CHAPTER 6: THE EMPATHY PROBLEM

If you just learn a single trick, Scout, you’ll get along a lot better with all kinds of folks. You never really understand a person until you consider things from his point of view...Until you climb inside of his skin and walk around in it. [Atticus Finch in Harper Lee’s To Kill a Mockingbird (1960, p. 30)]

Introduction

In this chapter I identify a fundamental problem that large group size poses for combating opportunism through harm-based moral restraint. I will show that if one’s implicit theory of morality maintains that wrongfulness is derived solely from harm, then in large groups many acts of opportunism will simply not feel wrong. This is because in large social groups it is often the case that no perceptible harm will come to any person because it is spread over too many people to affect any individual’s welfare perceptibly. As such, there is no harmed person to empathize with, sympathize with, or feel guilty about. This is a very important point, for it is because of this “empathy problem” that neither increasing our concern for others nor increasing our predilection for feeling guilty will effectively combat opportunism where doing so matters most.¹

Consider a single mother of three who has to work two jobs to earn $25,000 per year. Even she would not notice the loss of 20 cents if she did not observe the loss directly. What is going here is very simple. Life is filled with many random factors. Therefore we often cannot even perceive of relatively small losses unless they are pointed out to us directly. This is for good reason. To be so fastidious as to be able to notice all of them would be inefficient as it would risk having us be so preoccupied with minor problems that we will miss large problems.
When harm from opportunism is spread over all members of a large group, then something like the above occurs. The empathy problem refers to the phenomenon of negative moral acts not feeling wrong if moral restraint is solely harm-based in nature and the cost of such acts is spread over so many people that no one is perceptibly harmed. This is a problem because to many individuals embezzling $1,000 represents a significant improvement in their welfare, but since not a single shareholder has his wealth reduced by even a cent as a result no feelings of guilt results. Similarly, when an individual inflates an insurance claim by $10,000, no person on the planet is actually harmed at the margin. It is “The Insurance Company” that is harmed, but it is owned by a great many persons, none of whom are perceptibly harmed. The crux of the problem is that if moral restraint is limited to harm-based moral restraint, then when the harm from opportunism is spread over so many people that there is no person with whom to empathize, there is no one to sympathize with and therefore nothing to feel guilty about.

This is such a simple and obvious problem that one wonders why more hasn’t been made of it. One possible reason is that modeling it mathematically in a straightforward way tends to mask its effect. Although the harm per person falls with the size of the group, the number of persons harmed rises proportionately at the same time. At first, this appears to produce an equivalent mathematical effect – 100 people harmed by a $1 each or 1 person harmed by $100 is still $100 of total harm. This is certainly true, of course, but it leaves out important details about the psychological mechanisms that actually give rise to harm-based moral restraint. In reality, no matter how many terms are in a sum, if each and every term is not perceived so they are all zeroes then the total perceived effect on the part of the opportunist will also be zero. So the fact that a great many people are
harmed ends-up not offsetting the fact that the harm done at the margin is small.\textsuperscript{2} This comports with Gneezy’s (2005) observation that a focus on the harm done to an individual may explain why “…people are more accepting of fraudulent behavior directed at large organizations or rich counterparts than at individuals: the monetary cost may be identical, but the damage to the individual is perceived as greater.” (p. 391)

In the movie Office Space a plan is hatched to skim digits from transactions that would normally be rounded. If you were a “victim” of this scheme you would have had to transact dozens of times just to have the harm add up to one cent. This does not mean you are not worse off, of course, but that is beside the point. The point is that the person stealing the pennies would know that your utility level has not fallen, so you won’t feel badly so the opportunist has no bad feeling to empathize with at the margin. Increasing the number of people who are not perceptibly harmed doesn’t change this result even though it does pile up an impressive sum of money for the opportunist.

In the previous chapter I explained why we might be naturally inclined to obey moral prohibitions that provide moral restraint in those instances where prudential restraint that is derived from repeat play effects, social relations, or formal institutions is insufficient. In this chapter I now explain why to be effective in large groups, the actual content of moral beliefs also matters. In short, our implicit moral theory for why it is wrong to disobey moral prohibitions must go beyond the consequences our actions have on the welfare of others because in large groups the practical effect of such consequences often disappear.\textsuperscript{3} This should surprise no one. Our utterly consequentialist harm-based sense of moral restraint has no reason to work well in large group contexts because it evolved in very small social groups.
A Model of the Empathy Problem

In many cases opportunistic actions impose harm that is spread among the members of an entire group. For example, when a team member shirks, the harm may be spread among all team members in the form of lower total output. When a person embezzles money from a corporate firm, the harm that results is born by all of the firm’s shareholders. When a person exaggerates a deduction on his income taxes, the harm that results is born by all of the nation’s taxpayers.

Let’s therefore consider a single act of opportunism whose harm is spread equally over all of the members of a group. As we consider groups of increasing size, the number of individuals over which the harm from a given negative moral action is divided increases. Obviously, for any single act of opportunism undertaken by B, the harm experienced by any other individual with whom the opportunist can empathize grows closer to zero as n grows larger. As some point the harm done is imperceptible so the common query “Who is it going to hurt?” can be honestly answered “no one.”

The golden rule is often touted as the ultimate moral calculus. The golden rule demonstrates the power of empathy. But even the golden rule is susceptible to the empathy problem. If by doing x₁ person B is able to promote his welfare and harms no one in the process, he can rationalize that were he an “A” (someone in the group over which the cost of opportunism is spread) he would not be harmed and therefore not object. This is indeed what happens when someone supports another’s cheating of an insurance company because “nobody gets hurt” even though that supporting individual pays premiums to or owns stock in the company being scammed.
Note also that harm-based moral restraint disappears regardless of how prone an individual is to feeling guilty (that is, no matter how large the g(x) term is). This implies that inculcating a stronger predilection for feeling guilty does not solve the problem because being culpable for an action that produces no harm means there is nothing to feel guilty about if one’s theory of wrongfulness is based solely on harm. So no matter how large g(x) is, if n is sufficiently large and harm is spread evenly over all members of the group, B will know he is responsible for the outcome but will not feel not feel guilty about it.

For negative moral acts, x, undertaken in a group of size n, we can incorporate the discussion above into the model by rewriting net utility as:

\[ V(x, n) = U(z(x)) - C(x, n) = U(z(x)) - g(x)c(x, n). \]

Recall that since we are limiting our attention to golden opportunities the C(x, n) term does not account for the expected cost of retaliation or the expected cost of a ruined reputation. The key point is that U(), which accounts for the benefits of undertaking an opportunistic action, is not a function of n, but c() is. As the number of people over which harm is spread increases, the harm suffered by any individual with whom the opportunist can empathize vanishes, so cost in the form of experiencing feelings of guilt vanishes. Again, how much one cares about the individuals involved is irrelevant because if they cannot even perceive of there being a change in welfare then clearly no harm has come to them. How prone one is to feeling guilty is similarly irrelevant. Note that this result does not require any heroic assumptions about the nature of individual preferences or knowledge of preferences. One need not know the utility function for every individual.
to know that as \( n \to \infty \) moral restraint vanishes if one’s moral restraint is ultimately derived from feelings of sympathy and guilt derived from doing harm.\(^6\)

Let us now consider how you might actually think through the rationalization of opportunism when empathy effects are driven to zero. You really want something, but to get it you have to do \( x_1 \). You don’t want to do \( x_1 \) because \( x_1 \) is a negative moral act and, being a moral person, you try to stick to moral rules of thumb. You know that when you deviate from moral rules of thumb that you normally feel guilty. So you generally don’t do things like \( x_1 \) because of the feelings of guilt you expect to experience.

But wait a minute. Why do you regard \( x_1 \) as wrong in the first place? If you regard it as wrong solely because of the harm it causes because your implicit moral theory of wrongfulness is solely harm-based, then things you deem as wrong are wrong because they harm others. If you know with certainty that no one will be harmed by doing \( x_1 \) in this particular circumstance, then while doing \( x_1 \) is generally wrong and a prohibition against doing \( x_1 \) is therefore generally a good rule of thumb, the spirit of the rule is not violated by doing \( x_1 \) in this instance because clearly no one will be harmed. Given your implicit moral theory of wrongfulness this is clearly a valid exception to the rule.\(^7\) This does not mean \( x_1 \) drops from the list of negative moral acts. You still understand that doing \( x_1 \) is \emph{generally} wrong. It simply means that \emph{in this instance} you believe that you are confronted with a genuine exception to the moral rule of thumb. Doing \( x_1 \) therefore simply does not \emph{feel} wrong. After all, who’s it going to hurt?

You might even still feel a little guilty about doing \( x_1 \) because of your habit of mind, but you will experience far less guilt than if you knew your actions would actually harm someone. As long as the utility loss from experiencing that twinge of guilt is outweighed
by the utility gain from spoils of undertaking \(x_1\), you will still undertake \(x_1\). This example is precisely the problem we face in large groups. As a rule of thumb we regard \(x_1\) (stealing) as a negative moral act. As such we think it is wrong and we know we are not supposed to do it. We don’t give much thought to why we believe it to be wrong. But if we believe that what makes \(x_1\) wrong in the first place is solely the harm it causes others, then in a circumstance in which no person is harmed (taking paper home from our large employer) we have a clear exception to our moral rule of thumb against doing \(x_1\).

This has important implications for opportunistic behavior in general. All opportunistic acts ultimately involve taking advantage of a trust. Taking advantage of a trust is a negative moral act. We don’t give much thought to why we believe this, but in reality this carries moral force with most people because taking advantage of a trust nearly always results in harm. As a result, a general prohibition against taking advantage of a trust is a moral rule of thumb in all societies. But if your implicit theory of morality maintains that wrongfulness is itself ultimately derived solely from a prohibition against doing harm, then if taking advantage of a trust in a particular circumstance is expected to harm no one you will believe that you are confronted with a genuine exception to the moral rule of thumb that prohibits taking advantage of a trust. You therefore won’t feel very guilty about behaving in an opportunistic manner in that instance.

This result is hardly shocking or surprising. It comports with our personal life experiences and is exemplified repeatedly in our literature. In *The Adventures of Huckleberry Finn*, for example, Huck’s acts of complicity with petty cons show that his moral restraint was largely limited to his natural reluctance to harm others. So as long as the cons harmed a great many by very little, his conscience was not terribly burdened.
But when the king and duke were going to take everything left to a man’s daughters and nieces, it was too much for him to bear. Why? In this case he could directly empathize with the harm he was doing to actual people because the harm was no longer divided over many people.\(^8\)

Given his willingness to mock those driven by facile eagerness to obey rules I suspect that Mark Twain would have us believe that a man whose conscience is driven by rules rather than sympathy is a less moral man. I would be inclined to agree, as far as that goes. But we will find shortly that while the *sine qua non* of our basic decency may be ultimately derived from sympathy actuated by empathy, this is merely the foundation for a decent society as a whole. A foundation does not a house make, and we will soon find that because of the empathy problem harm-based moral restraint is simply not enough for a society to enjoy the full fruits of prosperity.

As was pointed out in chapter 2, even if no perceptible harm is done at the margin by a single opportunistic action undertaken by a single opportunist, when everyone behaves as an opportunist great harm is done to everyone. When every employee of a large firm embezzles just a little, the cumulative effect is devastating. The crux of the problem is that at no individual has the power to change the outcome through his own moral restraint. Indeed, any individual opportunistic act the individual takes does not by itself bring about the general result. As we saw in chapter 2, opportunism is a pervasive problem for human societies precisely because it is often quite rational given the nature of the incentives involved.

The empathy problem makes matters worse because when *n* is very large, it is often that case that there isn’t any perceptible harm done to any victim. But if everyone always
behaves opportunistically when they think they can get away with it then the combined effect is devastating. It can be rational yet harmful because, as was demonstrated in chapter 2, that is the nature of the commons dilemma associated with opportunistic behavior. The effect that the empathy problem has on the commons dilemma problem in terms of its canonical matrix payoffs is presented in the Appendix.

Two Illustrations of the Cost of the Empathy Problem

To see how a diminished ability to empathize with harmed individuals can reduce general prosperity, let’s first consider what happens when firms are limited in size because moral restraint is solely harm-based. Suppose that for firms of size \( n > n^* \) workers, empathy effects are so weak that behaving opportunistically is generally optimal for workers so agreements to refrain from opportunism inevitably break down for the reasons discussed in chapter 2. Now suppose that each of two societies is comprised of 100 individuals and they produce one good, whose output is measured by \( Q \). Suppose it is more efficient to produce a great deal of this good in one firm than it is to produce a small amount in many smaller firms. In other words, bigger as better, as argued by Adam Smith and as discussed in chapter 3. To reflect this, let \( Q = L^2 \) in both societies, where \( L \) is the number of workers. Increasing returns is reflected by the exponent on \( L \) being greater than one.

Now suppose that in the first society people are unable to solve the empathy problem for firms larger than 10 individuals so \( n^* = 10 \), whereas in the second society the empathy problem is better solved so groups can be as large as 100. In the first society output is \( 10(10^2) = 1000 \). In the second society output is \( 1(100^2) = 10,000 \). If the first society could solve its empathy problem as well as the second, then its output would rise tenfold.\(^9\)
Here is another example. I used to frequently get advertisements for cable television splitters that enable one to get premium channels for free. Since the signal is being sent anyway, the marginal cost imposed on the cable company is zero, so I would not harm anyone if I were to buy the splitter. Suppose everyone is willing to pay $30 a month for premium cable channels. Suppose the cable company has to pay the premium channel suppliers a flat fee of $100,000 per month for the right to provide these channels and it charges $20 per month to its customers. If all 10,000 people in the subscription area want the premium channels, then the cable company can profitably provide what all the customers want and are willing to pay for. What costs the cable company $100,000 a month to provide it can sell for $200,000 a month to its customers.

Now suppose that anyone can buy a cable splitter for $50 that allows free but illegal access to premium channels. If moral restraint is solely harm-based in nature, then no one would pay the $20 per month to buy the premium channels – they would all buy splitters that pay themselves off in three months. But in this case the cable company will not be able to cover the $100,000 it needs to get the signal in the first place, so no one gets the premium channels. This leaves everyone worse off (the cable company makes no additional profit and each customer fails to get something worth $30 to him per month by paying only $20 per month). To argue that harm is in fact done in equilibrium is to completely miss the point about why social dilemmas in general are so daunting. An individual’s decision to buy a splitter for himself does not bring about the terrible outcome – at the margin the loss in revenue to the cable company is not sufficient to cause it to terminate service.
Other Problems Associated with Harm-Based Moral Restraint

*Differences in Income or Wealth*

Harm-based moral restraint likely weakens the richer is the victim relative to the opportunist and/or those benefited. If utility is diminishing with respect to income and wealth, then the richer is the victim, the less the victim is harmed for any given amount taken. For this reason, it is simply difficult empathizing with those who possess much more wealth or earn much more income than we do. Consider a poor man who robs a billionaire of $10,000. If his moral restraint is derived solely from a reluctance to harm others, then he will not feel particularly guilty because he knows the reduction in the billionaire’s wealth is imperceptible. But even though he is poor himself, he might feel terribly guilty about cheating a homeless person out of $10. Moreover, the more that disparities in income/wealth are believed to be random or the winnings of a zero-sum game, the easier it is to rationalize theft precisely because empathy/guilt effects are reduced by such hypotheses (we wouldn’t likely feel as guilty about stealing what is believed to be either an arbitrary or possibly ill-gotten gain).10

This may present a problem for poor people who want to move up the income distribution because it makes it hard for rich people to trust them. If A believes that B’s moral restraint is solely harm-based, then A will not genuinely trust B unless he has s strong personal connection to B. Since many high paying jobs and careers require being genuinely trusted by others, this can be a daunting impediment to economic advancement by those who are poor. To the extent that poor people are discriminated against in this way, poor people are more likely to become discouraged and demoralized, concluding that “the system” is unfair and thereby more easily rationalizing future opportunism.
**Differences in Time Horizon**

The ability to empathize with others can also be weakened when the benefits of opportunism are realized in a time horizon that differs from the time horizon over which harm is experienced by victims. In the 1980s many people who were otherwise moral copied computer programs purchased by others even though doing so was illegal. The problem was that when people copy software it harms no one at *that* margin. The cost to software producers of a single individual copying their software is essentially zero to them so that individual knows he has not harmed them in any way. No harm was suffered by the software manufacturer so there was no one to empathize with and therefore nothing to feel guilty about. To those who lacked a principled reluctance to copy software illegally it simply didn’t feel wrong.

But doesn’t the software firm lose the profit it would have otherwise earned? The answer is no if the opportunist can convince himself that he would not have otherwise bought the product. This, of course, is an easy thing for an opportunist to convince himself of and he has ample incentive for doing so. Moreover, even if the opportunist knew that he was willing to buy the software at the market price, he could rationalize that the software firm could have improved his welfare at zero marginal cost but refused to. Instead it engages in what he regards as opportunistic hold-up by making him pay for something that costs it nothing to give away by allowing unlimited copying.  

We have all heard this kind of rationalization before and many of us made such rationalizations in our youth. It is a fallacious rationalization of course, because it considers the issue in an artificially narrow way (if everyone copies software illegally, then similar to the earlier example about cable splitters software creation becomes an
unprofitable business so we all are harmed by not having software produced). One does not have to be evil to make such rationalizations – one need only desire the object and possess an implicit theory of morality that fails to make it feel wrong.

*Fixed Versus Variable Costs*

A closely related problem is political opportunism that takes advantage of the fact that most voters do not understand the distinction between fixed and variable costs. If we possess only harm-based moral restraint, then the benefits that accrue to society from the production and sale of high fixed cost, low variable cost goods such as drugs will disappear if such goods are under produced because government only allows drug companies to charge a small amount above variable costs to keep the company from going bankrupt. The problem is that such a policy often results in revenues falling far short of covering prior development costs.

This is essentially an act of social bait and switch (the firm makes long-term investments because it believes it will make large profits over many years only to discover that once the investment is in place the government delivers a political benefit to voters who are eager to rationalize a policy that reduces current drug prices). The problem is that once this lesson has been learned by drug companies, investment in developing new drugs drops dramatically so products we can’t even conceive of yet fail to ever emerge in the first place. Political opportunists are never held responsible because no one misses drugs that never came to be.

The problem is really that the time scales over which the decision is made to buy or illegally copy software or to buy a drug is different than the time scales over which investments that make software or drugs possible are made. In the moment of
opportunistic appropriation, the assets are fixed and therefore moot. If people could, at that moment of self-serving rationalization, fully appreciate the entire time scale associated with production of these kinds of goods, then it would no longer appear to them that the marginal cost imposed on the seller of their actions is zero.

Counterfactual Losses

Our ability to empathize and therefore sympathize with others is also diminished when harm is counterfactual in nature. Suppose $B$ undertakes an act that does not reduce $A$’s welfare from what it is presently but fails to increase it by as much as some other action would have. In many cases we are not obliged to increase the welfare others, and certainly not to increase it as much as possible. But in cases in which, in return for something of value like a job, $B$ has promised to do his best to maximize $A$’s welfare, failure by $B$ to take the most beneficial action because another action produces more benefits for $B$ amounts to reneging on a contract. As such it is an act of third-degree opportunism as defined in chapter 2.

The reason why empathy effects are likely to be low or non-existent with respect to counterfactual harm is that it is hard to think of a person as being harmed when his/her welfare has not been reduced and it is especially hard when his/her welfare has actually been increased. In this case, harm does not fit the normal pattern of cause and effect ($B$ does $x$ to $A$ [cause], therefore $A$’s welfare falls [effect], so it follows that $B$ harmed $A$ by doing $x$). Suppose, for example, that a store manager considers taking advantage of an opportunity to steal $500 from the store’s owner. This was the kind of moral dilemma faced by Ethan Allen Hawley in John Steinbeck’s *The Winter of Our Discontent*. If he is the manager of a very small store, empathy effects are generally strong enough to assure
that sympathy for the owner will be strong enough to keep the manager from stealing directly from the store’s owner and thereby reducing his welfare.

But suppose the manager could choose actions $x_1$ or $x_2$ in response to an event and neither action is known to the owner as they constitute local knowledge possessed by the store manager only. What if $x_1$ and $x_2$ solve the problem so well that in both cases profit rises, but $x_1$ makes it rise even more than $x_2$, while $x_2$ produces more benefit to the manager than $x_1$? In this case the harm done to the owner by the manager’s choice of $x_2$ is counterfactual in nature so the manger might not feel guilty because his choice of $x_2$ did not reduce the owner’s welfare from what it was – indeed it increased it. In other words, putting himself in the store owner’s shoes he could rationalize that if he were the owner he would not be made unhappy by the outcome because his welfare would have actually been increased at the margin.\textsuperscript{12}

But shouldn’t the manager’s ability to empathize with the store owner lead him to imagine himself as a store owner and conjecture that were he the store owner, he would want the manager to choose the most profitable action? Perhaps. But this alone is not enough to produce harm-based moral restraint. It is still the case that there will be little actual harm to sympathize with and therefore feel guilty about. In some cases, of course, the level of sympathy the store manager has for the owner might be so great that the manager will always choose the most profitable action. This is certainly plausible if the store manager and owner are very close, which may explain why in most of the world a store owner will only delegate significant decision making discretion to a relative or close friend.
The weak or non-existent empathy response to counterfactual harm is of particular relevance to the use of relational contracts for which third-degree opportunism is a problem. Recall that the larger is the firm, the more localized is knowledge and therefore the more important are relational contracts. But relational contracts create innumerable opportunities for third-degree opportunism that produce harm that is often only counterfactual in nature because the self-serving action still increases profit at the margin.

In large firms that are crucial for achieving general prosperity, the use of relational contracts is particularly important to deal with the local knowledge problem in a way that supports entrepreneurial decision making throughout the firm. But the local knowledge problem that occasions their use and the flexibility they afford produces many more opportunities for third-degree opportunism than otherwise – many of which produce only counterfactual losses. In such cases if moral restraint is based solely on a reluctance to do harm, then moral restraint will be virtually non-existent.

**Solving the Empathy Problem**

Most humans have understood for a very long time that there are tremendous benefits to being able to organize social, political, economic, and military activity in large groups. In chapter 3 I explained why a fundamental obstacle to large group cooperation was the problem of increasingly localized knowledge. Now that we know that there is another, even more fundamental problem: since our natural, harm-based sense of moral restraint withers in large groups, to build large civilizations humans had to address the empathy problem.

Some moral concepts may have evolved in part, therefore, to solve the empathy problem by supplementing harm-based moral restraint. For example, before societies had
formal governments they were able to organize social, political, economic, and military activity into larger groups by addressing the problem of opportunism informally. Public shaming, for example, was often used to sanction those who were caught behaving opportunistically (Lal 1998). The possibility of having to experience shame and feelings of embarrassment drives up the expected cost of undertaking negative moral actions even in those cases in which no one could possibly be perceptibly harmed, so a social norm of shaming those who are discovered to have been behaving opportunistically can be viewed as a means of overcoming the empathy problem.

In even larger groups, opportunism that is not suppressed by self-restraint in the form of harm-based moral restraint or incentive effects due to shaming was suppressed by institutions that either preclude it through strict routines and procedures or deter it through prudential restraint derived from incentives produced by monitoring and punishment. These were particularly effective means of combating first-degree opportunism because it involves potentially observable acts so harm is usually not counterfactual in nature as is common with third-degree opportunism.

But as successful as these approaches were, they had serious shortcomings. Shaming has little effect in a society that is so large that people are effectively anonymous. This left formal institutions to take up the slack, but the harsh punishments that were common in all civilizations just a century ago are for the most part politically impossible in modern democratic societies. Judges and juries are normally inclined to punish opportunists in proportion to the damage done to people (judges and juries are, after all, hard-wired to automatically conjecture that doing so is just). This means that for expected costs to be sufficiently high to discourage opportunism, the reduction in harshness of
punishment must be made up for by increasing the probability of detection. But in many poor societies this is simply not possible.

An even more fundamental problem is that of third-degree opportunism. Relational contracts are the key to addressing the problem of local knowledge in large groups but relational contracts are highly susceptible to third-degree opportunism, which is by nature largely beyond the reach of institutions. This means that societies that are limited to relying on institutions to combat opportunism will face a tradeoff between firm size and the efficient use of local knowledge and, therefore, a tradeoff between the benefits of scale and the benefits of entrepreneurial rather than bureaucratic direction throughout firms. As I noted in chapter 3, this may explain why most civilizations have been able to produce large organizations or entrepreneurial organizations, but not large and entrepreneurial organizations.

In chapter 4 I argued that moral restraint could provide a way out of this problem. Recall that agents will not, by definition, be discouraged from acting on golden opportunities if their restraint is solely prudential in nature. Moral restraint overcomes this problem because it does not require even the possibility of detection because it is internalized. This renders the inability to detect golden opportunities present moot. But in this chapter I have shown why even moral restraint is not necessarily enough. This is because if moral restraint is based solely on a reluctance to do harm it will wither in large group settings.

Can’t societies circumvent this problem by dividing large groups into many small groups? Rubin (2002, p. 125) has pointed out that in large groups we tend to divide ourselves into smaller groups. Within large firms, for example, many workers are
organized in small teams. Could this be the result of a need to solve the empathy problem through subdivision of large groups into smaller ones to actuate our natural, harm-based sense of moral restraint? Kandel and Lazear (1992) have indeed argued that this may be explained as an effort to harness the power of guilt to minimize shirking. By having some of each team members’ reward made at the team level and also keeping the team small, shirking noticeably harms one’s teammates thereby actuating harm-based moral restraint.

But shirking is only one form of opportunism and it is not even the most problematic form since it can often be addressed with external mechanisms that don’t weaken with group size. As we learned from chapter 3, the most problematic form of opportunism in large organizations is third-degree opportunism since it usually cannot be addressed by external mechanisms and it can potentially make the use of relational contracts impossible. Dividing workers into teams is especially impertinent to the problem of third-degree opportunism for a decision maker in a very large firm. This is because harm is generally counterfactual in nature, so in many cases not choosing the best possible action does not actually reduce any team member’s welfare from its current state.

Moreover, the empathy problem has little to do with the size of the group within which cooperative production activity takes. What really matters is the size of the group over which the costs of opportunism are spread. Dividing workers into teams will do nothing to limit opportunistic consumption of health insurance that is provided by the firm or to limit petty theft from other units of the firm. So while dividing workers into teams may deal with some forms of opportunism, particularly shirking, it does not come close to fully addressing the empathy problem.
The most obvious means of effectuating moral restraint of sufficient strength to address the empathy problem is to drive up regard for others. We generally do not, after all, take advantage of those we care deeply about. In a recent book, *The Empathy Gap*, J.D. Trout (2009) makes essentially this argument, echoing a common view that appears to be gaining in popularity. So couldn’t a sufficiently strong desire to be a nice person overcome this problem? A desire to be nice certainly makes one more willing to undertake positive moral actions and less willing to undertake negative moral actions. As such, it could induce people to try to be more empathetic so they are nice to greater effect. But if one’s unwillingness to take negative moral actions is based solely on a reluctance to harm others, then if no one is perceptively harmed we are right back to the empathy problem.

With no perceptible harm there is nothing to empathize with, no reason to feel sympathy, and therefore no reason to feel guilty. Therefore increasing $B$’s regard for others in general by increasing $\theta_0$, the lower bound for $\theta$ in a given group, does not solve the problem. If no one is perceptibly harmed in the first place then the level of sympathy for others is simply irrelevant so driving up $\theta$ is irrelevant – the harm-based moral restraint story never gets off the ground. So driving up regard for others doesn’t change the fact that our natural sense of moral restraint breaks down precisely where it is needed most – economic activity organized in the context of large groups.

So moral restraint _per se_ is not enough – the kind of moral restraint people have also matters. Harm-based moral restraint is clearly insufficient in many large group settings. Is there some other way to make people reluctant to undertake negative moral acts when they believe there is little or no chance of being caught? One possibility would be to
teach people moral beliefs that compel them to not undertake negative moral acts because they believe undertaking them is wrong as a matter of principle and is therefore wrong irrespective of the harm they may do to others. This would produce what one might call *principled moral restraint*.

**Principled Moral Restraint**

My claim is that if the empathy problem is not addressed, it presents a daunting obstacle to the process of economic development and thereby impedes the maximization of general prosperity. The good news is that it can be plausibly overcome with moral beliefs that deem undertaking negative moral actions as inherently wrong and therefore wrong a matter of principle and therefore wrong even if no harm is done to others.

Principled moral restraint is particularly relevant to combating the problem of opportunism. One generally recognized negative moral action is that of breaking one’s word. If no one ever broke his or her word, including when it is implicitly given in the context of a social contract, then opportunism simply ceases to be a problem for society. But even if breaking one’s word is recognized as being generally wrong, if the implicit theory of wrongfulness is based on harm and in a particular instance no one is perceptibly harmed, in that instance breaking one’s word won’t feel wrong, it will feel like a legitimate exception. The key to solving the problem of opportunism in a general way, then, is to make breaking one’s word *always feel wrong* because it is inherently wrong and therefore is wrong as a matter of principle.

But regardless of the nature of the arguments that might be made for having principled moral restraint, such restraint must ultimately be effectuated through the psychological mechanisms humans actually have – otherwise we risk developing a theory
of moral restraint that is based on unsound assumptions about human behavior. This is why I carefully developed a theory of moral restraint based on hardwired moral intuitions in chapter 5. I will now explain how principled moral restraint can plausibly work through a psychological mechanism that humans already possess, feelings of guilt attached to negative moral acts themselves by moral beliefs. This is not a heroic assumption – just ask anyone who foregoes an opportunity to improve his welfare when he could have done so with no chance of detection and without harming anyone else, and invariably that person will say something like “…honestly, it didn’t even occur to me but even if it did, I still wouldn’t have done it because it still would have been wrong.”

If the feelings of guilt an individual expects to experience from doing something he believes to be wrong as a matter of principle are sufficiently strong, then the resulting guilt cost will make undertaking any negative moral act irrational even in such a large group that harm-based moral restraint is rendered inoperative. This can be added to our model of rational opportunism in a very straightforward way. Recall that if moral restraint is solely harm-based and therefore solely consequentialist in nature, net utility can be expressed as:

\[ V(x, n) = U(z(x)) - C(x, n) = U(z(x)) - g(x)c(x, n). \]

Recall that as \( n \) grows larger, the effect on \( U() \) of undertaking \( x \) is unchanged but the cost of experiencing feelings of guilt vanish. In this case, the commons dilemma associated with opportunism can only be overcome with external mechanisms that will inevitably prove insufficient to provide moral restraint with respect to golden opportunities.

We can now incorporate principled moral restraint to solve the empathy problem directly, thereby alleviating the problem of third-degree opportunism in large groups such
as large firms. Suppose people possess moral beliefs that attach feelings of guilt to the
decision to undertake negative moral actions irrespective of harm done to others because,
given their moral beliefs, they believe that disobeying moral prohibitions against negative
moral actions is wrong as a matter of principle. In this case there is a separate channel of
guilt so we must distinguish between guilt derived from being responsible for having
done harm (g1) and guilt derived from the act itself (g2). Such a non-consequentialist
source of guilt can be added as follows with a term that reflects guilt associated with
negative moral acts irrespective of harm, that is:

\[ V(x, n) = U(z(x)) - g_1(x)c(x, n) - g_2(x). \]

The main point is that while g1 can often be rendered moot by a sufficiently large n, g2 is
unaffected by n because this form of guilt is attached to the act itself. This dramatically
reduces the risk of opportunistic exploitation in the context of large groups. The original
C(x, n) term now becomes

\[ C(x, n) = g_1(x)\theta_1E[\Delta U_1(x)] + g_1(x)\theta_2E[\Delta U_2(x)] + \ldots + g_1(x)\theta_nE[\Delta U_n(x)] + g_2(x). \]

But for \( n > n^* \), we have:

\[ C(x, n > n^*) = g(x)\theta_1[0] + g(x)\theta_2[0] + \ldots + g(x)\theta_n[0] + g_2(x) = g_2(x). \]

Obviously, if \( g_2(x_1) \) is sufficiently large for a given individual, then it will be
irrational for that individual to undertake negative moral action \( x_1 \) regardless of how large
\( n \) is. Therefore the empathy problem can be solved through moral beliefs that tie guilt
directly to \( x_1 \) thereby making \( g_2(x_1) \) very high. By doing so, the commons dilemma
associated with opportunism disappears.\(^{14} \) It follows that from society’s point of view
that a key to enjoying the lowest possible transaction costs is to make \( g_2 \) as high as
possible for as many people as possible over the largest set of negative moral actions
possible. Of particular importance would be to make \( g_2 \) as high as possible for the negative moral act of breaking one’s word as this provides an overarching mechanism to combat all forms of opportunism.

An obvious way to drive up \( g_2 \) is to drive up moral conviction. One does not need much conviction to obey hardwired moral intuitions because for normally functioning adults their gut already tells them that hurting others is wrong. Abstract moral ideas that provide the basis for believing a given type of act is inherently wrong, however, likely require much more conviction because our hardwired moral intuitions give us little to go on. It is one thing to refrain from stealing $10,000 because you are not willing to harm another person. It is quite another to refrain from stealing $10,000 if you know with certainty that no specific person – not even a complete stranger – will be harmed in any way. In this latter case, if one knows one cannot be detected, one must, with great conviction, believe that stealing is inherently wrong if sufficient principled moral restraint is to be effectuated.

Religious beliefs can obviously play a role in driving up conviction and perhaps this may be one reason why religion is such an important part of human history.\(^{15}\) It may also be a reason why in large societies religions are noticeably more abstract and value oriented.\(^{16}\) Many religions deem \( x_1 \) to be wrong because God said it is wrong. Since God is the ultimate moral authority, such a pronouncement makes it inherently wrong and therefore wrong as a matter of principle and therefore wrong even if no harm is done. Religious beliefs could have also improved the function of formal institutions by providing a moral basis for meting out punishments that are excessive given the nature of the transgression, but that are necessary given the difficulty of increasing expected costs.
through increased probability of detection. It is easier to hang someone for a minor crime if you believe it is your moral duty to do so because God commands that you do so.

A person goes beyond the consequentialist, harm-based approach to moral restraint when he adds a non-consequentialist, principled approach to moral restraint because he regards the act as being wrong as a matter of principle and therefore feels guilty about undertaking the act even if no one is harmed in any way. But if the expected guilt cost is nevertheless low and the benefits arising from opportunism is high, then he will still act in an opportunistic manner even though the reason for feeling guilty was that he believed the act was wrong as a matter of principle. In short, principled moral restraint does not necessarily imply sufficient moral restraint to fully discourage opportunism.

This is very realistic. Even criminals who defend their actions by arguing that no one was harmed will often agree that what they did was still wrong as a matter of principle and that they feel some guilt for having done it. But they did it anyway because they want what they want so badly. The utility gained from the pleasure derived from the expected payoff of the opportunistic act simply outweighed the utility lost from any feelings of guilt for knowingly having done wrong.

Conclusion

Our ability to empathize and therefore sympathize with others is not just the basis of our natural sense of moral restraint – it is the very foundation of basic human decency. It is a deep part of who we are. I personally find great comfort in the idea that an important part of our goodness is written in our genes.

In chapter 3 I argued that prudential restraint effectuated by externally produced incentives is not enough to minimize transaction costs and thereby maximize general
prosperity. To most fully deal with the local knowledge problem it follows that moral restraint is also required to overcome the problem of third-degree opportunism because relational contracts open the door to third-degree opportunism.

In this chapter I have argued that even moral restraint *per se* is often not enough. This is because the larger is the group over which the costs of opportunism are spread, the more likely no actual harm is done to anyone with whom we can empathize, sympathize, and therefore feel guilty about harming. If moral restraint is solely harm-based, it winds-up withering in the large groups whose local knowledge problems occasioned the need for relational contracts in the first place. Because of the empathy problem, to discourage opportunism where doing so matters most – in the context of large group economic activity – we need more than harm-based moral restraint. We need principled moral restraint. This again points to the fact that we should not take too much solace from naturalistic arguments made in books by thoughtful authors such as Hauser (2006), Shermer (2008) and Trout (2009). The universal factors they seek to explain moral behavior are very important, but they are simply not enough.

This chapter is titled “The Empathy Problem” and not “The Sympathy Problem” for a very good reason. Contrary to what many modern moral theorists appear to believe, a high trust society cannot be produced by simply increasing our regard for others. How much we care about others is simply irrelevant if no harm is done to them in the first place. Therefore striving to increase moral earnestness by driving up the $\theta_0$ and $g_1()$ terms for all members of society is a misguided application of small group thinking to a large group problem. Indeed, in chapter 8 I will explain why attempts to produce moral
restraint by increasing moral earnestness may actually backfire if moral restraint is limited to that which comes most naturally to us.

The analysis presented in this chapter has troubling implications for those who take comfort in the fact that much of our moral behavior is written in our genes. For example, Francis Fukuyama, in his book *The Great Disruption*, argues that much of our moral nature is biologically hardwired and therefore quite robust. Marc Hauser, in his book *Moral Minds*, makes a similarly compelling argument. I, too, have already stated that I take great comfort in the idea that much of our decency is hardwired. But the more natural our moral sense is, the truer it is that the capabilities that produced it evolved in very small groups. It would only be by dumb luck that these capabilities also happen to work well in large groups.

While harm-based moral restraint is biologically grounded and therefore robust, I have explained why it is not enough if a society wishes to enjoy a condition of general prosperity. Such societies also require principled moral restraint, and it does not follow that the abstract moral beliefs that produce principled moral restraint are similarly robust. Even powerful ideas are merely ideas. Ideas are powerful precisely because they are adaptable and they are adaptable because they are not hardwired. As such they are fragile compared to hardwired moral intuitions. This suggests that solace taken in the fact our morality is to some extent hardwired is unwarranted.

Finally, let me emphasize yet again that the discussion above should in no way be taken to suggest that our ability to empathize and sympathize with others is not important. Empathy and sympathy are not enough to achieve a large and generally prosperous society, but they are absolutely essential to having a decent society. While
general prosperity requires moral sensibilities based on abstract moral ideas, the greatest horrors of human history were made possible by leaders that used abstract ideas to supersede what their citizens’ natural moral instincts told them very clearly was wrong.
Appendix

Recall from the appendix of chapter 2 that the canonical commons dilemma game is given by the following:\(^\text{17}\)

---

**The Commons Game**

<table>
<thead>
<tr>
<th>Individual cooperates</th>
<th>Fewer than (M) others defect</th>
<th>Exactly (M) others defect</th>
<th>More than (M) others defect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\text{COOPERATE} + \text{REWARD})</td>
<td>(\text{COOPERATE} + \text{REWARD})</td>
<td>(\text{COOPERATE})</td>
</tr>
<tr>
<td>Individual defects</td>
<td>(\text{DEFECT} + \text{REWARD})</td>
<td>(\text{DEFECT})</td>
<td>(\text{DEFECT})</td>
</tr>
</tbody>
</table>

All \(n\) players get a \(\text{REWARD}\) as long as there are no more than \(M\) defectors, but the payoff that defectors get \(\text{(DEFECT)}\) is always higher than the payoff obtained by those who cooperate \(\text{(COOPERATE)}\). However, every player is better off if they all cooperate than if they all defect, because \(\text{COOPERATE} + \text{REWARD} > \text{DEFECT}\). The existence of a \(\text{REWARD}\) payoff term produces a kind of “tipping point” equilibrium in that when a sufficiently high proportion of the population is not opportunistic there are qualitative changes to society that produce a significantly lower level of transaction costs generally. The problem is that any individual nevertheless does better playing \(\text{DEFECT}\). So what’s rational at the margin is what produces the worst outcome in equilibrium.

An important but heretofore unrecognized element of the commons dilemma problem associated with opportunism is that the absolute number of players, \(n\), can affect the payoffs because of the empathy problem. Specifically,

1. The smaller \(n\) is, the more likely it is the case that:
a. Choosing *COOPERATE* is self-enforcing so there is no dilemma in the first place.

b. Choosing *DEFECT* will significantly harm some persons and the resulting guilt costs will be sufficient to drive the net payoff of choosing *DEFECT* too low for *DEFECT* to be the dominant strategy. In this case harm-based moral restraint is sufficient to overcome the commons dilemma.

2. Conversely, the larger $n$ is, the more likely it is the case that:

   a. No one will be perceptibly harmed by choosing *DEFECT*.

   b. If moral restraint is solely harm-based, no guilt will be experienced and therefore the net payoff from *DEFECT* will be unaffected, so a tragedy of the commons outcome may occur *even if people are highly moral*. In this case harm-based moral restraint is insufficient to overcome the commons dilemma.
Endnotes

1 In the next chapter I will explain why misguided attempts to solve the empathy problem in these two ways also worsens another problem – that of greater good rationalization.

2 One could argue that as a philosophical matter it is the total harm that should matter. I do not contend that such an argument is wrongheaded but I believe that we are not likely to be hardwired to think of harm in this way. With the exception of a rare philosopher, such a conclusion would have to be derived from an abstract moral belief that one is taught.

3 That this might be a conjecture of fundamental importance is evidenced by the fact that the distinction between consequentialist and nonconsequentialist theories of morality is one that is of fundamental importance in modern moral philosophy. We will explore this point more fully in the next two chapters.

4 Although I didn’t know it at the time I developed the model below, some of the basic ideas involved had already been addressed long ago in a fascinating paper by James Buchanan (Buchanan 1965).

5 This is an important distinction. Opportunism is less common even among complete strangers when the dealings are one-on-one precisely because in such cases we know our action may significantly harm the other person.

6 Mathematically, as \( n \to \infty \), \( \Delta U_i(x) \to 0 \) and therefore \( g(x)\theta_i E[\Delta U_i(x)] \to 0 \).

7 Now it may be true that if everyone did x all the time it would produce measurable harm, but you are not proposing that. Your doing x in this instance does not constitute everyone doing x all the time or even everyone always doing x in this instance.
Hunt (2007) argues in *Inventing Human Rights: A History* that novels had a dramatic effect on human rights. Her argument illustrates the power of empathy. By reading stories of harm done to others, we identify with them and are compelled to feel as though the event directly involved us. The more engaging is the story and the more realistic are the characters, the stronger that our sense of empathy is aroused. Echoing this sentiment, Richard Rorty (1989) explained in *Contingency, Irony, and Solidarity* that “books…helps us become less cruel.”

One might object that the exponent of 2 is unrealistically high, but that is mistaken. The increase in output actually compares well with Adam Smith’s example of the pin factory, which was based on empirical data for a very simple production process (obviously one would expect even stronger increasing returns for a complex production process). In that example, forcing pin production to occur in firms no larger than 1 worker resulted in at most 20 pins produced per day (Smith considered this to be an upper bound). Allowing pin production to occur in firms that are larger, in his example 10, produced 48000 pins per day! This increase in productivity would require an exponent larger than 4. If Adam Smith was right about the importance of organizing economic activity in large groups, it pays handsomely to address anything that limits the size of groups within which production occurs.

If, instead, I believe that people who earn or have lots of money do so because they either work very hard and/or have made large investments in their own human capital, then stealing from them is not taking something that just as easily could have been mine to begin with; it amounts to stealing the most precious thing any human has – his time.
It is not surprising that we might have this reaction. Through most of our evolutionary history, an unwillingness on the part of A to promote B’s welfare when doing so costs A nothing was likely interpreted by B as a signal of hostility. Given the obvious benefits of reciprocity, B knows that A’s dominant strategy is to always promote B’s welfare in such cases. The exception, of course, is if A believes he benefits from harming (or at least not helping) B.

Another contributing factor to weak response to counterfactual losses is that we generally don’t know with certainty what the alternative gains would be, so must estimate them. Since that estimate is a subjective one made by the person who possesses the relevant local knowledge, it is easy to see how someone can convince himself that while the potential counterfactual cost is high, the expected counterfactual cost is less so because we don’t know for certain.

Ken Binmore (2005) has rightly stressed the inadequacies of moral theories based on what he calls “philosophical skyhooks,” which are nothing more than bald assertions without scientific basis. He advocates a naturalistic approach, which requires analyzing moral behavior vis-à-vis the psychological mechanisms humans actually have, not those that moral philosophers tell us we should have.

In more detailed terms as used in the Appendix, suppose that $n$ is so large that harm-based moral restraint is zero, producing the “tragedy of the commons” outcome. If we add sufficiently strong principled moral restraint, then the dilemma is eliminated completely for any $n$, no matter how large. This is because with sufficiently strong principled moral restraint, $g_2$ is pushed so high that the net payoff for DEFECT will be too low for choosing DEFECT to be a dominant strategy (indeed DEFECT can even
become negative regardless of $n$). This means it is never rational to choose $\text{DEFECT}$ regardless of what anyone else does and regardless of $n$, so the commons dilemma simply disappears because a necessary condition for the commons dilemma to exist is that $\text{COOPERATE} < \text{DEFECT}$. 

15 See Barro and McCleary (2002, 2003) for a discussion of how belief in heaven and hell can affect economic growth. See McCleary and Barro (2006 JEP) for recent and broad review of this growing literature.

16 Religious beliefs in primitive, small group societies are largely preoccupied with explaining the physical world and with currying favor with the gods for better treatment. Comparatively speaking, the religions of modern large group societies are remarkably preoccupied with matters of right and wrong and how people should treat each other.

17 This closely follows the presentation provided by Izquierdo, Gotts, and Polhill (2004).