*	*A's Outcome:* Information discounted by small risk of being eaten *B's Outcome:* Information discounted by small risk of being eaten	*A's Outcome:* Information discounted by large risk of being eaten *B's Outcome:* Information with no risk of being eaten
	Stall or Retreat	*A's Outcome:* Information with no risk of being eaten *B's Outcome:* Information discounted by large risk of being eaten	*A's Outcome:* No information, no risk of being eaten *B's Outcome:* No information, no risk of being eaten

In the upper left corner, two guppies on a mission gain valuable information, but the value of that information is discounted by a small chance of being eaten. In the upper right corner, Guppy A advances toward the predator, but Guppy B shirks its responsibility. In this case, Guppy A gets information that is discounted steeply by a high risk of being eaten, but Guppy B gets the information for free—that is, without having to worry about being eaten. In the lower left corner, the outcomes are reversed: Guppy A, who is the free rider in this scenario, gets the information for free, whereas Guppy B has to work to get the information—a benefit that is made less valuable by the fact that it had to take a large risk to obtain it. The lower right corner depicts a situation in which neither guppy advances toward the predator. Neither gets the desired information about the sunfish's appetite, but neither gets eaten, either. When the effect of one's behavior on one's fitness depends on the behavior of somebody else, what's a guppy to do?

DILEMMAS, DILEMMAS EVERYWHERE

The predicament I've just described is a special case of a more generic problem that scientists call the "prisoner's dilemma." Here's the make-believe quandary from which the prisoner's dilemma derives its name. Two criminals are arrested on suspicion that they've committed a major crime. The police don't have enough evidence to support a conviction, so they interrogate the two criminals separately in hopes of obtaining testimony that they can use to charge one of them with the principal crime. If neither suspect betrays his accomplice, both criminals will be charged with a lesser crime, and they'll eventually get one-year sentences. If one criminal offers testimony against the other, he won't be charged with anything and the other criminal will be charged with the principal crime, which carries a three-year sentence. If they implicate each other, they'll both get two-year sentences. What's the smartest course of action in a situation like this? That's the prisoner's dilemma.

Science writer William Poundstone has called the prisoner's dilemma "one of the great ideas of the twentieth century, simple enough for anyone to grasp and of fundamental importance."[5] Because of its seemingly limitless ability to reveal surprising truths about social behavior, the political scientist Robert Axelrod has called it "the *E. coli* of the social sciences."[6] In addition to the many things the prisoner's dilemma can teach us about deception, trust, self-interest, and rational action, it also has a lot to teach us about forgiveness. In particular, it can show us how the capacity to forgive evolved to help people benefit from cooperation with non-kin.

Prisoner's dilemmas are all around us. Any time you find yourself in a situation in which the average outcome for everyone involved is highest if everyone works together, but the best *individual* outcome comes from defecting against everybody else, you're probably in a prisoner's dilemma.[7] If both prisoners stay quiet about their accomplice's guilt, their *average* outcome is the best, but they'd get the very best *individual* outcome by betraying a faithful accomplice. If both predator-inspecting guppies do what they're supposed to, their *average* outcome is better than any other average outcome, but the very best *individual* outcome is to be obtained by loafing while your partner does the inspecting.

The world's nations would have the best *average* outcome if they disavowed all aspirations for nuclear weapons, but the best *individual* outcome would result from being the only nation in the world with a nuclear bomb.

In the 1940s, a Hungarian mathematician named John von Neumann developed *game theory* to bring some mathematical rigor to these sorts of strategic dilemmas. Game theorists try to identify the courses of action that self-interested actors should pursue when they're locked into competitions, known generally as "games," with other self-interested actors whose future behavior is uncertain. The holy grail of game theory is the "Nash equilibrium," named after John Nash, a mathematician at MIT who worked out the details in the early 1950s. (Four decades later, after recovering from a long battle with severe mental illness, Nash would receive a Nobel Prize for this work, and would later be the subject of the book *A Beautiful Mind* and the film of the same title.) The Nash equilibrium occurs when both players have chosen a strategy from which neither has any rational incentive to deviate, irrespective of the other player's choices.[8]

Back to the guppies for a moment. If I'm a self-respecting guppy, I want to get the data about a sunfish's appetite with minimal risk of getting eaten in the process. My inspection partner wants exactly the same thing. The grim conclusion of game theory (a conclusion that led game theorists to recommend that the United States launch an unprovoked nuclear strike against the Soviet Union during the Cold War[9]) is that a self-interested actor competing with another self-interested actor should always defect. The Nash equilibrium for the prisoner's dilemma—the thing you should always do when your first priority is saving your own skin and you can't completely trust your partner—is to loaf during predator inspection, to rat out your accomplice, to unleash your ICBMs on your Cold War adversary. To do otherwise is to invite the misery that is the fate of patsies everywhere. Unless your opponent is extraordinarily stupid or insane, count on him to come to the same conclusion. It would be good for the fish if they engaged in predator inspection, good for criminals if they didn't betray each other, and good for the world if we refrained from nuclear first strikes, but we should expect that self-interested actors will always defect. Right?

Well, we should thank our lucky stars that U.S. presidents don't take the advice of mathematicians too seriously: in real life, fish really do inspect predators cooperatively. In real life, there often is honor among thieves. And we actually managed to avert mutually assured nuclear destruction with our Cold War enemies. Why don't these silly fish, thieves, and commanders-in-chief just do what the mathematicians recommend?

DO IT AGAIN

The reason why cooperative action in the prisoner's dilemma is more common than the early game theorists would have predicted is that in real life, we can't insulate our choices from their effects on people's behavior toward us in the future. In real life, by and large, social organisms don't roam the world at random in search of new partners with whom they can play one-shot prisoner's dilemmas. Instead, they play a much smaller set of *iterated* games—multi-round games—with a much smaller set of partners. The individual fish in a group of guppies or sticklebacks know each other, so they can reward and punish each other in the next round of a predator-inspection "game" depending on how they behaved in this round. Most of humans' prisoner's dilemmas are iterated games too. We share refrigerator space with the same co-workers day in and day out. Therefore, it's in our best interest to keep the refrigerator clean so that our co-workers will appreciate our efforts and try to reciprocate. Any single round of play is really a game within a larger game.

It was a political scientist named Robert Axelrod who worked out the math of the iterated prisoner's dilemma. Axelrod suspected that when the prisoner's dilemma is extended from a one-shot game to an iterated game, the "all-defect" strategy, which is the Nash equilibrium for the one-shot game, might not work so well. Axelrod began to explore this possibility back in the late 1970s by asking his game theory colleagues to devise strategies for playing the prisoner's dilemma that would score the most points over many rounds of play.

Fourteen entries were submitted for the tournament. The simplest strategy (called tit-for-tat, which was submitted by the Canadian game theorist Anatol Rapaport) was so simple that

it only required four lines of code in the Fortran computer language. The most baroque strategy (submitted by a theorist who remained anonymous) required seventy-seven lines of Fortran code. Each of the strategies played against each of the others in five different matches. Each match was an iterated game consisting of two hundred prisoner's dilemmas. Every strategy's performance was based on the average number of points it scored in all of the matches that it played.

Axelrod found something surprising: the best performer was the humble (and concise) tit-for-tat. Tit-for-tat started the first round of each match by cooperating, and then it continued cooperating as long as its partner cooperated on the previous round. If tit-for-tat's partner defected on a given round, then tit-for-tat would defect on the next round. If the partner ever went back to cooperating, then tit-for-tat would resume cooperation as well. This is strikingly similar to the strategy that the Trinidadian guppies appeared to be using in cooperative predator inspection: start out nice, respond to cooperation with more cooperation, and respond to defections by defecting.[10] Axelrod had discovered something through computer simulation that guppies and sticklebacks had acquired through natural selection.

To make sure these results were reliable, Axelrod advertised the results of the first tournament and then solicited entries for a second tournament. Sixty-two strategies were submitted for the second tournament. Rapaport submitted tit-for-tat again. This time, the most complicated strategy was thirty times as long as tit-for-tat.

In the second tournament, all sixty-two strategies played against each other and against a strategy that chose cooperative moves and defective moves at random. Tit-for-tat won again. One of the interesting things about tit-for-tat's success in Axelrod's tournaments is that it wasn't particularly strong in any of the matches it played. Tit-for-tat performed so well over the long run not by beating its opponents into the ground, but by tying its opponents in many high-scoring matches. In fact, tit-for-tat can't beat *any opponent* because all it does is start out nice and then copy its opponent's moves. Tit-for-tat's strength comes not from the exercise of raw power but from its ability to encourage win-win behavior.

FROM ITERATED GAMES TO EVOLUTIONARY ONES

The history of game theory can be divided into two eras: Before Axelrod and After Axelrod. Before Axelrod, game theory was an effort to understand the behavior of people. After Axelrod came on the scene, game theory became a window into the behavior of all living things. Having conducted his two tournaments, Axelrod's next step was to turn the tournament into a computer simulation of "survival of the fittest." He took the sixty-two strategies submitted for the second tournament and allowed them to compete with each other, against themselves, and against a strategy that responded randomly. According to the rules that Axelrod established to simulate evolution, these sixty-three strategies existed in equal proportions at the beginning of the simulation, and at the end of each generation of competition, the strategies "reproduced" in proportion to their success in winning points during that generation. In this manner, strategies that were highly successful against their partners had a lot of offspring; those that were less successful had relatively few offspring. Axelrod thereby managed to simulate the evolutionary proposition that success breeds success.

Again, tit-for-tat came out on top. After a few hundred generations of competition and reproduction, most of the highly exploitable strategies (that is, strategies that didn't respond to defection by defecting in successive rounds) had been killed off by the nastier strategies (strategies that defected often or took advantage of cooperators). However, retaliatory strategies such as tit-for-tat could keep up with the nastier strategies by stalemating them in match after match.

After the nasty strategies had killed off all of the suckers, there was nobody left for them to pick on, so they picked on each other. As they did so, their numbers declined because their matches against each other were low-scoring matches; thus they failed to reproduce in sufficient numbers. Tit-for-tat continued to stalemate the nasty strategies, but in its competitions with itself, it could rack up high scores over and over again. As the nasty strategies suffered from low-scoring games and tit-for-tat benefited from high-scoring games when playing its twin, tit-for-tat increased its representation in the population.

Eventually, tit-for-tat became the most prolific strategy in Axelrod's computer-simulated ecology. By the end of the thousandth generation, almost 15 percent of the organisms were using tit-for-tat (recall that it constituted only 1 out of every 63 strategies at the outset), and its market share was still growing. Thus tit-for-tat appeared to be an evolutionarily stable strategy—evolution's version of the Nash equilibrium—that couldn't be overtaken by any other strategy. In Axelrod's semiconductor world, tit-for-tat seemed to be on its way to genetic immortality.[11]

WHAT MAKES TIT-FOR-TAT SO SUCCESSFUL?

Tit-for-tat has four characteristics that made it such a winner. First, tit-for-tat is a *nice* strategy: it always begins its matches by cooperating. As a result, tit-for-tat is always ready to benefit from mutual cooperation if its partner is similarly disposed. Second, tit-for-tat is a *retaliatory* strategy: if its partner defects in a given round, tit-for-tat will retaliate reflexively during the next round. By doing so, tit-for-tat prevents nasty strategies from capitalizing on its niceness. Third, tit-for-tat is a *forgiving* strategy: if tit-for-tat's partner returns to cooperation after a defection, tit-for-tat will also resume cooperation in the next round. Fourth, tit-for-tat is a *clear* strategy: it starts out nice and then repeats whatever its partner did on the previous round. In other words, it plays nice when its partners are playing nice, it's vindictive when its partners are playing nasty, and it's forgiving when its partners mend their nasty ways. That's it. Tit-for-tat doesn't overthink things. It's just the Golden Rule followed by the norm of reciprocity.

Should you retaliate when your cooperation partners harm your interests to maximize their own? Well, sometimes, yes. Tit-for-tat is a role model for this sort of eye-for-an-eye retaliation. Retaliation teaches your cooperation partners that they shouldn't try to take advantage of you and that they'd better not renege on their commitments. As we've seen in previous chapters, a propensity for tit-for-tat's style of revenge has evolved in many animal species as a method for enforcing social contracts.

But tit-for-tat also shows that if you want to be *really* successful, you can't hold on to that grudge forever. Tit-for-tat counsels

us that the key to long-term success—to say nothing of genetic immortality—is a willingness to forgive partners who defect but later return to cooperation. There's no point in holding a grudge: cooperation leads to much higher payoffs than does an interminable string of defections. So if you can get back to cooperation after defection, which tit-for-tat can, then your interactions will be high-scoring interactions for you and your partner both. A willingness to forgive periodic defections—if one's partner has demonstrated a desire to return to cooperation—is indispensable for creating win-win games.

In conducting these initial studies, Axelrod intentionally omitted some of the complications of evolution, not to mention the messy wow and flutter of real social interactions. But it's possible to add some messiness back into these computer simulations, and by doing so, to make them more realistic. As it turns out, the more we tweak these "tinker-toy models"[12] of evolution so that they more closely resemble the social conditions in which ancestral humans might have evolved their ultra-cooperativeness, the more they seem to favor the evolution of strategies that are even *more* forgiving than tit-for-tat.

BRING ON DA NOISE

Most of us, I think, have had the experience of trying to say something clever, and then watching in horror as our good-natured sentiment comes out of our mouths sounding rude or insulting. Even when we really are trying to play nicely with our cooperation partners, our best intentions sometimes backfire. Organisms of all shapes and sizes occasionally make mistakes in executing their cooperative strategies.[13] Game theorists call it "noise."

Noise—the possibility that you might accidentally defect when you mean to cooperate, or that your partner might read your genuinely cooperative behavior as a defection—is a big problem for tit-for-tat. Even when the error rate is very low—say, when there's only a 1 percent chance of making an error in implementing your intentions or in reading your partner's intentions—here's what happens when tit-for-tat plays its twin. Both strategies start out nice, but eventually one of them makes a mistake. Player A wants to cooperate but instead does

something that harms Player B, or else Player B misreads Player A's cooperative intentions. In either case, B will then defect against A as (what it considers to be) a justifiable act of retaliation. And because player A is also playing tit-for-tat, A will then defect on the following round in response to B's defection, and so on. Two players who started out playing the Golden Rule end up locked in an endless cycle of retaliation. Axelrod called this absurd scenario the *echo effect.* He recognized that when the echo effect is a possibility, tit-for-tat isn't forgiving enough.[14]

Martin Nowak and Karl Sigmund—a team of mathematical biologists—were among the first to investigate the evolution of cooperation in a noise-laden world. Like Axelrod before them, Nowak and Sigmund created a computer simulation of the evolutionary process. But unlike Axelrod, they allowed the organisms in this computer-simulated world to make mistakes in implementing their own rules and in interpreting the actions of others. Their simulation ran for millions of generations; every hundredth generation, they added mutant strategies to see if the mutants could invade the evolving system.

Nowak and Sigmund found, as Axelrod had, that evolution progressed in stages. Early on, the nastier strategies (those with a penchant for defecting) quickly established firm footholds in the population by winning lots of matches against nice partners (those that were a bit too keen to cooperate). However, after defeating most of the nice guys, the nasty strategies were left with nobody else to take advantage of. So they preyed upon each other. As they did so, they typically became locked into low-scoring game after low-scoring game, so eventually they too went the way of the dodo bird. The demise of the nasty strategies allowed tit-for-tat (which seemed to be an optimal compromise between strategies that are too nice and strategies that are too nasty) to increase its market share.

Tit-for-tat's days were also numbered, however. After many generations, even tit-for-tat began to struggle because of the small probability of making a mistake in implementing its intentions. Remember that even a single false move causes two players using tit-for-tat to become locked in a ridiculous cycle of negative reciprocity. This inability to overcome implementation errors was tit-for-tat's fatal flaw in Nowak and Sigmund's noisy world.

So what follows the reign of tit-for-tat? When they initially explored this question, Nowak and Sigmund crowned a strategy called "Generous tit-for-tat" as the ultimate evolutionary winner. Generous tit-for-tat, like its namesake, forgives a partner who returns from defection to cooperation, but Generous tit-for-tat has an additional feature: it grants forgiveness *unconditionally* (that is, without first requiring a subsequent round of cooperation from its partner) about one-third of the time. In other words, if you defect while playing a match with Generous, there is a one-in-three chance that Generous will turn the other cheek and continue cooperating. This tendency to forgive unconditionally keeps Generous from becoming hobbled by noise. Pleased with these results, Nowak and Sigmund wrote that after Generous appears on the scene in sufficient numbers, "Evolution then stops."[15]

But eighteen months later, they had to eat those words. More extensive simulations had subsequently convinced Nowak and Sigmund that tit-for-tat isn't the end of evolution. Instead, it merely paves the way for invasion by more unconditionally cooperative mutant strategies, which in turn paves the way for a reinvasion by mutant strategies that are more prone to defection, which takes the evolutionary process right back where it started. Undeterred, Nowak and Sigmund kept searching for an evolutionarily stable strategy. They eventually discovered that a strategy called "win-stay, lose-shift" could indeed become evolutionarily stable.[16] Win-stay, lose-shift always cooperates on the first trial. Then it follows the simple rule, "win-stay, lose-shift." That is, it repeats what it did on the previous round if it won the "temptation" payoff (the big payoff that comes from defecting while your partner cooperates) or the reward for mutual cooperation (which yields the second highest payoff), but it switches actions if it received the punishment for mutual defection (the next-to-worst payoff) or the sucker's payoff (the worst possible payoff). Win-stay, lose-shift seems to "learn" by viewing its bad outcomes as punishments and its good outcomes as reinforcers. Nowak and Sigmund gave it the nickname "Pavlov."

Pavlov forgives, but only after a mutual defection. If Pavlov and its partner both defected on the previous round, Pavlov "repents" and returns to cooperation on the next round. If Pavlov cooperated

on the previous round and its partner defected, Pavlov retaliates during the next round (recall that one-third of the time, Generous tit-for-tat will forgive in this situation). Therefore, Pavlov is clearly less forgiving than Generous tit-for-tat is.

Axelrod, of course, read Nowak and Sigmund's Pavlov paper. Having spent the previous fifteen years describing the virtues of tit-for-tat and its ilk, he felt duty-bound to defend Generous tit-for-tat's sullied honor. So Axelrod and a colleague ran yet *another* simulation to see if slight changes in Nowak and Sigmund's assumptions (I'll spare you the details) led to better evolutionary performance for Generous tit-for-tat. Indeed, Generous looked like a real contender for evolutionary stability when other strategies it encountered hadn't yet adapted to noise. But when the other strategies had already been winnowed on the basis of their ability to tolerate noise, the evolutionarily stable strategy was not Generous tit-for-tat. Instead, it was a form of tit-for-tat that, if it has defected without justification, allows its partner to defect in retaliation without itself retaliating in return. Axelrod and his colleague called it "contrite tit-for-tat."[17]

Change yet another assumption about how the game is played, and you get evolutionary stability for a strategy called "firm-but-fair."[18] Firm-but-fair starts out in a cooperative frame of mind, and then it makes its next decision about how to behave based on its most recent choice and its partner's most recent choice. If those two choices were cooperative ones, firm-but-fair keeps cooperating. If its most recent choice was cooperation but its partner's most recent choice was defection, firm-but-fair retaliates. However, if firm-but-fair's partner responds to firm-but-fair's defection with a retaliation of its own, firm-but-fair returns to cooperation. Finally, if firm-but-fair most recently chose to defect, and its partner most recently chose to cooperate, firm-but-fair returns to cooperation on the next round. Thus firm-but-fair is nice, vindictive, willing to let bygones be bygones, and responsive to mercy (an alternative interpretation is that it's unwilling to exploit suckers). Firm-but-fair will take one on the chin as punishment for its own bad behavior, and it readily accepts chastened sinners back into fellowship.

The fact that a few changed assumptions lead to such dramatically different evolutionary results might not inspire much

confidence that these simulations can tell us anything at all about forgiveness and evolution. However, what these studies do allow us to say is this: "all defect" isn't evolutionarily viable. Ever. And in a world with noise, even moderately forgiving tit-for-tat isn't evolutionarily viable. In fact, *all* of the strategies that have a claim to evolutionary stability forgive their partners' defections *some of the time,* and some of these strategies forgive a lot of the time. Generous tit-for-tat forgives unconditionally one-third of the time and always forgives if its partner has returned to cooperation after defecting. Pavlov forgives defection if it has defected too. Contrite tit-for-tat forgives righteous retaliation. Firm-but-fair is willing to forgive its partner if they both defected in the previous round and if a previously selfish partner ever returns to cooperating. Whatever the details of the evolution of human cooperation, it seems, the organisms that survived the evolutionary winnowing process had forgiveness in their cognitive toolkits.

EVOLUTION'S "THREE STRIKES" RULE

Still other twists on these evolutionary games deserve our attention, for they suggest that evolution may have shaped social creatures such as human beings to be hyperactively forgiving of a small circle of good friends and neighbors, but rather stingy in dispensing forgiveness to strangers and members of outgroups.

Let's start with a thought experiment. Open your address book and find the names of the eight people (other than relatives) with whom you have to cooperate most often to accomplish your daily goals (let's assume you're currently on good terms with all of these people). Now imagine that one of those eight people does something offensive or harmful to you. Perhaps your next-door neighbors go away for the weekend and leave their dog outside to bark for three days and two nights. Perhaps your office mate, with whom you work on a lot of projects, tells an embarrassing story about you at a staff meeting. Perhaps a roommate can't come up with her share of the rent for another week. Quick—how would you respond?

The odds that you'd retaliate harshly against these people are pretty slim. If your neighbors' dog barks all weekend, you're unlikely to throw eggs on their car or dump the contents of their

garbage can in the street. If a co-worker tells an embarrassing story about you, you're probably not going to spend a ton of time looking for a way to embarrass him at the next one. If your roommate is late with her share of the rent (assuming this is a first offense), you're not going to kick her out. What you'll probably do is (a) nothing, or (b) confront the offender in a constructive way and try to get the problem fixed. If you're really upset, you might sulk and avoid the person for a couple of days. But after that, most of us would continue on with the relationship as if nothing had happened.[19] For all intents and purposes, we'd forgive.

We tend to let these sorts of offenses go for three reasons. The first reason is trivial: we don't typically seek revenge against our cooperation partners in such situations because the harms they cause us are usually not very severe. The second reason, at this point in the chapter, should be easy to see: Axelrod and company have taught us that real-life interactions are noisy, and too-hasty retaliation in a noisy world creates unnecessary cycles of revenge and counter-revenge. One swallow doesn't make a summer, and one "defection" doesn't turn your cooperation partner into your enemy.

But a third reason why we avoid retaliating against our closest cooperation partners is that we're stuck with them (up to a point, anyway). If we turn these friends into enemies when they occasionally do things that harm us, then we'll too frequently be back out in the friendship market looking for new people to play our prisoner's dilemmas with, and good cooperation partners can be hard to find.

Granted, revenge sometimes has useful social effects. But if things get out of hand with my next-door neighbor—as anyone who has ever been involved in a dispute with a neighbor knows—home life can become astonishingly unpleasant. In the end, selling my house and moving somewhere else could become my only alternative for ending the rancor, and that's often highly impractical. If your relationship with a co-worker goes off the rails because you've decided to seek revenge, you'll have to find someone else to help you get your projects done. Even finding a new roommate involves the costs of putting an advertisement in the paper, the effort of interviewing a bunch of people, and the discomfort that comes with having to break in a new roommate. Better the devil you know than the devil you don't. For this reason, we tend to be especially forgiving of the people with whom we share our daily lives. In fact, evolutionary simulations suggest that when dealing with good friends and neighbors, we may use a sort of "three strikes (or four, or even five) and you're out" rule.

Patrick Grim, a philosophy professor who tinkers with computer simulations and game theory, was the first to figure this out. He noted that in real life, living things reside somewhere in two-dimensional space. We've all got addresses—spots on the face of the earth where we can be found working, shopping, eating, and sleeping. And we've got company—a relatively small set of friends, neighbors, and co-workers with whom we conduct most of our important day-to-day interactions. Grim wanted to capture this reality of social relations, so he tweaked the noisy prisoner's dilemma so that the organisms all resided at specific points on the surface of a two-dimensional map. Picture a 64×64 grid made up of 4,096 squares. If you were one of the pixilated creatures in Grim's simulation, you would occupy one of those squares, and your square would be touching the squares of the eight people with whom you can cooperate.

Grim added a second bit of realism to his simulation: at the end of each round of play, he had the organisms look around to see how their neighbors were doing, and then they all changed strategies to mirror those of their most successful neighbors. If the organism itself did better than any of its neighbors, it stuck with the strategy that it had been using. This new little wrinkle was an important one for Grim to add because many organisms—humans, nonhuman primates, and birds, for example—are good at learning new tricks from other individuals. The human mind learns new innovations as quickly as it does because it has acquired a few learning biases, or rules of thumb. One of those rules of thumb is "copy the person who does it better than you" or "copy the successful."[20] So it's reasonable to imagine our ancestors sorting out their strategies by looking around them to see who does well, and then simply copying the one neighbor who fares the best.

With these new conditions in place—giving each strategy an address (and some neighbors) and a "copy your most successful neighbor" rule—Grim found an evolutionarily stable strategy that's even more forgiving than Generous tit-for-tat. Remember Generous tit-for-tat, which unconditionally forgives about one-third of the time? In Grim's analysis, the evolutionarily stable strategy forgives unconditionally about *two-thirds* of the time. In other words, for every three unjustified slaps it received, it turned the other cheek to two of them.[21]

More than a decade after Grim published his results, two other evolutionary modelers came to a similar conclusion without even having known about Grim's earlier work. Dan Hruschka is a newly minted anthropologist (he was finishing up a post-doc at the time of this writing), and Joseph Henrich is an anthropology professor at Emory University. I won't even try to summarize the technical details of their work, but suffice it to say that they started with the assumption that social creatures have a social instinct to form "cliques"—small groups of good buddies who will prefer each other as cooperation partners over other potential partners. When you make this assumption, what evolves is a complex set of decision rules that dictate the following: "If the person who hurt you is one of your good buddies, forgive that person about 80 percent of the time. When playing against strangers, defect most of the time, unless you are short on good buddies. If you are, then take chances with strangers and start out by playing nice to see if you can turn those strangers into friends."[22]

Assuming that these modeling efforts by Grim, and now Hruschka and Henrich, really do a better job of simulating the ecological and social conditions in which humans evolved than do the models that came before (and I think they do), then it's fair to conclude that humans may have evolved to be almost promiscuously forgiving of their neighbors and good friends. The take-home message is clear: forgiveness isn't something that evolved to smooth over our relations with just anybody. Instead, it exists in large measure to help us preserve a relatively small number of geographically close neighbors or a small clique of trusted associates. According to these models, forgiveness helps us develop a social environment in which we can benefit from direct reciprocity.

ENTER GOSSIP

All the same, we do sometimes forgive people with whom we aren't friends or close relatives—even people who haven't distinguished themselves as particularly good cooperation partners. Theoretical biologists can explain this, too, but to do so, they go beyond concepts such as neighborhoods and social learning so that they can invoke other characteristic features of human society—features such as social norms, reputation, and, most surprising of all, gossip.

Before language, humans' ancestors had to rely on nonverbal means for learning about others. If you wanted to know what someone was like—whether she was aggressive, docile, selfish, timid, generous, trustworthy, forgiving, or whatever—you had to learn it for yourself, either by relating to that individual directly (that is, by learning the hard way), or by observing with your own eyes how that person treated other people. However, once language came online, new options opened up. Most important, it became possible to learn about others by talking to other people about them. There was no more need to rely on learning the hard way, no more need to rely on seeing things with your own eyes. Gossip became an option for figuring out what other people were like, and consequently, for figuring out how to interact with them. The evolutionary biologist Robin Dunbar has proposed that this evolved facility with language explains why humans' social groups exploded in size relative to the size of the social groups in which other primates live.[23]

Through the social sharing that language makes possible, people acquire reputations. These reputations have cash value: if you have a good reputation, people will be inclined to cooperate with you and treat you with respect. If you have a bad reputation, people will steer clear of you or actively work against you. Some theoretical biologists have surmised that reputation might enable cooperation to evolve in a population of self-interested individuals—even if those individuals interact with each other only once. In this evolutionary scenario, reputations are used to establish cooperation not through direct reciprocity (as when two people play an iterated prisoner's dilemma game), but through *indirect* reciprocity. Under indirect reciprocity, if you defect against me during our prisoner's

dilemma, I won't retaliate against you by defecting the next time we play each other (as has been the case in all of the simulations we've been discussing until now), because there won't be a next time. Instead, I'll retaliate with my mouth: I'll tell everybody (that is, all of your future cooperation partners) what a scammer you are, which will cause them to treat you differently in the future. If you help me by cooperating during our prisoner's dilemma, I'll return the favor by telling everybody what a mensch you are.

This leads to an interesting change in how the prisoner's dilemma influences fitness. After every round of play, your fitness changes according to the usual prisoner's dilemma contingencies (if you defect against a cooperator, you earn more than if you cooperate with a cooperator, and so forth), but it also changes according to how your reputation is affected. Thus the effect of any single choice on your fitness might have more to do with how its reputational consequences influence how people treat you in the future than with its direct costs and benefits.

In this artificial world that the indirect reciprocity theorists envision, there are also complex reputational dynamics at work. You enter an interaction with a good or bad reputation, and you interact with someone who has a good or bad reputation. These social facts shape your behavior and your partner's behavior during the game, which affects not only your immediate outcome but also your reputation after the game. With a potentially new reputation in place, you might make a different choice in your next prisoner's dilemma, which would in turn influence your subsequent reputation, and so on. Through this dynamical process by which behavior affects reputation, which in turn affects behavior, ad infinitum, evolution causes norms to emerge. These norms dictate two things: the choices that people should make in their own prisoner's dilemmas, and the rules by which people's behavior influences their subsequent reputations.

In a system such as this—with all of the possible prisoner's dilemma strategies and all of the possible rules for how to respond to people with different types of reputations, a staggering number of combinations are possible—4,096 unique pairs of dynamics and strategies, to be exact. Which ones become evolutionarily stable—while also yielding high payoffs for cooperation? Two biologists at Kyushu University in Japan

explored all 4,096 possibilities. As usual, they assumed that errors sometimes happen—that people can accidentally defect when they mean to cooperate, and that people can accidentally mistake someone with a "good" reputation for someone with a "bad" one, and vice versa.

The scientists found that under these conditions, only eight social norms become evolutionarily stable while providing high payoffs for cooperation. I don't have to describe each of these "leading eight" individually to convey how the entire package works. First, if you have a good reputation and you meet someone with a good reputation, you should cooperate. You'll both benefit from mutual cooperation and you'll both maintain your positive reputations. Second, if you have a good reputation and you interact with someone who defects—no matter what his or her reputation—that person is assigned a bad reputation for the next round. Defecting against a good person earns you a scarlet letter. Third, if you have a good reputation and you interact with someone with a bad reputation, you should defect. This punishes people for the selfish actions that led to their bad reputations in the first place, and you get to keep your good reputation. Fourth, if you have a bad reputation and you interact with someone who has a good reputation, you should cooperate. No matter whether your partner cooperates or defects, your good behavior will restore your good name in the next round.

Isn't that a tidy little ethical system? Good people should cooperate with other good people. People who defect against good people lose their good reputations and should be punished. If you have a good reputation and you choose to punish someone who has a bad reputation, it's credited to you as righteousness. If you have a bad reputation and you cooperate with someone who has a good reputation, expect punishment because of your bad reputation, but take heart because you can look forward to forgiveness after that. Forgiveness is an evolutionarily vital part of this ethical package because there has to be a way to restore people to good standing so that they'll be motivated to return to cooperation with all of the other cooperators in the population. If forgiveness weren't available, the average gains of cooperation would slowly decline in the population with each successive generation.

What's so striking about this set of norms is how intuitive it all seems. Those gossiping, status-conscious, computer-dwelling organisms sound just like humans! The intoxicating possibility here is that these norms seem so intuitive to us because our social instincts really *did* evolve in the way that these results suggest. Of course life is more complicated than any world we can simulate with zeroes and ones. But perhaps ethical systems such as this one, which are tough on defectors while nevertheless creating a way for bad people to be forgiven, are the inevitable outcome within populations of evolving, gossiping individuals who need each other's help but can't always use direct reciprocity to get it.[24]

FORGIVENESS IS THE BRIDESMAID; COOPERATION IS THE BRIDE

I think it's only fitting that these evolutionary simulations, with their iterations and their noise, have led to seemingly interminable cycles of scientific inquiry with no clear end in sight. Even though Bob Axelrod has been thinking about the evolution of cooperation for over a quarter of a century, he tells me that he's still not sure what to expect from natural selection. Nobody is. What you get from these models depends utterly upon your basic assumptions about how the games are played and the additional faculties with which you endow your game-playing organisms. Even so, studies with living, breathing, sentient human beings show that Pavlov and Generous tit-for-tat—two moderately forgiving strategies—are quite popular in real life, and that people fare quite well when using them or when playing with partners who do.[25] And it bears repeating that all of the strategies in the running for evolutionary stability tend to forgive at least *some of the time* and some of them forgive *a lot of the time*—especially when playing with good buddies and neighbors, and especially when reputations and gossip enter the picture.

If we distill all of these insights down to the bare essentials, we end up with a sort of recipe for the evolution of forgiveness. First, put some game-playing organisms together in the same niche. Second, let the organisms play one-shot games with lots of their neighbors, but also let them play iterated games with a more restricted set of partners who live close by. Third, make them cliquish creatures that prefer to limit their iterated endeavors to a small circle of trusted neighbors or friends. Fourth, let the organisms make occasional mistakes in implementing their intentions and in reading the intentions of others. Fifth, give them the ability to learn by observing what works for their neighbors. Sixth, give them communicative powers so they can tattle on each other and sing each other's praises. Let this blend cook for many generations, and you're likely to end up with creatures who are almost manically forgiving of their good friends and neighbors, and who are even willing to cut a break for a reformed sinner.

The evolutionary theorists who do this research, like all celebrity chefs, make cooking look easy. But is this recipe for forgiveness really any good? If so, it ought to have produced modern-day species that have a penchant for forgiving as a way of keeping their cooperative relationships intact. Predator-inspecting guppies are retaliatory, to be sure, but remember that they're also willing to forgive the repentant slacker. Are there other species that natural selection has endowed with a similar propensity to forgive? Given the evolutionary advantages that can accrue to organisms that are inclined to forgive under certain circumstances (those three words again), we should expect that many other flesh-and-blood creatures that populate our planet today—particularly those that benefit from cooperation with a long-standing group of associates—are inclined to forgive. As it turns out, such creatures aren't hard to find.

CHAPTER SIX

THE FORGIVENESS INSTINCT

"The discoverer of the role of forgiveness in the realm of human affairs was Jesus of Nazareth," wrote the political philosopher Hannah Arendt. "The fact that he made this discovery in a religious context and articulated it in a religious language is no reason to take it less seriously in a strictly secular sense."[1] In a 2003 interview, the late author Kurt Vonnegut made essentially the same observation: "Two radical ideas have been introduced into human thought. One of them is that energy and matter are pretty much the same sort of stuff. That's Einstein. The other is that revenge is a bad idea. Revenge is an enormously popular idea but, of course, Jesus came along with the radical idea of forgiveness. If you're insulted, you have to square accounts. So this invention by Jesus is as radical as Einstein's."[2]

If you buy into the disease metaphor of revenge that I introduced back in Chapter One, then Arendt's and Vonnegut's "creation story" may sound right on target to you. If revenge really is a disease, then maybe it's appropriate to imagine a time, long ago, when a revenge-weary human race was struggling under the burden of its own vengeful impulses, just waiting for some Einstein

of the moral realm to arrive on the scene and come up with a solution to the problem—to "discover" or "invent" forgiveness.

However, if you find yourself buying into the evolutionary account of forgiveness that we've been exploring, then you have to take the "Arendt-Vonnegut hypothesis" with a grain of salt. According to the evolutionary perspective, forgiveness isn't an idea that Jesus or anybody else had to invent (although, as we'll see in a couple of later chapters, Jesus did have some things to say about forgiveness that were pretty innovative). The evolutionary perspective holds out the possibility that organisms that are motivated purely by self-interest can develop a penchant to forgive solely through the action of natural selection. Forgiveness enables them to promote their own inclusive fitness by avoiding a self-defeating tendency to over-harshly punish their genetic relatives, and it enables them to forge ahead with their bumbling efforts to cooperate with each other, despite the mistakes that they and their cooperation partners will inevitably make. The simplest social organisms we can imagine—creatures whose only social instincts can be summarized in a few lines of computer code—may be capable of "discovering" or "inventing" forgiveness solely through natural selection's action upon their behavioral repertoires, without any help from moral exemplars, religious leaders, political theorists, or novelists, thank you very much.

LET'S GET REAL

Theoretical biologists' research on the prisoner's dilemma certainly leaves one with the impression that natural selection could have programmed a propensity to forgive into many animal minds, including human minds, but aside from some fish playing a tit-for-tat strategy during their predator-inspection games, we haven't yet seen much scientific evidence for a "forgiveness instinct" in real, living creatures—certainly not any in human beings. Theoretical biology shows that forgiveness *might* have evolved to help organisms achieve fitness by promoting the fitness of their genetic relatives and by cooperating with nonrelatives, but did it *actually* evolve in this way? Do human beings or any other organisms really have a built-in penchant to forgive (friends and family in particular), courtesy of natural selection?

We can't go back in time to see how forgiveness evolved, but all is not lost. If natural selection acted upon ancestral humans to turn us into a species with a natural inclination to forgive our family members and cooperation partners, then we should find evidence for it by looking at how, when, where, and why humans and other social animals engage in the business of forgiveness today.

In the past three decades, animal researchers have made a profound and unexpected discovery that has revolutionized our understanding of conflict, aggression, and peacemaking: when group-living animals get into aggressive conflicts with their kin and their friends, many of these conflicts end with friendly reconciliations. This discovery has tremendous implications for understanding the human capacity for forgiveness.

A brief semantic digression into the concepts of "reconciliation" and "forgiveness" now seems inescapable.

RECONCILIATION AND FORGIVENESS

Philosophers and scientists often bend over backward to make fine-grained distinctions between the concepts of reconciliation and forgiveness, but for the purpose of trying to understand the evolutionary origins of forgiveness, I think that the laborious distinctions are mostly a tempest in a teapot. Reconciliation and forgiveness aren't conceptually *identical*, but they probably have the same evolutionary roots.

In defining forgiveness, scholars usually focus on the idea that when people forgive an offender, they come to feel less vengeful and less bitter, and they experience the return of positive motivations and good will—perhaps even love—toward the offender. Forgiveness, therefore, is a private process of getting over your ill will and negative emotions, and replacing those "negatives" with "positives" such as wishing the offender well or hoping for a new and improved relationship. Of course, these motivational and emotional changes often lead to better behavior toward the offender. If you've forgiven somebody, at a minimum you've stopped wanting revenge and you wish that person well, at least in a limited sense. You might not want to invite the person who injured you over to your place for a barbecue, but you don't crave

a slow, painful death for that person, either. This is the bare-bones forgiveness of the prisoner's dilemma, with some positive feelings and intentions added into the mix.

Reconciliation, many scholars insist, is a different kettle of fish. Biologists define reconciliation in a straightforward, behavioral way: a "friendly reunion between former opponents" that "supposedly serves to return the relationship to normal levels of tolerance and cooperation."[3] Primatologists don't focus on intentions or motivations or feelings when defining reconciliation, perhaps because chimpanzees aren't very good at filling out questionnaires. Psychologists, however, who *do* consider people's feelings and intentions along with their actions, tend to define reconciliation as the restoration of a fractured relationship that happens because the victim has forgiven the offender *and* because the offender has mended his or her evil ways.[4]

So forgiveness is an internal process of getting over your ill will for an offender, experiencing a return of good will, and opening yourself up to the possibility of a renewed positive relationship with the offender. Reconciliation, in contrast, is a friendly reaching out to the person who harmed you (or the person whom you've harmed) that's supposed to fix the relationship breach. Or, if you're a psychologist, it's a relationship that's been repaired because the victim forgave and the offender repented.

I suppose this forgiveness-reconciliation distinction isn't completely useless. For one thing, it acknowledges that a friendly-seeming gesture from somebody you betrayed last month doesn't necessarily mean that all has been forgiven. Maybe I'm being nice to you in hopes of lulling you into complacency so that I can seek my revenge against you at a more opportune time. I can treat you nicely and still hate you. There's also a second reason why the forgiveness-reconciliation distinction might be useful: "forgiveness" is a morally loaded concept—"good people" are supposed to be forgiving. For this reason, it's nice to define forgiveness in such a way that people can be "forgiving" (by releasing their vengeful impulses and by wishing the offender well), even if it's ill-advised for them to resume relationships with offenders who haven't shown any remorse or any desire to change their nasty behavior (which would make reconciliation difficult and perhaps even dangerous). Defining forgiveness as something private

therefore allows people to be "forgiving" without also having to be doormats.

However, reconciliation and forgiveness have much in common. Getting over your grudge and starting to feel positively again toward someone who harmed you ("forgiving") must surely be one of the most important psychological causes of reconciliation. If you had to guess why two people have had a "friendly reunion" that supposedly returned "the relationship to normal levels of tolerance and cooperation," you'd usually be right to guess that it was because the victim had forgiven the transgressor. Indeed, relationship restoration is probably the most basic social effect of forgiveness.[5] Similarly, if a fractured relationship hadn't returned to "normal levels of tolerance and cooperation," you'd suspect that one of two ingredients was missing: either the victim hadn't forgiven or the offender hadn't repented (or both).

All to say that reconciliation, at least in humans, seems to be the point of forgiveness. If natural selection really did outfit people with an ability to forgive (you can be the judge of that soon enough), it wasn't because natural selection was concerned with cheering people up or helping them to release their pent-up negative emotions (even though forgiveness often does exactly that). Instead, as we've seen, the main adaptive function of forgiveness seems to be helping individuals preserve their valuable relationships.[6] Because forgiveness so reliably precedes reconciliation in human beings, it seems safe to hazard a guess that something like forgiveness (an internal motivational and emotional change) precedes reconciliation in nonhuman animals too, if only we could measure it.

Well, we can't—not exactly, anyway. Nevertheless, by giving some close attention to post-conflict behavior in nonhuman primates and other mammals, we'll start to get a picture of something that looks an awful lot like a "forgiveness instinct."

WOLFGANG KÖHLER'S LITTLE RAP

In 1913, the gestalt psychologist Wolfgang Köhler was appointed director of the Prussian Academy of Science's primate research station, which was located on the island of Tenerife in the Canary Islands. In 1917, after several years of research at the station, he published a book on the nature of intelligence in nonhuman primates, *The Mentality of Apes*. In a lengthy postscript, Köhler goes beyond the book's primary subject matter and provides a detailed account of chimpanzees' social sensibilities. Here Köhler describes the reaction of a young chimpanzee that he had just punished with "a little rap" for repeatedly snatching food away from a weaker chimpanzee. What follows is, as far as I know, the first suggestion in the entire field of biology that nonhuman primates might possess a forgiveness instinct:

> The little creature, which I had punished for the first time, shrank back, uttered one or two heart-broken wails, as she stared at me horror-struck, while her lips were pouted more than ever. The next moment she had flung her arms round my neck, quite beside herself, and was only comforted by degrees, when I stroked her. *This need, here expressed, for forgiveness, is a phenomenon frequently to be observed in the emotional life of young chimpanzees* [italics mine]. . . . Even animals, who when they have been punished, at first boil with rage, throw one glances full of hate, and will not take a mouthful of food from a human being: when one comes again after a time will press up close . . . pressing one's fingers affectionately between their lips and making all other protests of friendship.[7]

It's a cute little anecdote, but it didn't exactly set the agenda for the next generation of research on primate social behavior. Chimpanzees that are raised around humans are different from those in the wild. For this reason, Köhler's story of a juvenile chimpanzee that wanted a hug from a caregiver after a conflict was probably interpreted by other researchers as no more than a behavioral oddity that arose from the animal's close contact with humans—if the story was noticed at all. It would be six more decades before scientists were ready to seriously consider the possibility that forgiveness and reconciliation might have a place in the social repertoire of a nonhuman animal.

THE KISS

The year 1979 was a watershed for scientific research on forgiveness. At the University of Michigan, a young associate professor named Robert Axelrod was soliciting entries for his prisoner's

dilemma tournaments, and research to come from this work would soon reveal the indispensability of forgiveness for the evolution of cooperation. On another continent, a young primatologist named Frans de Waal was publishing the first scientific study of reconciliation in nonhuman animals.

Four years earlier, de Waal had begun a postdoctoral fellowship at the Arnhem Zoo in The Netherlands—one of the largest colonies of captive chimpanzees in the world. One November day in 1975, de Waal noticed a male and a female chimpanzee kissing each other. Chimpanzee kissing is not particularly rare, but what struck de Waal on this particular occasion was that only moments previously the male had attacked the female during a showy display of his physical strength and dominance. Even more unusual was the noisy ruckus that had erupted in the colony. As de Waal recounts the episode, "Suddenly the entire colony burst out hooting, and one male produced rhythmic noise on metal drums stacked up in the corner of the hall. In the midst of this pandemonium, two chimpanzees kissed and embraced."[8]

Why would two chimpanzees engage in affectionate contact just moments after one had assaulted the other? De Waal wondered if the friendly post-conflict interaction was intended to help them to undo the damage that the assault had inflicted on their relationship. Could it be that chimpanzees kiss and make up in the same way that people do?

De Waal was picking up where Köhler had left off, and he suspected he was onto something important. To verify that chimpanzee reconciliation was real, he began to gather proper scientific evidence. In 1979, de Waal and a co-worker published results from their observations of the Arnhem colony. They had discovered that friendly behaviors such as kissing, submissive vocalizing, touching, and embracing were actually quite common after chimpanzees' aggressive conflicts. In fact, they were the chimpanzees' *typical* responses to aggressive conflicts. The researchers observed 350 aggressive encounters and found that only 50, or 14 percent, of those encounters were preceded by some sort of friendly contact. However, 179, or 51 percent, of the aggressive encounters were followed by friendly contact. This was a staggering discovery: friendly contact was *even more common* after conflict than it was during conflict-free periods.[9] This finding

sent a shock wave through the small community of researchers who studied primate social behavior.

The methods for measuring reconciliation have become more sophisticated since de Waal started this area of research back in 1979. Today, primatologists calculate the "Conciliatory Tendency," or CT. CT values range from 0 to 1. Values of 0 mean that the animals from a certain group aren't any friendlier toward each other after conflicts than they are during peacetime. In contrast, if the animals within the group had friendly encounters after every conflict, but never during peacetime, their CT would be 1. However, most primates do have a fair amount of friendly contact during nonconflict episodes (many primates spend hours each day grooming each other), so primate groups' CT scores rarely exceed .50.[10]

A Law of Attraction

Chimpanzees' CT estimates have ranged from a low of around .18 to highs in the .40s;[11] among chimpanzees in the wild, the estimates are toward the low end of that range.[12] These figures may look small (remember that they theoretically can go from 0 to 1), but they're bigger than zero, which means that friendly contact is *more* likely after conflict than it is during peacetime. What we ought to conclude from these figures is that chimpanzee conflicts lead most commonly to friendly contact, not interminable cycles of revenge or alienation. Wrangham and Peterson's "demonic" chimpanzees and de Waal's "good-natured" chimpanzees are one and the same. Conflict and aggression among chimpanzees (if they're from the same living group—which is a big "if") don't cause the combatants' relationships to end, and they don't cause them to become locked in interminable feuds. Instead, conflict and aggression seem to make combatants *more attractive* to each other. It seems perverse, but it's true.

Chimpanzees aren't the slightest bit unique in this respect. Other great apes, such as the bonobo and the mountain gorilla, also reconcile.[13] Several peaceable macaque species have conciliatory tendencies at least as high as those of chimpanzees. Even rhesus macaques, which are renowned for their nasty temperaments (on average, they're involved in eighteen aggressive

episodes during every ten hours of observation,[14] and in my limited experience, they seem to love nothing better than throwing poo at visitors), show a tendency to reconcile after conflicts. Indeed, of the thirty or so primate species that have been studied, only a select few (for example, the ring-tailed lemur and the red-bellied tamarin) appear not to reconcile. Each reconciling species has its own signature style: chimpanzees kiss and embrace, bonobos partake of a seemingly endless variety of sexual activities, stumptailed macaques show each other their rear ends,[15] and baboons grunt at each other.[16] Most species also use heavy doses of grooming to patch things up. Humans are hardly the only primates that rely on hugs, backrubs, and make-up sex to iron out their conflicts.

It gets more interesting still, for reconciliation isn't even limited to primates. Goats, sheep, dolphins, and hyenas all tend to reconcile after conflicts (rubbing horns, flippers, and fur are common elements of these species' reconciliation gestures). Of the half-dozen or so nonprimates that have been studied, only domestic cats have failed to demonstrate a conciliatory tendency.[17] (If you own a cat, this probably comes as no surprise).

THE CONCILIATORY TENDENCY IN HUMANS

How does the conciliatory tendency of human beings compare to those of other species? Unfortunately, there isn't a single study of adult human beings that would allow us to directly compare humans and nonhumans. However, reconciliation has been studied in human children, and these studies clearly show that children as young as three or four have a strong conciliatory tendency. The CTs of young children from many countries (including Russia, the United States, and Japan) hover around a very respectable, chimpanzee-ish .40. A study of six- and seven-year-old children from the peace-loving Kalmyk (a minority ethnic group in Russia, descended from the Mongols, who practice a form of Buddhism and are known to be incredibly pacific) revealed a CT of .7, making Kalmyk children the most actively conciliatory critters of any species known on the planet today.[18]

The strategies that preschool children use for reconciling conflicts are a lot like the ones that we adults use when we've offended someone at work, angered a neighbor, or hurt our spouse's feelings. They explicitly apologize, invite each other to resume playing, offer to share the objects or goodies that they were fighting over, hug each other, and hold hands. These same basic strategies are used by children across cultures, but there's some cultural variation, too. Japanese preschoolers use apologies as their main strategy, whereas Swedish preschoolers use "invitations to play" as their main strategy,[19] and so on. But these cultural differences are trivial. What really matters is not how preschoolers' reconciliation gestures differ, but how they're all the same: little tykes from around the world seem to be working from the same basic palette of options for resuming positive relations after they've hurt each other's feelings.

This still leaves us wondering whether reconciliation, forgiveness, or both are universal among human adults. Until now, scientists have left one important stone unturned in their search for an answer.

FORGIVENESS AND RECONCILIATION: HUMAN UNIVERSALS?

In Chapter Four, I told you about Martin Daly and Margo Wilson's survey of the ethnographic data on a representative sample of sixty distinct world cultures, which showed that blood revenge following homicide is a "statistical universal." Their research showed that blood revenge has emerged as an important social phenomenon in 95 percent of the cultures they examined. This fact supports the notion that the human propensity for revenge is a product of evolution: if violent revenge were merely a "cultural artifact," rather than an intrinsic attribute of human nature, then why does it pop up in virtually every culture?

Daly and Wilson's results led me to wonder about the ethnographic data on those same sixty cultures and the story they might tell about the cross-cultural universality of forgiveness and reconciliation. After examining those ethnographic data, I discovered that the concepts of forgiveness, reconciliation, or

both had been documented in fifty-six, or 93 percent, of the sixty cultures in the HRAF Probability Sample. The only four cultures in this sample whose tendencies to forgive or to reconcile have escaped anthropologists' notice are the Chukchee of the Arctic Circle, the Bororo of Brazil, and the Pawnee and Klamath Indians of North America. Across cultures, the concepts of forgiveness and reconciliation were considered appropriate in a variety of relational contexts, including spousal relations (the most common relational context in which forgiveness and reconciliation were discussed), relationships between children and their parents, relationships between warring communities, and relationships between neighbors embroiled in the mundane conflicts of daily life.

Is it possible that forgiveness and reconciliation really didn't exist among the Chukchee, Bororo, Pawnee, and Klamath? Sure, I suppose anything is possible. Or perhaps the anthropologists who studied those cultures just failed to notice the forgiveness and reconciliation that was occurring right under their noses. This second possibility is much more plausible. The evolutionary biologist David Sloan Wilson has observed that "It is actually difficult to find descriptions of forgiveness in hunter-gatherer societies, not because forgiveness is absent but because it happens so naturally that it often goes unnoticed."[20] I think Wilson may be correct, and not just about hunter-gatherers but about all cultures. Forgiveness and reconciliation may be so common and so taken for granted by anthropologists as to be regarded, quite literally, as nothing to write home about.

In either case, the 93 percent hit rate for evidence of forgiveness and reconciliation in those sixty cultures is tantalizingly close to the (somewhat arbitrary) 95 percent threshold that the anthropologist Donald Brown proposed as a standard for concluding that a behavior or psychological process is a "statistical" human universal.[21] I'm inclined to think that for subtle and often private processes such as forgiveness and reconciliation, a .930 batting average is close enough to the .950 mark that we're safe in treating these results as supportive of the idea that forgiveness and reconciliation really are standard-issue social instincts. Granted, revenge is probably a human universal, but reconciliation and forgiveness seem to be universal as well.[22]

The methods that people from these fifty-six "forgiveness-cultures" use for seeking forgiveness, granting forgiveness, and reconciling are fascinating, both for the commonalities across cultures and the diversity across cultures. Bottom-holding and horn-rubbing were found to be in short supply, but public apologies, gift exchanges, attempts to compensate injured parties, animal sacrifices, religious rituals, and third-party mediation are common elements of forgiveness and reconciliation in many cultures.

Of course, diversity abounds as well: the mystical Dogon people of Mali, for example (one of the three societies for which, you might recall, Daly and Wilson were unable to find evidence of blood revenge), have a wide variety of social mechanisms for making forgiveness happen. These include a ritual in which a contrite offender clasps the ankles of the person harmed, a ritual in which the perpetrator takes three bites from a piece of charcoal and then spits them back out in the presence of the victim, and the intervention of third parties who actively work to effect a reconciliation between feuding parties.[23] According to a Serbian tradition, the Sunday before the beginning of the Christian season of Lent was called "Forgiveness Day"—a day when young people were supposed to go around to their elders in order to mend any quarrels that had accumulated during the previous year.[24] The ethnographic evidence shows that even the Yanomamö people of Venezuela and Brazil, renowned among social scientists for their bellicosity rather than for their peacemaking prowess (thanks in large measure to the writings of the anthropologist Napoléon Chagnon[25]) have the potential to work out their violent conflicts in a conciliatory way, *under certain circumstances.*[26]

Needless to say, the proposition that forgiveness and reconciliation are human universals doesn't imply that forgiveness and reconciliation are practiced the world over in the same way, or with the same frequency—not any more than the fact that every society has a language implies that they all use Shona or Urdu or Aramaic or Esperanto. What's universal across cultures about language is that every culture has one. Likewise, although there are cultural differences in *what* people are willing to forgive, and *how* they go about doing it, it seems a fairly safe bet that human beings from every culture understand the concepts of forgiveness and reconciliation, appreciate the value of these processes, and

under the right social conditions, will take the time and trouble to put them to use.

WHY *DO* GROUP-LIVING ANIMALS FORGIVE AND RECONCILE?

Biologists have introduced two hypotheses to explain why most group-living animals (humans included) have developed propensities to forgive or reconcile with each other. The first of these hypotheses, which has been championed by the UCLA anthropologist Joan Silk, is that animals reconcile in order to signal that they're sick of fighting and are ready to start treating each other nicely again.[27] According to this hypothesis, which Silk calls the "benign intent" hypothesis, the function of reconciliation is to convey to former enemies that they can drop the defensive attitude, lay down their arms, and resume lives of peace. By Silk's lights, then, reconciliation delivers two freedoms: first, it delivers a freedom from fear, second, it delivers a freedom to resume normal peaceful relations.[28]

The second hypothesis, espoused by Frans de Waal and many other primatologists, is the "valuable relationship" hypothesis: animals reconcile because it repairs important relationships that have been damaged by aggression. The very act of being nice to each other after a conflict "undoes" the relational damage that the aggression caused. By undoing this damage, the animals can preserve the relationships upon which they rely for their own fitness.[29]

If I had to pick *just one* of these two hypotheses (although, best I can tell, they're not really mutually exclusive), I think I'd go with the valuable relationship hypothesis. The idea that reconciliation gestures have the function of helping animals to restore their valuable relationships fits neatly with three independent lines of evidence. First, it squares with what the theoretical biologists have been saying about the adaptive value of forgiveness: ancestral organisms that were willing to forgive their kin ended up with better inclusive fitness than did those organisms that couldn't resist taking their pound of flesh whenever a genetic relative harmed them. And as the countless computer simulations

show, organisms that were willing to forgive their cooperation partners were better at gleaning the benefits of cooperation. The evolutionary deck is stacked: natural selection leads self-interested organisms toward the acquisition of behavioral processes that allow them to forgive so that they can benefit from cooperative friendships and family relationships. This is exactly what the valuable relationship hypothesis says.

Second, the valuable relationship hypothesis fits nicely with the fact that the most conciliatory animals are also highly groupish. The great apes (excluding orangutans), lots of macaques, and many other mammals such as goats, dolphins, and hyenas are bound to their groups in important ways. They simply can't survive on their own in the wild because natural selection has made them interdependent. For example, group-living apes and monkeys assist each other in finding food, grooming, alerting each other to predators, raising young, climbing the social ladder, and hunting. Among dolphins, cooperation occurs in the context of reproduction (males work together to isolate females for sex). Goats and sheep rely on the other members of their herds for safety in numbers against predators.

Which brings us back to those nonreconciling domestic cats. Cats' only natural social groups are their birth families. Yes, a bunch of unrelated cats might look like a group because they live under a single owner's roof, eat from the same dish, scratch at the same post, and play with the same little toy mouse with the bell inside, but adult cats don't *need* other adult cats for much of anything. They're one of the few mammals whose conciliatory tendencies have been studied to date that truly are "bowling alone." And that's why they don't reconcile after conflicts.

The idea that some species became conciliatory as an adaptation to group living has received a special kind of direct support. In a head-to-head comparison of the conciliatory tendencies of two species of monkeys, de Waal and a colleague found that stumptailed macaques, whose communities are highly cohesive (probably because they have an evolutionary history marked by the need to defend themselves against external threats), are much more conciliatory than are the foul-tempered, scat-throwing rhesus macaques, whose communities are not particularly cohesive.[30]

A third bit of evidence for the valuable relationship hypothesis is this: the hypothesis implies that reconciliation will be more frequent among relatives and close allies than among nonrelatives and unrelated individuals who are not otherwise very important to each other. Research has supported this prediction very well. Primatologists found that the conciliatory tendency of a particular group of captive chimpanzees was about 60 percent among friends, but only about 20 percent among nonfriends. (How do you measure "friendship" among chimpanzees? You figure out who spends the most time socializing with whom. Friends spend a lot of time sitting in each other's presence and grooming each other. Nonfriends don't.)[31] In stumptailed macaques, the conciliatory tendency is less than 25 percent among nonfriends, but around 50 percent among friends.[32]

In perhaps the most striking demonstration of how the value of a relationship affects whether a conflict will be reconciled, a couple of scientists calculated the conciliatory tendencies of seven pairs of female long-tailed macaques before and after an experimental manipulation of relationship value. In the first phase of the experiment, the researchers simply examined how often these seven pairs of individuals reconciled. Averaging across the seven pairs, about 25 percent of their conflicts got reconciled. In phase two, the seven pairs of individuals were trained to cooperate with each other in order to get food. If one partner wanted to eat, she had to wait until the other one wanted to eat. Then they could work together to gain access to the food. No cooperation, no food. In other words, the researchers used experimental methods to turn the macaques' relationships into *valuable* relationships. After they had been trained to work together in order to obtain food, the average rate of reconciliation doubled to about 50 percent. When group-living animals are given the choice between (a) reconciling with a valuable relationship partner who has harmed them, or (b) holding on to their grudges but going hungry, they generally choose the reconciled relationship and the full belly.[33]

So we're starting to cook up a good just-so story for why group-living animals forgive and reconcile: by doing so, they can preserve valuable relationships with blood relatives and cooperation partners. This is exactly the kind of just-so story we're looking for: one with reams of scientific evidence to back it up.

ANXIOUS TO FORGIVE

After Terry the stumptailed macaque has had a fight with his buddy Joe, Terry probably doesn't start thinking, "My friendship with Joe is really important to me. And the pain I've had to suffer because of what he did to me doesn't outweigh the benefits that I'm likely to enjoy in the future by mending our friendship. Maybe I should try to patch things up with him. Maybe I ought to go grab his rear end—just to let him know that I want to be friends again." It can't be a rational thought process that motivates Terry to mend things, because stumptailed macaques don't have the capacity for rational thought. Instead, anxious tension seems to be the motivating force. Anxiety is an unpleasant feeling for humans and nonhumans alike, and we're motivated to find ways to rid ourselves of it. For the nonhuman primates, reconciliation does the job quite nicely.

We can't ask nonhuman animals to tell us whether they feel anxious, but we can infer it from their behavior. For many species, so-called self-directed behaviors such as scratching, yawning, and shaking seem to be good indicators that an individual is anxious.[34] A primatologist discovered that primates who have had recent conflicts scratch themselves furiously. He also discovered that the stronger the relationship between the two individuals prior to the conflict, the more furious the self-scratching afterwards. Finally, when the aggressive episodes are reconciled, self-scratching subsides, which suggests that reconciliation reduces anxiety.[35] Another good indicator of anxiety is increased heart rate. When rhesus macaques have a conflict, their heart rates go up; after the combatants reconcile, their heart rates go back to normal.[36]

Nonhuman primates' choices to reconcile, then, seem to be driven by their feelings. When a valuable relationship is disrupted by aggression or conflict, they get anxious, they try to patch things up, and presto! They become less anxious as a result. But we humans aren't such slaves to our emotions. Unlike

apes and monkeys, we're capable of making conscious, rational decisions about whether to forgive someone who has harmed us. If somebody has injured you, it's easy to sit down and draw up a list of the costs and benefits of forgiving and a comparable list of the costs and benefits of holding a grudge. You can also reflect on abstract moral principles—principles such as justice, retribution, and care—to help you figure out what to do. After all of this soul-searching, you can then choose whether to forgive or reconcile in a thoughtful, rational way. This could be a much better model for how the human forgiveness process works, except for how incorrect it is.

Now granted, maybe there really are some people out there who could reason their way into forgiveness in such a fashion, but just because we *could* base our decisions to forgive on rationality and moral principles doesn't mean that we actually *do*. Moral choices are deeply influenced by emotion and intuition, perhaps even more strongly than they're affected by reason.[37] This is probably true of most instances of forgiveness, too: you're more likely to forgive a brother, sister, parent, or good friend because "it just felt right" or because "I missed spending time with her" or because "I felt sorry for him" than because you concluded that it was the most rational or morally defensible thing to do.

Rationality and moral reasoning do have a role to play, but it's not the role you might think. If I were to ask you after the fact why you forgave somebody or chose not to, you'd probably have a good rationale at the ready, but I'm betting that the rationale didn't cause your choice. Instead, I'd wager that your choice caused your rationale—you probably used your powers of higher-level reasoning to shore up your justification for doing whatever it was that you felt like doing in the first place. As is the case with the nonhuman primates, it's our emotions that are central, and anxiety is one of the biggies, for kids and grown-ups alike.

Everybody knows that little kids (and many adults, too) suck their thumbs and bite their nails when they're anxious. When preschoolers have had a conflict with a peer, the thumb-sucking and nail-biting increases to a fever pitch.[38] However, once the aggressor and victim have reconciled, the thumb-sucking and nail-biting cease. If the children don't reconcile, the thumb-sucking and nail-biting continue. In fact, children who have just had a conflict also

experience a flood of stress hormones including cortisol (which is closely linked with fear and anxiety) and DHEA-S (which, some researchers think, might be the body's efforts to keep cortisol and its effects under control). If the conflict gets reconciled, the circulating levels of these hormones go back down to their pre-conflict levels. If the conflict doesn't end with reconciliation, the cascade of stress hormones continues.[39]

A laboratory experiment showed that similar things happen when adults recall occasions from their past when valuable relationship partners (mostly friends, romantic partners, parents, and siblings) did something to hurt them. When the researchers asked the participants to think unforgiving thoughts about their transgressors (for example, to think about their grudges or to imagine what it would be like to take revenge), the participants got anxious and tense. They had more muscle tension in their faces. In addition, their heart rates, blood pressure, and sweating all increased. These tension-related symptoms were much reduced after participants were instructed to think about their transgressors in a forgiving light.[40] In another study, researchers found that when they asked people to describe occasions when a friend or a parent had harmed them, those who reported that they had already forgiven the transgression experienced smaller increases in blood pressure than did people who hadn't forgiven. Lack of forgiveness for close, valuable relationship partners who have harmed us in the past is associated with more anxiety, tension, and physiological arousal.

Pencil-and-paper measures of anxiety and stress tell a similar story. When people report that they've forgiven a particular person who harmed them at some point in the past, they experience lower levels of self-reported stress and anxiety. In addition, the extent to which they've forgiven at one point in time predicts how much anxiety and stress they're going to be experiencing several months later.[41] These results from studies of children and adults, then, are very consistent with the sorts of conclusions that the primatologists have been drawing about the emotional factors that motivate reconciliation in nonhuman primates. Aggression and conflict lead to stress and anxiety, which motivate social animals to forgive or reconcile, which in turn alleviates their stress and anxiety.

Know forgiveness, know peace. No forgiveness, no peace.

VALUABLE RELATIONS, AGAIN

The links of forgiveness and reconciliation to anxiety are strongest when the person who hurt you is a close, valuable relationship partner. The same is true of our primate cousins. People who fail to forgive a close, important relationship partner will continue to feel anxious tension when they think about that person; when they forgive, the anxious tension disappears. In non-close, unimportant relationships, this isn't the case: forgiving someone who's not very close or important has no effect on people's levels of anxious tension. Some social psychologists demonstrated this phenomenon in several clever experiments. In one experiment, they gave people a psychological test that supposedly revealed whether they had actually forgiven "deep down" for something that a specific person had done to them in the past. Half the participants were told that the test revealed that they really had forgiven. The other half were told that the test revealed that they were still holding a grudge.

The researchers wanted to know which participants would feel upset by "learning" that they really hadn't forgiven their offenders. It turned out that if the person who harmed them was a stranger or an acquaintance, finding out that "you really haven't forgiven them after all" didn't create much anxiety. However, if the person who hurt them was a close, committed relationship partner (a good friend or a loved one, for example), then "finding out" that they really hadn't forgiven created psychological tension and negative emotion.

Conclusion: people get anxious when they haven't forgiven a valuable relationship partner precisely because the relationship is a valuable one.[42] Of course, this is exactly what the prisoner's dilemma, thirty years of primate research, and adaptationist thinking about forgiveness would lead us to expect. Improving our inclusive fitness and maintaining a stable set of cooperation partners are the *ultimate* causes of our desire to forgive and reconcile. These are the reasons why we possess tendencies to forgive and reconcile in the first place. The motivation to feel less tense and anxious is one of the *proximal* mechanisms that natural selection put in place to make sure that we actually follow through on those evolutionary mandates.

AN INTELLECTUAL BOMBSHELL MADE OUT OF TINKER-TOYS AND FLESH AND BLOOD

David Sloan Wilson has called the prisoner's dilemma a "tinker-toy model" of natural selection because of the simplistic way in which it models animals' social instincts.[43] "Always resume cooperation if your partner does the same." "If your partner is a good friend, forgive him unconditionally 80 percent of the time." "If your partner is a stranger and she betrays you, let her have it." "Trash the reputations of people who attack you if you're not going to get a chance to retaliate against them directly." This sort of simplicity just seems, well, too simple.

But maybe the theoretical biologists get the last laugh here. The research on reconciliation and forgiveness in real, live, flesh-and-blood creatures shows that group-living animals seem to live by social instincts that aren't much more complicated than what the tinker-toy models suggest. Is a social rule that says, "If that guy who just stole your food is a good friend, go up to him and see if he'll let you groom him" (which is what reconciliation, interpreted through the valuable relationship hypothesis, actually looks like) really any more sophisticated than a rule that says, "Forgive your good friends unconditionally 80 percent of the time?" Maybe real life isn't always more complex than the tinker-toy version.

The idea that conflict and aggression attract animals to each other might not seem like an intellectual bombshell, but it is. Group-living mammals don't simply scatter to the four winds or beat each others' brains out after conflicts, as scientists assumed for many years. Instead, they often come together to actively undo the negative effects of conflict and aggression on their relationships. Reconciling and forgiving aren't passive enterprises. Reconciling animals are as sincere and hard-working in their efforts to make peace with each other as they are, at other times, in their efforts to make trouble for each other. Humans are also group-living animals, and by all indications we're just as prone to reconciliation and forgiveness as are the nonhuman species that have received so much attention in recent years. This gives us

cause for optimism that humanity really does possess a "forgiveness instinct."

OPENING THE TOOLKIT

Natural selection seems to have outfitted us with a forgiveness instinct because it helped our ancestors preserve relationships that had biological utility. But they had to have utility, or at least the promise of utility. Natural selection most surely did *not* create a forgiveness instinct because it was useful for our ancestors to try to preserve each and every relationship—just the valuable ones. When the potential benefits of forgiveness are low and the potential costs are high, such as when a victim is figuring out whether to forgive a stranger or a sworn enemy who still seems dangerous, or of little likely value in the future, or undeserving of care and concern, we should anticipate that people will favor the alternatives to forgiveness—revenge being one of them.

Revenge and forgiveness, then, are *conditional* adaptations—they're context-sensitive. Whether we're motivated to seek revenge or to forgive depends on *who does the harming*, as well as on the advantages and disadvantages associated with both of these options. We don't weigh these considerations consciously, of course, but our brains perform the necessary computations behind the scenes. Then those brains motivate us in the direction that they think we need to go. But what, exactly, is going on behind the scenes? A journey into the human brain might be in order.

Now, you might have your doubts about whether scrutinizing a three-pound jumble of neurons can teach us anything useful about why the fathers and mothers of murder victims sometimes forgive their children's killers, or why feuds sometimes end peaceably, or why nations sometimes heal after civil wars, but this is definitely a journey worth taking. Natural selection is the ultimate cause of revenge, but people who are contemplating a single act of revenge are not thinking about their fitness. They're driven by feelings and thoughts that are generated within the brain. By studying the brain systems that come online when people are contemplating revenge, making plans for how to enact revenge, or basking in the warm glow of consummated

revenge, we can better understand what a vengeful person is really trying to accomplish, and what humans might need to control those vengeful impulses. Likewise, if natural selection created human beings with a "forgiveness instinct," it did so by building a set of computational tools that crunch the numbers to figure out whom, what, where, and when we should forgive. These tools are worth trying to understand. To find them, you have to look between your ears.

CHAPTER SEVEN

THE FORGIVING BRAIN

The human brain is the most powerful information processing device in the known universe. It consists of one hundred billion neurons that are joined together by at least one hundred trillion interconnections. Thanks to recent technological breakthroughs, our scientific understanding of how the brain works is light years ahead of where it was even three decades ago. Techniques that record images of the brain's activity as people think, feel, talk, behave, and experience life have enabled scientists to examine the neurological basis of some of the most human-seeming aspects of our existence. In the past ten years, not even the most intimate of our traits—not love, language, sex, or even spirituality—has escaped neuroscientists' probing and prodding.[1] Neuroscientists have even got some important things to teach us about the neural circuitry that motivates revenge and forgiveness.

THE SEEKING SYSTEM AND THE RAGE CIRCUIT

Your brain has a system for telling you whether something out there in the world is good for you—a system that the neuroscientist Jaak Panksepp has called the "seeking system."[2] It doesn't matter what

that something is: if your experience with an object, a substance, or a person in your environment has produced positive consequences for you in the past, the seeking system will create enthusiasm and a feeling of anticipation when a new opportunity to interact with that object or substance or person arises. The seeking system leads you to expect that the upcoming interaction is going to be worth your while.[3]

People who are in the midst of a satisfying and cooperative interaction with another person experience high activation in this so-called seeking system. Neuroscientists know this because a brain structure called the caudate nucleus, which receives a lot of input from the seeking system, is highly active during cooperation. The better a social interaction is progressing, the more your brain (courtesy of the caudate nucleus) seems to be saying, "This is going well for you. Keep it up."[4]

When someone harms you, though, the seeking system shuts down in an instant. Anticipating upcoming rewards is not the most important thing on your plate. You need to think about saving your hide, and, later, about ensuring that you won't be harmed by that person again. That good feeling of eager anticipation is gone, but anger, fear, psychic pain, contempt, and even disgust are there to take its place. Say you're feeling hurt because you've been excluded from a group of people whom you considered friends. This distress is generated by the same areas of the brain that create distress when we're experiencing physical pain.[5] Say, instead, that you're feeling the angry contempt of someone who has been treated with less respect than deserved. This feeling leads you to protest the unfair treatment you've received, and it seems to be driven by a part of the brain that helps create negative emotions such as disgust.[6] The most common emotional consequence of an interpersonal harm, though, is anger.[7]

Just how do the negative emotions we initially experience after being harmed—hurt, anger, and so forth—morph into the searing, focused desire for revenge that creates so many problems for our species? You might expect the involvement of the so-called rage circuit, which is found in many mammalian brains. The German physiologist and Nobel Laureate Walter Hess helped to identify the rage circuit by applying electrical stimulation to the brains of live cats. Apply electricity to certain regions

of a brain structure called the hypothalamus (which is also responsible for regulating body temperature, sex drive, hunger, and thirst), and a previously docile animal is transformed into a seething, snarling, spitting rage machine with claws bared and hair erect. An animal who is receiving this sort of electrical stimulation will attack any living thing it can get its paws on. Apply electrical stimulation to the same part of human brains and people report feeling intense fury. Evidently, animals don't enjoy this electrically stimulated "sham rage" very much: when given the ability to turn off the electrical stimulation—say, by pressing a bar—they readily do so.[8]

The rage circuit leads to very quick and very focused aggressive responses to threats, so it seems reasonable to assume that it's also important for creating vengeful feelings. But as it turns out, revenge isn't primarily a product of rage: the neuroscientists tell us that it's actually a product of *desire.*

FROM VICTIM TO PREDATOR

Recall that if you use a probe to apply electrical stimulation to an animal's rage circuit, the animal will try to turn the stimulation off. If you apply electrical stimulation to another region of the hypothalamus, however, the animal seems to like it. In fact, if cats learn that they can turn on the electrical stimulation to this second hypothalamic region by pressing a bar, they go into frenzies of bar pressing, as if they believe that intense and abiding satisfaction is just one more bar-press away. They don't seem content: crazed is a more suitable description of their demeanor. When you activate this region of the hypothalamus, it turns out that you're actually stimulating nerve fibers from the seeking system that just happen to run through the hypothalamus.[9]

Now, stimulate the part of the cat's hypothalamus that produces all of that bar-pressing and then throw a mouse (even a dead one) into the cat's enclosure. What happens? Instead of blindly lashing out at the mouse in fury (which is what happens when the rage circuit is stimulated), the cat begins to quietly stalk the mouse. Panksepp calls it a "quiet-biting" attack. The rage circuit produces a cat that lashes out at the mouse as if trying to escape a predator, but the seeking system produces a cat that stalks

the mouse as if craving a good meal or, at least, a good hunt. It's the seeking system that's behind this stalking behavior.

Because revenge, by definition, isn't fundamentally about stopping an attack in progress or escaping from a predator that poses an immediate threat, it's a pretty safe bet that the rage circuit isn't so important for revenge after all. It turns out that the seeking system is a much better place to begin looking for the neurological foundations of the desire for revenge.

CRAVING REVENGE

People talk about "craving" revenge. This isn't just a linguistic oddity. It's a signpost to a deeper understanding of revenge and its neural foundations. The "craving" quality of revenge was brought to light by a pioneering neuroscientist named William Shakespeare, once again speaking to us through Shylock in *The Merchant of Venice.* When Salarino encourages Shylock to drop his insistence on taking a pound of Antonio's flesh in revenge for defaulting on a big loan, Shylock reveals that his desire to be made whole is driven not just by anger, but also by hunger. For what purpose could Shylock possibly use a pound of Antonio's flesh? "To bait fish withal: if it will feed nothing else, it will feed my revenge."[10]

When you've been injured by someone, the initial response is that familiar suite of negative emotions—anger, hurt, and the rest—but after those initial negative emotions give way, the seeking system calls for a fundamental change in course. The seeking system motivates people to turn from a desire to escape pain or threat toward a search for pleasure. Recent studies show that Shylock's comparison of revenge to a hunger was more physiologically accurate than Shakespeare could have possibly imagined.

THE "TRUST GAME" STUDY

In 2004, a team of Swiss scientists used positron emission tomography (a technique that involves determining which brain areas are active during a task by measuring how much blood they consume during the task) while a group of men played a "trust game" with what they thought were a series of other sentient

human beings (the participants were, in fact, playing against preprogrammed computer strategies). Economists invented the trust game to learn more about the conditions under which people are willing to trust strangers in social interactions.

The game is modeled loosely upon the relationships of an investor, someone entrusted with the investment, and the market in which the investor's money is invested. Both players start out with equal amounts of money (say, $10). At the beginning of the game, the research subject, playing the role of the investor, can choose to transfer some of his money to an anonymous player who is playing the role of the trustee. For each dollar the investor gives to the trustee, the researchers (playing the role of the market) quadruple it (as if the trustee has done a good job of managing the investor's funds). So, for example, if the investor gave all of his or her money to the trustee, the trustee would then have $50 (the trustee's own initial $10 plus $40 based on the investor's $10 investment, which grew to $40), and the investor would have zero.

Next, the trustee has the opportunity to return some of the $50 to the investor. You probably think, as most people do, that fairness would require the trustee to return $25 to the investor so that they both end up with $15 more than they started with; anything less is usually considered stinginess. But on four out of seven trials—each presumably with a different trustee—the trustees in this particular experiment didn't return any money at all. In those instances, investors reported feeling a strong desire to punish the trustees.

After each of the four betrayals, the investor was given one minute to decide whether to retaliate by taking up to $20 away from the trustee's earnings. It was during this minute that the scientists used positron emission tomography to determine what was going on in the short-changed investors' brains. By varying some of the details during the four rounds in which the computer strategy was stingy with the investor, the researchers were able to examine two different kinds of retaliation—an ability to take away up to $20 from the trustee's earnings "for free" (that is, at no cost to the investors themselves), and an ability to take money away from the trustee at a cost to the retaliator of $1 for each $2 of punishment.

In both the "free punishment" and the "costly punishment" conditions, the caudate nucleus was highly active during the moment of decision. Remember our friend the caudate? It's deeply involved in the seeking system, and it lights up when people are anticipating that they're about to receive a monetary award or a pleasant taste.[11] (You'll recall that the caudate is also highly active when people are interacting in a positive way with a cooperative stranger.) The caudate nucleus was even active in the costly punishment conditions (when the investor had to pay $1 for each $2 worth of punishment inflicted on the trustee). Moreover, the amount of activity in the caudate nucleus during the free punishment rounds was strongly related to the extent to which the investors chose to punish stingy trustees during the costly punishment rounds. This fact suggests that players who anticipated lots of satisfaction in the free punishment condition were also more willing to punish at a personal cost to them—presumably because they anticipated more pleasure to result.[12]

THE "BRAMITOL" STUDY

Some of the most striking evidence that people are actually seeking pleasure when they're seeking revenge comes from experiments by a team of social psychologists. They wanted to know whether people would retaliate against someone who had insulted them if they were led to believe that revenge wouldn't make them feel any better. The researchers convinced the research participants that they were about to take part in a study of how people form impressions of a stranger. Participants were first instructed to read an essay that either supported or refuted the idea that aggression makes people feel better. Next, they were asked to take a harmless pill that would supposedly speed up their reaction times (which, they were told, would be helpful for a later task). Next, one-half of participants were told that the pill (which went by the intriguing name Bramitol, although it was, in fact, just a vitamin B6 tablet) would also have the side effect of "freezing" their moods for about an hour. No matter how hard they tried, they wouldn't be able to change their moods after they took the Bramitol. The other half of the participants were given the same pill to speed up their reaction times,

but they were told that Bramitol wouldn't have any mood-related side effects.

Next, participants were subjected to that workhorse of laboratory research on revenge: they wrote an essay that the stranger (of whom they were soon going to be asked to record their first impressions) was going to evaluate. After writing the essay, the stranger supposedly evaluated it, and then the research participants received very insulting evaluations back from the stranger (poor organization, lack of originality, weak writing style, lack of persuasiveness and clarity, and so on). The stranger also attached a handwritten note that said, "This is one of the worst essays I have read!"

Next comes the part where the Bramitol becomes crucial. The participants and the strangers with whom they were paired then competed against each other in a reaction-time test to see who was faster at pressing a button after receiving a signal. If the participant won the race, the participant got the chance to administer a loud blast of noise to the other player. How loud? That decision was left up to the participant. If the participant set the sound blast device to the highest possible setting, it was ostensibly 105 decibels (about as loud as a jackhammer operating a few meters away). On the lowest intensity, it was supposedly 60 decibels (roughly equivalent to the sound of normal conversation). Participants could also control the duration of the noise by holding their buttons down longer. These sound blasts therefore served as a nice, unobtrusive measure of participants' willingness to deliver a painful stimulus to the strangers who had previously insulted them.[13]

Results showed that the participants who believed that aggression would help them feel better (because they had read the essay that argued that this was the case) gave louder and longer sound blasts to their provokers than did those participants who read the "aggression doesn't help people feel better" essay—but only if they hadn't taken the "mood-freezing" pill. In other words, people seemed to be interested in retaliating, via the sound blast device, to the extent that they believed it would cheer them up. If they thought it wouldn't have that effect (either because they were led to believe it was generally ineffective at doing so or because the Bramitol had ostensibly frozen their moods), they

didn't bother trying to retaliate. Without the prospect of pleasure, revenge just didn't seem worth the trouble.

PLANNING REVENGE

The prefrontal cortex sits right behind your forehead. Evolutionarily speaking, it's a rather young part of the brain, and it's important for a variety of advanced psychological skills such as reasoning, problem solving, and telling right from wrong. The most important thing to know about the prefrontal cortex for our purposes is that it helps people plan the steps for accomplishing their goals. Nature, it seems, has neatly divided the brain's goal-planning responsibilities between the left prefrontal region and the right prefrontal region. If you're pursuing a goal that involves moving toward a desired object ("How do I get something I want?"), it's your left prefrontal area that's most active in the planning process ("First, do Step 1. After that, go to Step 2. If you fail on Step 2, try Step 2b as an alternative."). In contrast, when you're pursuing a goal that involves staying away from something bad ("How do I avoid something I don't want?"), your right prefrontal cortex gets highly involved and your left prefrontal cortex goes on standby.

So guess which side of the prefrontal region is most active when people are plotting revenge? That's right: revenge is a left prefrontal kind of thing—a movement toward an object of desire.

In 2001, a group of social psychologists who study the brain brought undergraduate students into the laboratory and placed them in a rather typical-seeming social psychology experiment. They were asked to write an essay as a way of introducing themselves to another person with whom they were about to interact. As in the Bramitol experiment, after writing the essay, participants read an insulting evaluation of their essays that was supposedly written by the upcoming interaction partner. Afterward, in an ostensibly unrelated task, participants were instructed to choose one of six substances (either sugar, apple juice, lemon juice, salt, vinegar, or hot sauce) to mix with eleven ounces of water. The resulting drink was going to be given to the insulting participant as part of a "taste perception study." Like Bushman

and colleagues' sound-blast device, the opportunity to prepare a nasty drink for the insulter served as an indirect measure of aggression: if you felt like seeking revenge, you could mix up a really awful concoction for your insulter to drink during the upcoming taste perception task.

You won't be surprised to learn that the insults made people angry. They also led them to mix nastier brews for their insulting partners to drink. But what was particularly fascinating was what was going on in these avengers' brains. When they were plotting retaliation, they experienced increased activity in the left prefrontal cortex and reduced activity in the right prefrontal cortex. In fact, people who had the highest differences between left and right prefrontal activation were the ones who reported feeling angriest toward their transgressors. They were also the ones who prepared the most disgusting drinks for their insulting evaluators. So when people are planning revenge, the left prefrontal cortex seems to be egging them on.[14] Conclusion: we plan revenge using the same neural hardware we use to strive for any other outcome we really desire.

WHEN PLANS FOR REVENGE GET FRUSTRATED

What happens when those alluring revenge goals get thwarted? Apparently, it makes people feel pretty frustrated—so frustrated, in fact, that they'll try to anaesthetize their blocked goals with a stiff drink or two. Researchers at the University of Washington and the University of Wisconsin had social drinkers participate in what appeared to be a simple wine-tasting task.[15] The researchers randomly assigned these social drinkers to one of three experimental conditions that I'll describe in a moment. Before the wine tasting, all of the participants were asked to do a couple of other tasks.

First, they completed a set of difficult anagrams. In the room with the participants from two of the three groups was another "participant" (actually a stooge who was in cahoots with the researchers) who finished the anagrams in record time and then proceeded to ridicule the actual participant's intelligence, fashion sense, interpersonal manner, and overall physical appearance.

(One group was not insulted at all, and thus formed a control group). Later, participants from all three groups took part in a "learning task" in which they were supposed to use electric shock to "help" the stooge recall words from a list he had supposedly just memorized. When the stooge made a mistake, participants were supposed to provide a painful electric shock to punish the wrong answer (no shocks were actually administered, although participants apparently believed they were). In other words, participants thought they had been entrusted with equipment that could be used as a low-voltage taser weapon. However, for participants from one of the two groups of people whom the stooge had just insulted, the shock machine mysteriously malfunctioned just before the recall task began. As a result, those participants were denied the opportunity to administer painful electric shocks to their insulters.

Later, all of the participants participated in the wine-tasting task. Subjects were encouraged to drink as little or as much of each wine as they wanted to help them decide how much they liked it. What the wine-tasting task really did was provide researchers with an unobtrusive measure of participants' appetite for alcohol.

Participants who were insulted and then *denied* the opportunity to shock their insulters (because of the equipment failure) drank more alcohol than did the participants who were insulted but then given the opportunity to retaliate. Why? Perhaps because some of alcohol's most potent effects are in the prefrontal cortex—precisely where we make our plans for achieving our goals.[16] Given what we now know about the brain systems that govern the desire for revenge, it's likely that those frustrated avengers drank more alcohol because they were trying to put the left prefrontal cortex to sleep so that they could stop obsessing about their thwarted ambitions for revenge.

"EVERYBODY HERE IS HAPPY WITH THIS": THE REWARDS OF REVENGE

Pursuing a revenge goal is exciting, and having a revenge goal blocked is frustrating, but when we actually accomplish a revenge goal, it's positively exhilarating. This neurological reality was made

plain on March 31, 2004, when masked gunmen in the city of Fallujah, in Iraq's Anbar Province, killed four American security contractors and desecrated their remains in the most gruesome ways anyone could envision.

The four private contractors had been escorting a convoy of three empty trucks that were going out to pick up some kitchen equipment. After the gunmen stopped the SUVs with explosive devices, they opened fire on the SUVs. They then pulled several of the wounded contractors out of their vehicles and into the street.

A crowd of three hundred men and boys rushed to the scene and joined the mob. Somebody went out and found a can of gasoline. The mob doused the contractors and their vehicles and then burned the men alive. After killing the Americans, they pounded their corpses with pipes, shovels and shoes (the latter a characteristically Arab mode of humiliation). Loose body parts were kicked and thrown about like so much street trash. Cars were used to drag two of the bodies through the streets of Fallujah. When the drivers reached a bridge on the Euphrates, the two bodies were hoisted up onto the bridge's metal frame, where they were left to hang for the rest of the day. The men from the crowd took turns having their pictures taken with the carbonized cadavers, now scarcely recognizable as the bodies of human beings.

Every photograph of this horrific spectacle that made it into the American media outlets showed that the men and boys of Fallujah were having a really good time. They didn't look angry. They looked happy. Actually, they looked ecstatic. If you photoshopped the burning cars and the charred human remains out of the pictures, you could easily think they were celebrating a wedding, or perhaps a football victory.

And from the point of view of those in the crowd that day, it really was a cause for celebration. They had dealt the Western forces a humiliating blow. They had managed to retaliate against the vastly more powerful coalition forces, which had invaded their sovereign nation and disrupted their productive lives. Recall that Fallujah was a Sunni-dominated, pro-Saddam stronghold. Under Saddam's powerful patronage system, many of Fallujah's residents had enjoyed material security and status.

So in the midst of the carnage, the men and boys of Fallujah smiled, danced, raised their hands above their heads in exultation, and chanted, "Allahu Akbar" (God is great) and "Fallujah is the graveyard of Americans!" One particularly disturbing image shows three Iraqi men beating one of the burnt bodies with their shoes. They're surrounded by a ring of maybe three dozen men who are pumping their fists in the air, clapping, dancing, and smiling. In the foreground of the picture is a boy—perhaps ten or eleven years of age—who wears that unfakeable sign of genuine joy: the Duchenne smile. The award for understatement of the day goes to the taxi driver from Fallujah who summarized the local sentiment: "Everyone here is happy with this. There is no question."[17]

The fact that the residents of Fallujah had such a fine time that March day doesn't distinguish them in the slightest from the men and boys who inhabit the rest of the world. Geronimo, the fierce Apache warrior, described his elation when he finally took revenge on the Mexican forces that had, a year before, massacred his mother, wife, and three children: "Still covered with the blood of my enemies, still holding my conquering weapon, still hot with the joy of battle, victory, and vengeance, I was surrounded by the Apache braves and made war chief of all the Apaches. Then I gave orders for scalping the slain. I could not call back my loved ones, I could not bring back the dead Apaches, but I could rejoice in this revenge."[18] Anthropologist Chris Boehm recounts an observer's description of the tribal Montenegrin people's love of revenge: "When a Montenegrin takes vengeance, then he is happy; then it seems to him that he has been born again, and as a mother's son he takes pride as though he had won a hundred duels."[19] An ambitious young Philadelphia gangster named Eddie Scarfo enjoyed revenge so much that on one occasion, after murdering someone who had insulted him, he told his associates, "If I could bring the motherfucker back to life, I'd kill him again."[20]

Clearly, these are people who find satisfaction in their work. But you don't have to be a Saddam loyalist or a gangster or a fierce Apache warrior to get satisfaction from revenge. As we might expect from the fact that the seeking system sets revenge in motion, two-thirds of people report satisfaction after

they take revenge on someone who has harmed them.[21] Just as a good meal creates pleasure for a hungry person, drugs create pleasure for addicts, and a cold cup of water seems like an illicit treat when you're really thirsty,[22] seeing your perpetrators suffer for their transgressions also activates the brain's reward pathways.

In 2006, neuroscientists scanned the brains of people who had been treated either fairly or unfairly by another player in an economic game. After the game, participants witnessed their partners receiving painful electric shocks to the hand. While participants were watching the unfair players receiving shocks, they had lots of activity in the nucleus accumbens (interestingly, this effect only occurred in men). The nucleus accumbens is a central part of the brain's seeking system.[23] The more revenge the men said they desired after being treated unfairly, the greater their nucleus accumbens activity as they watched their transgressors suffer. It only makes sense that these vengeful participants were experiencing so much activity in the nucleus accumbens because the brain's seeking system was creating a satisfying, "rewarding" psychological state as they observed the suffering of the people who had treated them unfairly.

Mark Twain once wrote, "Revenge is wicked, & unchristian & in every way unbecoming. . . . (But it is powerful sweet, anyway)."[24] A twenty-first-century paraphrase might read, "Revenge pays neurochemical dividends." People who have been harmed by another person are goaded into revenge by a brain system that hands them a promissory note certifying that revenge, when it comes, will make them feel good. Upon receipt of this promissory note, the left frontal cortex goes to work to develop a plan for obtaining revenge. When avengers actually see their transgressors experiencing the pain they've planned for them, they get the pleasurable jolt that the seeking system had promised. A hard truth of human nature is that it's often pleasant to watch our enemies suffer, and it's a pleasure that we'll sometimes go to great lengths to acquire. Natural selection's logic here seems pretty easy to comprehend: by paying us back with pleasure, our brains ensure that we'll go to the trouble of seeking the social advantages that come from returning harm for harm. Injustice, modern neuroscience tells us, can make sadists of us all.

INSIDE THE FORGIVENESS INSTINCT

If the neuropsychological foundation for revenge is the desire for pleasure, then what are the neuropsychological foundations for the forgiveness instinct? It's no good to wave our arms around and insist that humans are naturally inclined to forgive if we can't point to the mental processes that enable them to actually pull it off.

A few theorists have taken stabs at neurological models of forgiveness,[25] but they've had to work without the benefit of very much hard data. However, a smattering of recent neuropsychological evidence, when paired with more standard psychological research, makes it clear that there are three psychological conditions that activate the forgiveness instinct: (1) *careworthiness* (people forgive transgressors whom they view as appropriate targets for kindness and compassion); (2) *expected value* (people forgive transgressors who, they think, might be valuable to them in the future); and (3) *safety* (people forgive transgressors whom they perceive as being unwilling and unable to harm them again).

CAREWORTHINESS

Humans are capable of experiencing deep and sincere concern for other people, but it's hard to care for every single person you come across. Caring is metabolically expensive. It consumes psychological and physical energy. And it can be personally dangerous—caring for others can literally cost you your life. Haldane's crack about his willingness to surrender his life for his eight cousins comes to mind here: the amount of care we experience for others is directly proportional to genetic relatedness. The closer the genetic relation, the more likely you'll help someone with a favor or rescue someone from a burning building. But we don't compute genetic relatedness on the spot to determine whether we should dash into that burning building. A more proximal mechanism is how close we feel toward the person in need. We care for people to whom we feel close,[26] and we feel closest to those with whom we share the most genes.[27] We also care more for the helpless and the innocent than for those who can help themselves and those who caused their own suffering.[28]

Forgiveness seems to be built on some of the same psychological scaffolding that the brain uses to generate care and concern for others.[29] This is a good news/bad news sort of thing. The good news is that we find it fairly easy to forgive our close relationship partners. The bad news is that many of the people who harm us in real life are people to whom we don't feel particularly close. Sometimes they're strangers. Other times they're people from groups that we've come to mistrust or hate.

So just how *do* you come to care for someone with whom you're not particularly close? One way is through empathic emotion. Empathy is not the warm and fuzzy emotion that it's often taken to be. It can actually feel somewhat aversive, especially when it's associated with another person's suffering. If you're feeling a lot of empathy for someone, you're likely to say that you feel "moved," "sympathetic," "compassionate," or "concerned" for that person. When you stumble into feeling empathic for someone in need, whether that person is a genetic relative or not, you'll be inclined to try and alleviate his or her suffering.[30]

One of the best ways to take all of the fun out of revenge, and promote forgiveness instead, is to make people feel empathy for the people who've harmed them. In 1997, my colleagues and I showed that when people experience empathy for a transgressor, it's difficult to maintain a vengeful attitude. Instead, forgiveness often emerges.[31] Empathy seems to promote forgiveness in relationships between co-workers, friends, romantic partners, Northern Irish Catholics and Protestants, and even perpetrators of crimes and their victims.[32] When you feel empathic toward someone, your willingness to retaliate goes way down.[33]

Neuroscience helps us understand why. In a study I described earlier in this chapter, men experienced high activity in the seeking system when they watched an unfair player receive painful shocks to the hand. However, women didn't experience the same uptick in the seeking system. Instead, when women watched an unfair player receive shocks, they experienced activity in a part of the brain that generates the distress we feel when we're in physical pain. In addition, neither men nor women experienced seeking system activation when watching a *fair* player receive painful shocks. In such instances, they also experienced activity in the brain's pain networks, and the higher their scores on a paper-and-pencil measure

of "empathy," the more pain network activation they experienced. Other research shows that when people feel empathy toward someone who has harmed them, they don't experience the increased activation of the left prefrontal cortex that typically accompanies the desire for revenge.[34] You can stand by passively and watch an enemy suffer, and sometimes that feels good. However, if the suffering of your enemy evokes distress in you instead, then revenge is going to feel hollow, pointless, and cruel, and forgiving is going to seem like the thing to do instead.

A brief story illustrates this point. Steven McDonald was a New York City policeman until one day in 1986 when Shavod Jones shot him in Central Park, paralyzing him from the neck down. Strangely, McDonald found himself completely devoid of any desire for revenge: "I was angry at him, but I was also puzzled, because I found that I couldn't hate him. More often than not I felt sorry for him. I wanted him to turn his life to helping and not hurting people. I wanted him to find peace and purpose in his life. That's why I forgave him."[35]

The Central Park encounter left McDonald forever confined to a wheelchair. Life as he knew it was changed irrevocably. Still, he experienced empathy for Jones, and that empathy led to caring, and that caring made forgiveness possible. However, care on its own is rarely enough.

EXPECTED VALUE

Remember the "valuable relationship" hypothesis from the previous chapter? That's the idea that people forgive (and that nonhuman animals reconcile) to the extent that they perceive their relationship with the transgressor to be a valuable one. Expected value is the second psychological foundation for forgiveness.

The forgiveness epidemic that has broken out among the Acholi people of northern Uganda illustrates the importance of expected value. For two decades, rebels calling themselves the "Lord's Resistance Army" have fought to overturn the Ugandan government. To build support for their cause, they've terrorized civilians. Thousands of preteen girls have been stolen from their villages and given as wives to rebel commanders. Thousands of other children have been taken captive, brainwashed, and turned

into the next generation of rebel soldiers, trained to attack and kill their own people. Villagers who have resisted the rebels have had their lips, noses, ears, hands, or breasts cut off to intimidate others into meeting the rebels' demands.

Fatigue has set in among the Acholi, many of whom have been displaced from their homes for years, so they've adopted an unorthodox strategy for peacemaking: welcoming the rebel soldiers back into their midst with offers of forgiveness. Since 2000, popular radio programs have promised the rebels amnesty if they'll simply lay down their arms and return to their communities. As time has gone on, the grassroots calls for unconditional amnesty—even for the LRA's leader Joseph Kony—have only become more insistent. One man who had been living in a camp for displaced civilians summarized how most Acholi feel about the situation: "Let [Kony] come back and live with the community because this is how reconciliation will be achieved." The International Criminal Court in The Hague has resisted requests that it drop its indictments against the LRA's leaders, but this hasn't deterred the Acholi. Indeed, the Ugandan government has officially offered amnesty to the rebels.

When rebels return home (sometimes in groups as large as eight hundred), they participate in a traditional forgiveness ritual in which they stick a bare foot into a raw egg—a symbol of innocence and new life. Next, they step over the long handle of a farming tool to symbolize their intention to return to a productive life in the community. As a final element in the ritual, they receive a figurative cleansing by brushing against the leaves of a pobo tree, "whose slippery bark catches dirty things." After the ritual, the repentant rebels must sit down with community leaders and formulate plans for confessing their sins and compensating the families that they've harmed—often by paying with livestock.

"What I'm after is peace," said one of the rebels' victims, whose nose, ears, and upper lip had been cut off more than a decade earlier. "If the people who did this to me and so many others are sorry for what they did, we can take them back." And it's not hard to understand why, especially when they repent and attempt to compensate their victims: the LRA turned children against their own villages and their own tribes, but those

children continue to have value to their families and their communities, even though they were brainwashed and intimidated into doing horrible things. As a Catholic nun who works among the Acholi put it, "They are all our children . . . there is no other way."[36] As I write, the soldiers continue to return home, the peace negotiations continue to limp along, and a cease-fire agreement continues to hold firm.

The post-conflict anxiety I described in Chapter Six seems to be one of the forces that motivates people to restore valuable relationships. Concerns about losing a valuable relationship create anxiety, and that anxiety motivates us to find ways to patch things up and restore the relationship. But we've also seen in this chapter that the brain has a dedicated system for computing "value": if we expect our upcoming interactions with someone to be positive, the brain causes us to anticipate rewards.[37] This also helps us forgive valuable relationship partners.

The problem is that when somebody harms you, the harm itself drains some of the expected value from the relationship. The Acholi children who were spirited away to join the Lord's Resistance Army did horrific things to their own people, so they're now regarded as potential agents of harm. Therefore, despite the returning soldiers' implicit value to their parents, siblings, and former neighbors, the Acholi have to reevaluate their relationships with them—quite literally, they have to reassess the value they can expect to derive from the returnees in the future. It's not safe to simply assume that their future interactions will be rewarding.

But transgressions don't necessarily drain *all* of the expected value out of a relationship. Even if you harm me, I might continue to assign our relationship a high expected value if it was really valuable to me up until now, and this will dispose me to forgive you. A social psychologist at Carnegie Mellon University and his colleagues proved this point in an elegant way. They experimentally manipulated the extent to which participants focused on their romantic partners' value to them by asking half of the participants to list ways in which their lives were linked to their partners, and by asking the other half of the participants to list ways in which their lives were independent of their partners. Then, in a supposedly unrelated task, participants were asked to

imagine how they would respond to twelve hypothetical acts of betrayal committed by their partners. People who had thought about the ways their lives were linked to their partners were much more forgiving of the twelve hypothetical acts than were the people who had thought about the ways that their lives were independent of their partners. The researchers went on to show that people with high levels of commitment to their relationships are much more forgiving of real-life transgressions as well.[38]

This bodes well for humans' ability to forgive people who had high expected value prior to the transgression. But what if your relationship with a transgressor had low expected value prior to the transgression (for example, if the two of you were in a long-standing conflict or didn't even know each other)? In such instances, endowing the relationship with some expected value after the fact is going to be more of an uphill climb.

An implication: if you want forgiveness from someone you've harmed, you have to overcome the fact that your victim might not be able to imagine that your relationship could have much value to him or her in the future. If you hurt your victim badly enough, he or she might view you as truly worthless. You have to change your victim's intuitions about your expected value. This is why victims around the world tend to respond positively to compensation as an overture to forgiveness.[39] Paying someone back for the harm you caused signals to the victim that your relationship has the potential to become rewarding once again. Compensation tells the victim's brain, "Remember me, friend? Even though I treated you badly, I'm back to my old, valuable self."

This can be a mixed blessing. Many people find it perplexing that battered women who manage to get away from their violent spouses often end up right back in the hellholes they spent months or years trying to escape. The problem is that despite the violence and terror that they and their children are forced to endure, battered women often perceive that their relationships with their spouses continue to have value (for example, when a husband is a woman's only source of financial support). When battered women feel tied down to an abusive partner by such constraints, they're more willing to forgive the abuse, and therefore more willing to return to the abuser.[40] A woman who forgives and later returns to an abusive domestic partner isn't out of her mind—more likely, she's at the end of her rope.

PERCEIVED SAFETY

Which takes us to the third psychological condition for forgiveness: safety. People are more inclined to forgive a transgressor whom they perceive to be unwilling or unable to harm them again in the future. It's a simple matter of trust: should you expect more pain from your transgressor in the future, or can you trust that his or her intentions toward you are basically benign? As we saw earlier, when children have conflicts, they experience increases in the stress hormones cortisol and DHEA-S; after reconciling, these hormones go back to their pre-conflict levels.[41] These hormonal changes reflect the fact that the prospect of having to endure more conflict and harm is stressful, whereas reconciliation leads to reduced uncertainty about the future of the relationship, and therefore reduced stress. With a conflict reconciled, there's less need to worry about the future, and therefore it's easier to forgive the past.

To evaluate a transgressor's safety, we try to understand why the transgressor harmed us in the first place. Did he intentionally injure you? Could she have avoided harming you in the first place? Could he have known that his actions would harm you? People more easily forgive transgressors whose behavior was unintentional, or unavoidable, or committed without awareness of its potential consequences for others. Malicious, intentional transgressions are much more difficult to forgive than are those for which one doesn't blame the transgressor.[42]

To evaluate whether an offender is safe, people are also interested in the offender's remorse and concern for the victim after the offense. Humans are better prepared to forgive a remorseful transgressor—one who seems to genuinely regret the harm she caused—than an unremorseful one. This makes good sense: the transgressor who is appalled by the consequences of her own behavior, or who is personally pained by the pain that her behavior caused another person, is advertising that she possesses psychological barriers—sympathy with the victim's suffering and a sincere desire to uphold society's

moral standards—that will deter her from treating her victim in the same way a second time.[43] Research suggests that nonvoluntary behaviors such as blushing after a transgression may serve a similar function. Blushing shows that you're aware of your moral infraction and that you're eager to distance yourself from it. As a result, blushing after some moral transgressions seems to make people more forgivable.[44]

There's a paradox here: by admitting fault (either verbally or through some involuntary signal, such as blushing), offenders lock themselves into accepting a certain amount of blame, which works against forgiveness in the short run. However, when they admit fault (especially when their admission of guilt is accompanied by remorse), they're also reaffirming the validity of the social rules that they violated and they're acknowledging the harm that their behavior caused. They may also be acknowledging the psychological pain that their transgression inflicted. In the long run, affirming society's laws and acknowledging the victim's pain make the transgressor more forgivable. But the fact that admissions of guilt and expressions of remorse are such two-edged swords—sometimes getting the wrongdoer into more trouble on the way to getting him or her out of trouble later on—explains why people are often afraid to admit wrongdoing and to apologize.

To evaluate whether an offender is safe, people are also interested in whether a transgressor possesses the *desire* to harm them again as well as the *ability* to harm them again. We usually view it as a good sign when transgressors profess that they've changed their ways and that they won't repeat their offenses,[45] but vows like these usually work only when the victim already trusts the transgressor.

There's another way for a transgressor to create the intuition that he or she can't and won't hurt the victim again: making it seem physically impossible for himself or herself to do so. In many cultures, reconciliation rituals involve the surrender of weapons—perhaps because of the powerful symbolism associated with giving up one's power to harm.[46] We may therefore find it especially easy to forgive transgressors who lack both the will to re-offend and the ways to do so. Without the will or the ways, forgiveness doesn't seem like such a sucker's game.

ACTIVATING THE FORGIVENESS INSTINCT

Evolution seems to have outfitted us with a forgiveness instinct because this instinct helped our ancestors preserve relationships that had reproductive, economic, and political utility. When you care for someone who has hurt you, or you experience your relationship with that person as a valuable one, or you feel safe around that person, those brain-generated feelings are cues that prod you to forgive, and by so doing, to reestablish a relationship that may be worth trying to salvage.

The flip side is that when a transgressor doesn't seem safe, or valuable, or careworthy, people will be naturally inclined to favor the alternatives to forgiveness—revenge being chief among them. At the risk of being repetitive, I repeat: revenge and forgiveness, like all adaptations, are *conditional* adaptations. Whether we are motivated to forgive or to seek revenge depends on *who does the harming,* as well as on what happens afterward.

And ay, there's the rub. The terrible things that humanity most desperately needs to forgive—violence, homicide, genocide, war, political persecution, and disenfranchisement based on religion, nationality, or race—are typically not perpetrated by our parents, brothers, sisters, loving spouses, good friends, or neighbors—people whom we most easily experience as careworthy, valuable, and safe. Instead, they're perpetrated by strangers, enemies, and people whom we hate. The people whom we most need to forgive are the people for whom the psychological building blocks of forgiveness are naturally in short supply.

So we have a serious problem. Serious, yes, but not hopeless. Who's to say we can't create social conditions that will conjure up the psychological ingredients for forgiveness even in situations in which those ingredients are, under normal circumstances, in short supply? Maybe we didn't evolve to forgive strangers who have tried to kill us or our children, but domestic dogs didn't evolve to raise orphaned squirrels and tiger cubs, either. Remember Mademoiselle Giselle—the little papillon back in Chapter One who couldn't resist the urge to take care of an orphaned squirrel along with her own little puppies? Remember that dog from the Thailand zoo who was doing such a great job

of raising a couple of little tiger cubs? The simple presence of small, defenseless, furry, four-legged creatures in need of milk was enough to activate mechanisms in these new mothers' brains that motivated them to treat the strangers as if they were their own young. Can we apply the same logic to helping people forgive? Can we evoke something natural from the human brain by creating unnatural social conditions? Perhaps. But if this approach is to stand any chance of success at all, we need to take a good, long look at those tried and true social behaviors that people use to signal careworthiness, value, and safety.

CHAPTER EIGHT

"TO PROMOTE AND TO MAINTAIN FRIENDLY RELATIONS"

Making Forgiveness Happen

Only weeks after declaring war on Nazi Germany, and just weeks before shipping the first American troops off to Europe, Congress passed the Foreign Claims Act on January 2, 1942. Congress recognized that the nation was going to need a legal mechanism for making reparations to the foreign civilians who would inadvertently (but inevitably) be killed, injured, or made to suffer property damage as a result of U.S. military action. The Foreign Claims Act provides that legal mechanism. Currently, the law allows for compensation payments of up to $100,000. Between 2003 and 2006, the U.S. Department of Defense paid $26 million under the Foreign Claims Act to settle more than twenty-one thousand claims coming out of the wars in Afghanistan and Iraq. Most of these claims were related to automobile accidents, physical injuries and property damage incurred during detention procedures, and accidental death or property damage due to weapons fire.