CHAPTER 9

RESTORATIVE JUSTICE AS EVIDENCE-BASED SENTENCING

LAWRENCE W. SHERMAN
AND HEATHER STRANG

The idea that justice should repair harm, and not just inflict it, has waxed and waned for thousands of years. In the past two decades, that idea—now called restorative justice—has enjoyed a resurgence of discussion and experimentation, especially in the common law countries of British heritage. The operational impact of those developments on sentencing and corrections has been relatively marginal so far. Yet a growing body of systematic evidence has stockpiled ample fuel to power a restorative justice (RJ) movement, if ever a sudden political "tipping point" could set it on fire. With rising awareness of the cost of imprisonment, an economic recession has added more heat to the most incendiary question of all: whether RJ could produce greater public benefit for less money than current practices, especially mass incarceration and probation.

Our conclusion is that RJ can reduce both reoffending and costs, as well as harm to crime victims, in a wide range of applications to existing criminal justice systems. Whether it can reduce imprisonment without increasing crime is a *causal* question that requires further testing, but the evidence so far is encouraging. Whether RJ *should* reduce imprisonment is a *moral* question that cannot be resolved by causal research, given the moral opposition to the utilitarian principle of what "works" found on one side of that debate. Most criticisms of restorative justice are moral rather than causal, despite morally driven attempts to portray RJ as "impractical" for causal reasons. Debates over RJ would be more transparent if each side would distinguish their moral and causal theories of justice, as this chapter attempts to do. Yet we conclude that all decisions about justice are ultimately framed by emotions (Sherman 2003; Sherman and Strang 2011). Any attempt to reform justice must therefore be as emotionally intelligent as it is "evidence-based."

This chapter reviews the conclusions from two decades of RJ innovations as they stand in early 2011. Our primary focus is on the most rigorously evaluated form of RJ: the face-to-face *restorative justice conference* (RJC) that brings together offenders, their victims, and their respective kin and communities, in order to decide what the offender should do to repair the harm that a crime has caused. This method of dispensing justice has been applied in criminal cases on at least three continents, in six major ways:

a. RJ Conferences instead of prosecution, with cautions or no criminal records;
b. RJ Conferences after a guilty plea but before sentencing, as "mitigation";
c. RJ Conferences as the process for making sentencing decisions, with court review;
d. RJ Conferences imposed by a sentence as part of probation or imprisonment;
e. RJ Conferences that precede reentry after prison;
f. RJ Conferences and Life or Death Sentences.

This chapter assesses the evidence for the comparative effectiveness of justice with and without RJ conferences in most of these applications. It sets the stage for that assessment with several building blocks. Section I examines the history and theories of restorative justice, including the definitions of the many varieties that RJ has taken. The history includes primordial and traditional systems of justice, many of which have continued uninterrupted for over 5,000 years. Our discussion of moral theories explains why some of them support RJ while others attack it. Our review of causal theories highlights only some of the many hypotheses about human behavior that fit the facts of restorative justice research. It concludes by showing how research has helped to reformulate theories, in keeping with the scientific method.

Section II summarizes the logic of evaluation research on RJ, which is a comparison of results. The question of whether RJ "works" is inherently comparative, requiring that one must first answer the logical pre-cursor question: *as compared to what*? Increasingly, the question of whether RJ "works" must be framed in terms of cost-effectiveness for government, and in terms of emotions for politics. Running through all these issues is the logic and methods of randomized controlled trials (RCTs), in contrast to more biased methods of making comparisons. We therefore exclude a large body of research on RJ that is potentially biased in favor of it, not by the researchers themselves but by the nature of their research methods.

Section III reports the available research on the six comparisons listed in points a–f above. Much of this research was conducted or designed by the authors, but with an independent evaluator assessing the outcomes in the UK tests, and independent authors designing and conducting the U.S. experiment. The research suggests that the greatest cost-effectiveness of RJ is found with some of the most frequent, serious, and persistent offenders. This conclusion sharply contradicts conventional wisdom among criminal justice policy makers that RJ is more appropriate for minor criminals than for major ones.

Section IV briefly reviews the global social movement promoting wider use of RJ. Led primarily by present or former public servants, this movement has attracted the attention of legislatures from Australia to Indiana. With the exception of New Zealand, however, the movement has failed so far to reshape sentencing and corrections. The key organizations and opportunities are identified, with Gladwell's (2006) description of a "tipping point" presented as the key question for the future of the movement.

Section V presents the eight major conclusions:

1. RJ conferencing is now the most evidence-based strategy in corrections.
2. That evidence is more than sufficient to implement it widely, even while doing more research to answer further questions about it.
3. RJ conferencing is especially cost-effective with some of the more serious criminals, yet
4. Any reduction in prison populations must insure incapacitation of the "worst" offenders.
5. The prospects of expanding RJ depend upon evidence that a person committing a notorious crime would not escape imprisonment by agreeing to an RJ conference.
6. RJ may need the added evidence of statistical risk assessments in determining which offenders and offenses would be eligible for RJCs that would be used instead of prison, rather than in addition to it.
7. Whether or not their offenders are imprisoned, victims can benefit from RJCs in ways that no conventional court process appears able to provide without an RJ conference.
8. Direct benefits of RJC to victims can justify its possible mitigation of retribution and general deterrence, eliminating most legal theory against its use.

I. History and Theories

The history of justice among primates, including chimpanzees and monkeys, is replete with restorative practices. That fact alone offers no logical basis for preferring restoration, since the history of primate justice is also replete with revenge. This section begins by reviewing the long history, which merely describes the larger biohistorical context in which several varieties of RJ may be defined. These definitions allow us to limit our review of current debates to one form of RJ, the RJ Conference (RJC). Those debates embrace both *moral* and *causal* theories of behavior.

History

In the twenty-first century, "restorative justice" is a broadly defined term. Ranging from compulsory financial payments (restitution) to voluntary apologies (remorse) and compulsory community service in marked jackets by "chain gang"–style work groups (called "payback justice" in the UK), "restorative justice" has become almost anything that someone thinks is a pathway to offenders' redemption.

A definition this broad is far too heterogeneous for meaningful policy assessments in sentencing and corrections. It also leaves out values and processes that are intrinsic to the concept (Braithwaite and Strang 2001). Yet that breadth is a useful starting point for surveying the long history of group responses to harms by their members. That survey then allows us to focus on one very specific modern version of the broad idea: restorative justice conferences (RJCs).

The broad concept of rebuilding social ties after aggression encompasses primate behavior that is millions of years old. Primatologist Frans de Waal (2009) describes monkeys who respond to fights between other monkeys by encouraging them to "make up" after the fight. The peacemaker monkeys encourage the winners to "groom" the losers (picking insects out of each other's fur), and vice versa, as an expression of mutual support. De Waal's equally clear descriptions of well-planned acts of group revenge on abusive group members shows that restoration is far from the only response to conflict. Yet the evolutionary advantage of membership in a group large enough to defend itself appears to support ongoing efforts to maintain group harmony. Even self-sacrifice during efforts to save the lives of other group members—such as those who may drown after falling into water—support the limitation on punishment in many species.

Other primate species besides humans practice banishment and even homicide, as well as intergroup warfare. Yet any effort to characterize vengeance as the overriding, genetically driven social response to harm (e.g., Diamond 2008) misses the equally compelling bio-social evidence for forgiveness and reconciliation (Sherman and Strang 2011; McCullough 2008). This is true whether we consider all primates, or only humans—whose capacity for speech broadens the range of possibility for restorative practices.

For millennia, humans have used speech extensively in repairing the harm of crimes. These practices still survive among indigenous people from New Zealand to Canada and the United States to the Middle East, Africa, and South Asia (Braithwaite 1998, 2002). The most common process is the template for the modern RJC. Communities that have suffered crimes assemble a substantial group of people to confer on what should be done. The composition of the group varies across cultures, subject at times to the exclusion of women, offenders, or anyone besides village elders. The conference group may span tribal or familial divisions, and sometimes even languages (Huxley 1939). Yet they all use talk to express emotions, consider evidence about damages, consider precedents, and (usually) agree upon a fair means to repair the harm.

European tribes in the first century C.E. (A.D.) developed highly standardized tariffs for compensation of damages. The systems of *bot* and *wergeld* ("man-price") in Anglo-Saxon law, for example, provided for standard sums in coins for each kind of injury to a man or woman: so much for the loss of an arm, an eye, an ear, or a life—a system that arguably persisted well into Tudor England (Musson 2009). African tribes did the same, using goats or oxen as currency (Huxley 1939). There seems to have been no question of punishing an offender by death or disfigurement in these systems unless the agreement was not kept. A healthy offender was necessary to provide the restorative payments, which even in eighteenth-century Scotland included lifetime support for widows (Braithwaite 2002). This system resembles what is today described as "civil" law, but in its context it was a complete system of "penal" law, with no other penalties imposed. Only the coming of strong nation-states and kings seems to have displaced this community-based system with the idea of crime as an offense against the "King's peace," rather than the victim (Christie 1977).

Definitions

Three key concepts must be defined in modern restorative justice: mediation, restitution, and restorative justice conferencing.

Mediation

Since the 1960s, many innovations in criminal justice have tested diversion of cases from prosecution by the state. One of the most widely used practices is *mediation*, which we define as an out-of-court agreement between a victim and an offender about what the offender should do for the victim (usually, but not always, a cash payment). This process is usually managed face-to-face (or in "shuttle diplomacy") by a single trained professional called a "mediator," whose work is modeled on civil and labor disputes. There is no formal interaction ritual (Collins 2004) that unleashes emotions. Nor are members of the family or friends of either the victim or offender (usually) allowed to be present. Use of the process is widely seen as a means of avoiding imprisonment, yet there is only weak evidence from comparing like-with-like cases treated by mediation or prosecution.

Restitution

Results similar to those from mediation can be imposed after prosecution by a judge as *restitution*. This decision will usually be imposed without direct contact between offender and victim. As in fines paid to the state, restitution for the victim is imposed on the offender with the threat of imprisonment for failure to pay (Schneider 1986). This does not deter millions of American offenders annually from failing to pay, with many going to prison as a result. Nor does it guarantee that victims receive the money, even when it is paid. Over $10 million in restitution payments to crime victims in Philadelphia remains in the hands of the courts, for example, due to their inability to locate the victims (Malvestuto 2009).

Restorative Justice Conferencing (RJC)

This third method of RJ, about which the best and most research is available, differs substantially from both mediation and restitution. Sometimes called "family group conferencing" (in New Zealand and Australia) or "youth conferencing" (in Northern Ireland), RJC is defined by three elements:

1. The presence in one room at the same time of one or more victims, the victims' offender(s) who accept(s) responsibility (if not legal guilt) for a crime, the respective friends or families of victims and offenders present, and a facilitator to guide their discussion;
2. The provision of enough time for everyone present (including the offender) to express their emotions about the crime and the harm it has caused, as well as to suggest what should be done about it;
3. The facilitator's effort to reach a conclusion of the conference that expresses a group consensus about what the offender should do.

There are many other procedures described as "restorative justice" that may look like RJC, but which lack one of the three defining elements. Victims who visit prisons to talk to groups of prisoners, as in the Sycamore Tree Program, do not create an RJC because the victims' own offenders are not present. Youth justice panels (Sherman and Strang 2007), in which juvenile delinquents appear before a group of community residents and criminal justice officials, fail to meet the RJC definition on three grounds: victims are not usually present, offenders are rarely allowed or encouraged to say very much, and stakeholders to the crime are excluded from the decisions made.

As noted above, *the rest of this chapter focuses solely on RJCs*, for several reasons. One is that RJCs are the modern equivalent of the most widely practiced form of restorative justice in human history. Another is that RJCs have been subjected to more rigorous testing than any other variety of restorative justice. A third is that RJCs have been used in six different ways across the criminal justice process, a far broader reach than any other form of RJ. Fourth, and most important, is that RJCs are the version of RJ that is closest to a "tipping point" of mainstream adoption in developed, common law nations.

Moral Theories

Two major moral theories dominate the policy debate on RJCs in sentencing and corrections. One is *utilitarianism*. The other is *deontologism*, or non-consequentialism. These theories both arose during the eighteenth-century Enlightenment, with the Italian philosopher Cesare Beccaria (1764) the original exponent of utilitarianism in sentencing and the German philosopher Immanuel Kant (1785) the exponent of non-consequentialism (but see Murphy 1987). Libraries are filled with the interpretation and debates about these competing perspectives. Space limits us to one paragraph to summarize these ideas, with another to apply them to RJC.

Utilitarians seek decisions that cause the greatest good for the greatest number. Deontologists seek decisions based on moral principles such as culpability and desert, regardless of the consequences of those principles in any given case (Moore 1987). Utilitarianism is the philosophy of empirical cost-benefit analysis, and may be "bounded" by such principles as the human right of offenders not to be tortured, and not to be imprisoned without a fair trial (Frase, chapter 5 of this volume). Deontologism is a philosophy of consistency in both how and what decisions are made, with equality of both procedural and distributive justice (von Hirsch 1976). This preference for consistency led Kant to reject mercy as unfairly inconsistent, a problem that directly affects restorative justice (Mason 2002; Ashworth 2002).

Broadly speaking, utilitarians support RJC *if and only if* it leads to less harm than not using RJC. Deontologists oppose it. Utilitarians therefore reject RJC where it causes more harm, but support RJC wherever it reduces harm to victims, offenders and communities, taking all stakeholders together in the calculation. But the inconsistency across cases that utilitarians allow is unacceptable to "just desert" deontologists, who refuse to use offenders (or victims) as a means to the end of less total harm. Desert theory prefers an exact tariff looking backward, not forward, with the goal of minimizing inequalities in punishment severity from one human being to the next (von Hirsch and Ashworth 2005). Deontologism, in its strong form, privileges moral judgments over possible facts, such as that this approach could cause more crime (or punishment) than utilitarianism, inflict more harm on offenders overall, or deprive children of health care because of high imprisonment costs. Such consequences are arguably not relevant to a principled Kantian analysis of fairness in sentencing. That is why the rest of this chapter is relevant only to a bounded utilitarian framework for sentencing policy (e.g., Dignan 2003). Pure Kantians (but see Murphy 1987) can stop right here.

Causal Theories

The effects of RJC on victims and crime embrace a wide range of theories. One theory about why *victims* benefit from RJC is that it reduces their post-traumatic stress symptoms. By discussing the memory of the crime in a safe place, especially with the criminal present, RJC may fit the hypothesis of desensitizing the victim's conditioned response of anxiety whenever recalling the crime (Foa et al. 1991). Another theory about why victims benefit is procedural justice theory (Tyler 1990), which holds that people find the legal system more legitimate if they are treated with equality and respect by legal authorities. Because victims are widely reported to feel anger at being silenced by common law adversarial procedures (Strang 2002), the experience of being heard by the offender and others may heal that anger in ways that a courtroom would never allow.

A third theory predicts the effect of RJC on both victims and offenders, as well as others present at the RJC. Collins (2004) hypothesizes that meetings like RJCs bring people together in a way that lifts their energy and their commitment to shared norms. Thus an RJ conference could help both victims and offenders to become more positive about their own lives, as well as to be more morally committed to obeying the law.

The central question raised by policy makers is the effect of RJC on repeat offending. Answers to this question embrace numerous criminological theories (Sherman 2003): reintegrative shaming (Braithwaite 1989), life-course "turning points" of desistance from crime (Laub and Sampson 2003), and even specific deterrence from the enormous trauma offenders can suffer from undergoing an RJC (see Woolf 2008).

The effect of RJC on crime rates in a community or society is the single issue that may appear to lack theoretical clarity. Yet clarity can be found easily in certain kinds of facts. These facts require comparisons of entire communities with and without RJC in widespread use. While this may appear difficult in a Western context, it might be easily achieved where community-level units of analysis are available in large samples. Using police stations in India, for example, would encompass huge populations in each station. There are about 100,000 such stations. Experiments assigning some stations to RJC would offer evidence on the three key components of general deterrence theory: community-level perceptions of certainty, severity, and speed of punishment.

General deterrence would predict that RJC *increases* crime if it reduces the certainty, severity, or speed of punishment. Yet RJC could just as easily *reduce* crime by increasing *certainty* and *speed* of sanctions (Sherman and Strang 2007) even if the *severity* appears to be less than in conventional justice (CJ). Moreover, a public perception that RJC constitutes a harsh enough punitive experience for rational offenders to avoid it could allow RJC to satisfy general deterrence doctrine. Consistent with de Waal's (2009) argument that all primates experience empathy for other people's pain, RJC may create traumatic stress for the offender, who must listen to a victim's tears and pain for hours. As one delinquent in Northern Ireland said after stealing an elderly woman's car, it was so painful to see her crying that he decided not to commit crimes against people any more (BBC Radio 4, 2010). This kind of specific deterrent effect can contribute to a general deterrent effect by word-of-mouth (Zimring and Hawkins 1973), especially if it is coupled with a high certainty of apprehension and speedy sanctioning in this way (Durlauf and Nagin, 2011).

II. Comparing Results

Far too few innovations in criminal justice—let alone long-standing practices—have been evaluated using rigorous methods. In that respect, RJC stands out as one of the most rigorously tested innovations in sentencing and corrections (Ruth and Reitz 2003, 56–66). The main reason for this distinction is the use of *randomized controlled trials*, or RCTs, which provide the strongest method possible for an unbiased test of the effects of RJC on victims, criminals, recidivism, and crime rates. While the central questions of corrections and sentencing have been almost devoid of such rigorous testing, RJC has now been subjected to 12 high-integrity experiments with a total of 3,000 cases, spread over three continents. We are unable to find another aspect of sentencing that has been subjected to an equal level of scrutiny, in which the exact same protocol has been used by personnel with identical training from the same trainer—which is exactly how the RJC evidence was developed.

This section briefly reviews the concepts needed to understand and assess the claim that RJC is supported by the best evidence available. Rather than RJC having "mixed evidence," as a senior official of the UK Ministry of Justice recently described it, the truth is that prison itself has far less evidence supporting its use than RJC (Durlauf and Nagin 2011). The central reason for the lack of evidence on most forms of sentencing is the absence of the most reliable method for testing causal theories: experimentation. Experiments differ from other kinds of research in their use of *manipulation* of the factors of interest, in addition to *observation* of behavior over time.

One purist adage holds that we can infer "no causation without experimentation." Others would argue that we can come close to inferring causes without manipulating key factors, especially with factors causing very large effects—such as effects of smoking on cancer. Yet when it comes to sentencing, the experimental evidence to date suggests no large effects. Hence progress in sentencing will arguably require a substantial investment in experiments that detect small to moderate effects. Given the millions of people sentenced annually, even small differences in sentencing practices can cause large numbers of crimes to occur (or not), with differences in savings or spending of billions of dollars.

RJC is an excellent example of this claim. Even though the effects of RJC on crime are small to moderate, they can produce enormous cost savings. Few policy makers may understand, however, *why* the evidence of small effects is so compelling, and so important. The answer depends on four key concepts: unbiased comparisons, statistical power, replication, and cost-effectiveness. Once the reader grasps these concepts, it is an easy step to multiply small differences across huge populations to conclude that RJC can create enormous benefits.

Unbiased Comparisons

Public policy debates constantly cite comparisons as evidence supporting a policy choice. Yet the rhetoric of such comparisons rarely acknowledges their flaw. Most such comparisons suffer from a *statistical* bias, or prejudice. Like racial and religious prejudice, statistical bias is a false claim made with much self-assurance, based on flawed empirical premises. Yet statistical bias is harder to spot. It is found wherever someone implies that a comparison is fair, when it is in fact biased in favor of one side over another. The best way to detect such bias is to ask the key question: *Are there any other factors that might explain the difference between two groups besides the one being cited?*

Racial bias itself is a good example of statistical bias. When someone cites the comparison of murder rates between white and black people in the United States, they may imply that black people are more violent than whites because of their race.

Such a claim obviously fails to consider differences between the races in the age structure of their populations (younger people are more violent than older people), their respective household structures (married men are less violent than single men), and the historic discrimination in education and employment that has reduced the "marriageability" of black men compared to white men. The list of other differences between blacks and whites, other than race, is almost endless. Hence to identify one and only one difference associated with different murder rates is a sort of statistical lie, or at least a statement that is indifferent to the truth about the number of competing explanations that are equally plausible.

The same is true when comparisons are made about sentencing. It is often claimed that as the imprisonment rate in the United States increased, the crime rates declined. This claim is actually incorrect, since in many states crime rates and prison populations have declined together. But to the extent that any general claim can be made about crime going down when imprisonment went up, it necessarily disregards hundreds of other factors that changed at the same time: the age structure of the population, the economy, immigration by low-crime populations (especially into high-poverty areas), and many more. To disregard these competing explanations is to make a claim that is *biased* in favor of one theory over others.

Even proponents of restorative justice are guilty of such bias. Many evaluations of RJ and RJC make comparisons between people who have volunteered for RJ and those who have not. These comparisons show that offenders who have been through RJ have lower rates of repeat offending than "similar" offenders who have not. But the "similarity" of the offenders is gravely in doubt unless they were both offered RJC and were willing to undertake it. Offenders who refuse RJC are by definition different from those who agree to it. A refusal to meet a victim could well predict a higher rate of offending. Hence the correlation may be backward: it is not that RJC causes less recidivism, but that people who are less likely to recidivate are also more likely to agree to RJC.

A "spurious" explanation may be defined as the explanation of a correlation between two facts (doing RJC and having low recidivism) by an underlying *third* factor common to both. Spuriousness is simply one kind of statistical bias. It is also the most important reason that RJC cannot be evaluated by comparing cases that "naturally" get RJC to those that do not. It also explains that *historical* comparisons between RJC cases this year to similar cases last year that were not offered RJC are biased: such comparisons usually leave out the people who refused it this year, but include likely refusers in the historical sample. Finally, spuriousness plagues the comparison of *completers* of a treatment to a sample of people similar to those like the *starters* of any treatment. Since completing the program may be caused by the same third factor that causes low recidivism, the comparison to completers only is biased in favor of showing that the program works.

Our insistence on accepting these brute facts of causal inference is often seen as an insult to the many hard-working evaluators who eke out the best evidence they can from innovative and courageous programs. We certainly mean them no disrespect. Our position is to praise the innovators, as well as the evaluators, for adding to the knowledge base of experience and evidence. But as Iain Chalmers (2003), a founder of the movement for evidence-based medicine, has observed, the main concern should not be people who are running or evaluating treatments. The primary duty for all interventions is to insure that they do less harm than good. Our duty to offenders and victims is to become as certain as possible about the likely effects of RJC, and then to make it clear to policy makers why such evidence is far less "mixed" than the highly biased claims they accept about prison and other untested policies.

The reason that the evidence is not "mixed" is that in 12 independent experiments called "randomized controlled trials," the average effect of restorative justice was a statistically significant *pattern* of reducing repeat offending among offenders willing to meet with their victims. These offenders were compared to almost identical offenders who were also willing to meet with their victims, but who were (by lottery) *not* assigned to do so. These experiments were undertaken at different points of the criminal justice system, and the samples were not identical across the 12 tests. Yet that is all the more reason to conclude from the evidence that RJC is likely to reduce repeat offending wherever it is used. That reason derives almost entirely from the lack of bias used in this kind of experiment (an RCT, or randomized controlled trial).

Statisticians generally consider comparisons in well-executed RCTs to be *unbiased* in relation to the program being tested. The reason that they are unbiased is that they usually rule out all other possible explanations. By using a lottery method of equal chances for each unit to be chosen for one treatment or another, RCTs make comparisons between groups that are close to identical in their distributions of all other factors that could explain a difference in results—except for the difference that is intentionally assigned to them. One group had RJC; the other did not. The difference in repeat offending between those two groups is unlikely to be caused by any other factor. As long as the difference is not due to chance, that is, it is "statistically significant," then that difference is likely to be due to RJC. That is what statisticians say in general, as long as the RCTs are well-executed.

Not all of the RCTs of RJC have been equally well-executed. Yet all of them arguably have less bias than any other evaluations of RJC that do not use random assignment as the basis for comparing offenders who did or did not participate in an RJC. That bias stems from creating a more level playing field at the outset of a comparison than used in study designs other than RCTs.

Unfortunately, by the high standard of the frequency rate of repeat convictions (as opposed to arrests), all of the *individual* results have been statistically "insignificant." Fortunately, the *combined pattern* of all the results taken together is clearly significant, or unlikely to be due to chance. This is true for both an independent evaluation of seven RCTs in the UK (Shapland et al. 2008) and for our own review of all 12 RCTs ever reported (Strang and Sherman, forthcoming). Understanding why the pattern can be significant while the individual tests are not requires that we address the concept of statistical power, as well as the concept of replication.

Statistical Power

Policy makers are often unduly impressed with the idea of "statistical significance." This view comes, of course, from researchers themselves, who have often used the idea of "significance" as a simplistic formula for passing judgments on programs. One example of misuse of the concept, seen in hindsight, is the Report to the U.S. Congress by Sherman and his colleagues (1997) that required at least two "significant" results to conclude that a crime prevention program is effective.

What is wrong with this approach? One answer is that it confuses *certainty* of a result with the magnitude of a result, or how large the effect size is (Weisburd et al. 2003, 2001). Another answer is that emphasizing *significance* may blind the debate to the equally important emphasis on *power*, or the capacity of a test to detect a true effect as "significant." Many results that are "statistically insignificant" may be unable to detect effects that are large enough to save millions of lives, or to prevent millions of crimes. The lack of adequate power for a test is increasingly recognized as another form of bias, one that can easily mislead policy makers into thinking a strategy is ineffective when it is actually very beneficial.

The three key elements of statistical power are complicated but important. The most obvious element is *sample size*. A small sample, other things equal, is less likely to detect a true effect than a larger sample. But other things are not usually equal, and other things matter. The two other factors in power are *effect size* and sample *heterogeneity* (Cohen 1988). The idea of effect size describes how big an effect is, such as heating a pot of water on a gas stove with a low flame (small effect) or a high flame (larger effect, heating the same amount of water more quickly). More complex is the idea of sample heterogeneity, which underlies such statistical concepts as variance and standard deviation.

A *heterogeneous sample* is comprised of units that are different from each other in characteristics that may affect the results of an experiment for each member of the sample. All samples have some such differences, but some samples are far more heterogeneous than others. Imagine a sample of first offenders in which all are white, female, and between the ages of 16 and 17. Compare that sample to a second one in which the offenders have anywhere from 0 to 50 previous criminal convictions, half are white, half are male, and the age range varies from 12 to 83. The first sample is far more homogeneous than the second. Other things equal (like sample size and effect size), the chances of a statistically significant result are far lower with the second sample than with the first (Weisburd 1993).

Taken together, sample size, effect size, and heterogeneity can bias a test toward or against a "statistically significant" result. In very large sample sizes, often found in medical research, tiny differences in rates can lead to changes in medical treatments of millions of people (Haas et al. 2004). Yet such results are only statistically significant because governments spent large sums of money on the medical research needed to obtain the sample size needed to detect the difference. Since governments rarely spend on RCTs in crime prevention even one-tenth of what they spend on medical RCTs, it can be said that government funding is biased against finding statistically significant results in crime prevention. This bias is entirely due to a lack of statistical power.

Criminologists have generally failed to educate policy makers, and even each other, about statistical power. Yet it is impossible to evaluate the evidence on RJC without grasping the essence of this idea. The fact that a series of randomized trials might be unable to obtain large enough samples (from change-resistant criminal justice agencies) to be *statistically powerful* in each test should not prejudice policy makers against the benefits of small to moderate effects of RJC. It is arguably more important to consider the overall pattern of results, using the greater statistical power generated by the combined analysis of repeated replications of the tests.

Replications

The goal of similar results from repeated experiments is often cited as a test of scientific truth. This test is known as "replication," whereby independent researchers undertaking the same experiment can get the same result, over and over again. The image of an apple falling on Isaac Newton's head suggests how often the experiment of dropping something has tested the gravity hypothesis. Never mind the fact that replication has become increasingly problematic in a wide range of sciences, for reasons that are not at all clear (Lehrer 2010). The fact is that RJC experiments undertaken by different researchers, in different countries, and in different parts of the criminal justice system, have generally obtained similar results. Thus we can conclude that this evidence has passed at least one test of replication: the test of identical methods under different conditions.

A second test of replication has not, as yet, been attempted: one that uses identical methods under the same conditions but with different samples (see, e.g., Sherman 1992). This could be an important addition to evidence-based policy. We know, for example, that in Australia the effects of RJC on Aboriginal offenders was the opposite of its effect on white offenders, even within the same experiments with the same police officers delivering the RJC (Sherman 2006). Each society that undertakes RJC may have alienated minority groups who might react differently to RJC's attempt to create a shared morality about obeying the law. Thus separate tests of RJC on majority and minority samples could be a vitally important kind of replication.

A more general problem of replication is to learn whether what works in one community (or nation) will work in another. Relatively few social interventions have been tested in the same way across a wide range of communities, let alone with different legal systems and languages.

The most fundamental purpose of replication remains to test whether a finding reflects a general pattern or a "fluke" result from an unusual set of conditions. These conditions could include the people leading the experiment, the people administering the treatment, the victims or offenders in the sample, or the communities in which the offenders may be committing repeat offenses. The more often a finding is repeated, the less likely it is that the finding was due to chance. The greater the proportion of all tests with the same results, the same conclusion can be reached. If 10 out of 12 tests reach positive results, then the risk of positive results being due to chance is greatly diminished over having just one test result.

Just as these facts of replication strengthen the general findings about RJC, they also weaken the specific findings about using RJC at different stages of the criminal justice system. There are 12 independent tests of RJC in RCTs, but only one test of RJC with drinking-driving violations. There are four RCTs of violent crime, but only two RCTs with an all-adult (over-18) sample. For each subgroup, the uncertainty about RJC effects for only that subgroup is greater than the uncertainty about RJC in general.

These and other caveats are required to prevent any bias in our presentation of results. It is also necessary for those biased against RJC to compare the evidence base that we discuss with the evidence for other sentencing and corrections practices. All evidence is relative. In an absolute sense, there is far more evidence to be gained about all aspects of life, let alone criminology. In a relative sense, the evidence for RJC is far in front of the evidence for any other new initiative that attempts to serve both victims and offenders.

Whether crime victims are important, however, is something that many civil servants in the United Kingdom would dispute, at least in private. While they might accept the unequivocal evidence that victims who consent to RJC are far more satisfied by experiencing it than if they do not (Strang 2002; Sherman et al. 2005), they might discount it because it does not matter to their budgets. Such issues show how important it is to reveal the definitions of cost-effectiveness, and to make them central to the policy analysis.

Cost-Effectiveness

How is the cost-effectiveness of sentencing and corrections to be defined? The United Kingdom is far ahead of the United States or Australia in even thinking about this question. The United Kingdom's answers pose a clear moral and political dilemma. While the cost of crime can be estimated from the kind of detailed studies that have been done on both sides of the Atlantic, the biggest cost of crime is to the victims of crime (see, e.g., Shapland et al. 2008). Reducing the cost of crime will therefore help victims far more than it will help governments to balance their budgets.

In consultations with UK civil servants—the permanent officials who run the government at the direction of a succession of Members of Parliament who direct the officials—advocates for RJC have been told that reducing crime costs to victims does not matter. In the view of these officials, all that should drive government policy is saving money for taxpayers in general, not for those who have become crime victims. That is, unless RJC can reduce the costs of policing, prosecution, sentencing, probation and imprisonment, *relative to the cost of providing RJC*, then it cannot be considered to be cost-effective.

This position can be argued morally but not technically. Yet even on its own terms, the question posed can be answered in the affirmative. Does RJC save the government money? According to calculations by Shapland et al. (2008), it does in some tests, if not all. Moreover, it can save so much money in some tests that the average effect of RJC across all tests in the United Kingdom is to save money for the government.

The simple formula for reaching these conclusions requires a complex menu. The menu assigns two prices to each type of crime. One is the overall cost of that type of crime, based on surveys of crime victims, their employers, the health care system, and welfare benefits drawn. The other is the cost of crime just to the government, which is currently calculated as just criminal justice system costs (arguably an error, since taxpayers pay for most health and welfare costs). For each crime of which an offender is convicted, the two prices can be applied from the menu. The price of all the crimes is then summed across each offender's history for two years after random assignment to RJC or not. Then the average costs of crime for all offenders in the RJC group is subtracted from the average cost of crime for all offenders in the non-RJC control group. This difference is defined as the "benefit" of RJC, either in total costs or criminal justice costs.

The average cost of producing each RJC is then computed across all RJC cases in the experiment. This includes largely the salaries and expenses of the people delivering RJ services, usually police officers. This cost includes meetings with offenders and victims who refuse to consent to RJC, which happened in over half of all cases in the UK tests. The average cost of all activities needed to produce the number of RJC cases in the experiment is defined as the cost. By dividing the cost into the benefit, as calculated above, the cost-benefit of RJC can be calculated in two ways—one for each definition of the benefit. The results of these calculations are presented below in figure 9.3, just for the seven UK experiments, after the following section presents descriptions of the applications for which the calculations were done.

III. Applications

The major applications of RJC have been spread across six basic stages of the criminal justice process under the common law systems of the Australia, the United Kingdom, and the United States. These applications take two basic forms: diversion from prosecution as an *alternative* to criminal justice, and in addition to prosecution as a *supplement* to criminal justice. Using a systematic review process to identify all RCTs of RJC using random assignment *after* asking victims and offenders for their consent to participate in RJC,[1] Strang and Sherman (forthcoming) have compiled the results of 12 RCTs with almost 3,000 cases. This section reports on the evidence from that review. That evidence is first presented across all the different points of criminal justice en masse, and then again on a stage-by-stage basis.

The overall evidence is presented in figures 9.1 and 9.2. Figure 9.1 displays a brief description and sample size of each of the RCTs, arranged roughly in the chronological order in which the experiments were conducted. Figure 9.2 displays the

	Offender N
1. Australia RISE Canberra <30 years violence	121
2. Australia RISE Canberra juvenile personal property	248
3. Australia RISE Canberra juvenile shoplifting	142
4. Australia RISE Canberra Driving Under influence	897
5. US Indianapolis juvenile property/violence	782
6. UK Northumbria juvenile property/violence	208
7. UK Northumbria adult property	63
8. UK Northumbria adult assault	44
9. UK London robbery pre-sentence	88
10. UK London burglary pre-sentence	167
11. UK Thames Valley violence - probation	64
12. UK Thames Valley violence - prison before re-entry	94
Total offender N =	2,918

Figure 9.1. Description: 12 RCTs of Restorative Justice Conferencing.

results of each of the RCTs, arranged in the order of their effect sizes using a standardized mean difference [data not displayed] called Cohen's D (Cohen 1988). Rather than explaining the complexities of this statistic, the chapter presents the much more accessible calculation of the percentage differences in each experiment.

For each RCT, figure 9.2 displays the percentage difference in the frequency of convictions over a two-year period after the random assignment. The comparison shows first the average number of convictions (crimes) per offender per year in that period among consenting offenders assigned to RJC. It then shows the same average for the consenting group *not* assigned to RJC. In the column on the far right, figure 9.2 shows the percentage difference in the frequency of convictions between the two groups.

The offenders did not always experience the treatment to which they were assigned. Some of the offenders assigned to RJC never completed it, for a variety of reasons. Almost all offenders assigned not to get RJC remained in that condition. The evidence therefore may understate the potential benefit of RJC, since so many cases without RJC were included in the group randomly assigned to it. This "intent-to-treat" approach is considered by statisticians as the least worst way to analyze experiments with relatively small sample sizes (Peto et al. 1976).

Figure 9.2 shows that the frequency of post-random-assignment convictions was lower for the offenders assigned to RJC than for those not assigned in 10 out of 12 RCTs. It also shows that the average of the differences between RJC and non-RJC assigned groups was 22 percent fewer convictions for those randomly assigned to RJC than for those assigned not to have RJC. While this method of analysis is not preferred by statisticians, it is preferred by policy makers for ease of communication to a wide range of audiences.

By displaying the raw averages of criminal convictions per offender in each group, figure 9.2 allows anyone to calculate the percentage differences on their own. They can also estimate the number of crimes prevented in each test, which varies according to the frequency of offending in each sample, independent of whether RJC was used. The latter can be estimated by simply inspecting the non-RJC conviction

Experiment	Mean Crimes Per Offender After RJC- Assignment	Mean Crimes Per Offender After No RJC- Assignment	Percent Difference: Mean of RJC compared to Mean of No-RJC
Canberra Drink-Driving	0.09	0.06	+50%
Canberra Juvenile Property	1.02	0.69	+48%
London Robbery	1.85	2.01	–8%
London Burglary	2.88	3.41	–16%
Thames Valley Prison	1.39	2.06	–33%
Canberra Juvenile Shoplifting	0.57	0.82	–30%
Northumbria Magistrates' Property	3.03	3.97	–24%
Indianapolis Juvenile	1.29	1.67	–23%
Thames Valley Probation	0.59	1.32	–55%
Northumbria Juvenile Warnings	0.41	0.79	–48%
Canberra Youth Violence (under 30)	0.42	0.88	–52%
Northumbria Magistrates' Assault	0.82	2.08	–61%
Average of Average Differences of Treatment Groups	NA	NA	–22%

Figure 9.2. Results: 12 RCTs of Restorative Justice Conferencing.

rates, which vary from .09 convictions per offender per year in the drinking-driving experiment to 3.03 convictions per year in the Northumbria Magistrates' Court property crime experiment.

The preferred method for synthesizing these results is called a "meta-analysis," in which a more complex calculation gives each experiment a different weight based upon the number of cases it contains. While this procedure has important advantages, it has one large disadvantage: the sample size of each experiment bears an unknown relationship to the proportions of cases for which RJC may be used. In these 12 RCTs, for example, almost one-third of all the cases come from a drinking-driving experiment in Canberra. Given the negative effect of RJC in causing more offending (but at a very low rate), it is unlikely that RJC will be used for drinking-driving cases in the future. It certainly has not been used in Canberra since the results of the experiment were announced. Yet the preferred method gives substantial weight to a type of crime now deemed irrelevant.

Displaying the raw data for each experiment, figure 9.2 still omits important information for statisticians, including the heterogeneity of each sample, as well as the overall statistical power of each test. Yet it still provides a reasonable basis for

				Cost: Benefit	
SITE	RJ COST	CJ Benefit**	Total Benefit	CJ Ratio	Total Ratio
London	598,848	1,817,426	8,261,028	1:3	1:14
Northumbria	275,411	70,420	320,125	1:0.26	1:1.2
Thames Valley	222,463	101,520	461,455	1:0.46	1:2
Total	1,096,722	1,989,734	9,042,608	1:1.8	1:8

Figure 9.3. * Cost: Benefit.
*** Computed from Shapland et al, 2008. All amounts expressed in Pounds Sterling. **CJ benefit estimated at an average 22 percent of total costs of crime.**

non-statisticians to assess the results of these experiments, individually and overall. It is especially important to see it this way in order to assess the cost-effectiveness of RJC, as in the data from the seven UK experiments across three sites presented in figure 9.3.

Figure 9.3 shows that across all seven UK experiments, RJC saved eight times as much money as it cost to deliver, earning that payback in just two years. That calculation is based on total costs of crime, including costs to victims. But even when we examine only the costs saved to the criminal justice system, the payback is still "profitable." RJC saved 1.8 times as much money in the costs of crime to the criminal justice system.

What figure 9.3 also shows is that for two of the sites, the cost of RJC was greater than the costs saved just to criminal justice. In Thames Valley and Northumbria, the experiments still saved money in total costs of crime. But delivering RJC to anyone other than the highly serious and active criminals pleading guilty in the London Crown Courts did not, on average, save the criminal justice system money. This finding is all the more striking because the percentage reductions in crime frequency were so much larger in Thames Valley correctional cases and in Northumbria non-prisonable court cases than in the London Crown Courts robbery and burglary cases, almost all of which entailed a prison sentence. What this means is that when the London offenders got out of prison, they committed less costly crimes to the taxpayer, as well as to victims of crime.

The data in figure 9.3 emphasize the importance of using a crime harm index (CHI) to evaluate sentencing and corrections, rather than simply counting crimes or convictions (Sherman 2011). A CHI approach abandons the myth that each crime causes equal harm to society, and that mere counts of crime have any meaningful role in policy analysis. By contrasting the count results of figure 9.3 with the cost results of figure 9.2, the reader can see stark evidence of the difference between assuming that "all crimes are created equal" and the more accurate premise that they are not. For the time being, however, it is almost impossible to discuss sentencing options without accepting raw counts of undifferentiated types of crime as the primary criterion for success or failure.

RJ Conferences Instead of Prosecution

The evidence from these 12 RCTs includes five experiments that diverted offenders from criminal convictions as a part of RJ. Two of these, however, were juvenile experiments in which diversion with RJC was compared to diversion without RJC (Northumbria and Indianapolis). These two experiments provide little basis for a more pressing policy question in an era of cost-cutting: Could RJC provide a low-cost alternative to prosecution of adults, without increasing crime?

The only answer to this question so far for adult offenders is a clear but equivocal statement: it depends. Both diversion experiments were conducted in Canberra, Australia, a jurisdiction with historically low rates of imprisonment. Almost no cases in the control group (prosecution) were sentenced to prison. That said, it is striking that the violence cases assigned to prosecution did so much worse than the cases diverted to RJC, especially for adults over 18; there was virtually no effect of RJC for the under-18s, thereby making the overall estimate dependent on the adult subgroup [data not displayed]. Since the adult subgroup is too small for adequate statistical power, we cannot make too much of this finding without replication. Unfortunately, no such replication was allowed by the central government in the UK experiments.

At the same time, the victimless RJC conferences in Canberra did worse than prosecution in relation to the low base rate of repeat offending. Technically, we could exclude the drinking-driving diversion to RJC from the analysis altogether, on the grounds that there was never a personal victim of crime present in the conferences to make offenders feel ashamed. But because we conducted the experiment, transparency requires that we include it the reporting and discussion.

If one concludes that the drinking-driving experiment is irrelevant, then there is a good case to replicate the diversion from prosecution to RJC for adult offenders in minor assaults. The large effect of adding RJC to such cases in Northumbria (figure 9.2) combined with the Canberra findings yields an average of 56 percent fewer reconvictions with RJC over the two experiments. Since neither the Northumbria nor Canberra control group received imprisonment, the case is even stronger for reducing reconvictions. Whether it would save money is a different question.

RJ Conferences Before Sentencing

Figure 9.2 shows four tests of using RJC after guilty pleas but before sentencing. In all four cases, RJC yielded fewer reconvictions. But it is not clear whether RJC was cost-effective in all four cases, measured only by criminal justice costs. Shapland and her colleagues (2008) did not separately report the cost of crime effects of each experiment in each site, so we simply do not know whether the results for RJC before sentencing are consistently cost-effective. Yet there is clearly a strong case for replications, and for a policy rollout in the serious burglary and robbery cases of the kind that we tested in London—where prison sentences are longest and most costly to a government presently trying to reduce the prison population.

RJ Conferences as Sentencing Decisions

A third application of RJC has virtually no unbiased evidence about its effects. There is not a single RCT on the use of RJC to decide what the sentence should be, including prison versus no prison. Yet that is exactly what is done in thousands of cases in New Zealand, Canada, Australia, and elsewhere. For that reason alone, we offer the following discussion, with no conclusions on the cost-effectiveness of this application.

While the RJ paradigm embraces very different values and processes from those of formal justice, maximizing the use of RJ for the benefit of both offenders and their victims requires them to intersect in a fashion that preserves the conventions of both. Only then can RJ be used for serious offences for which diversion alone would not generally be acceptable. Positioning RJ within the court system allows it to be used either in addition to formal justice processes—as a referral from court, or in conjunction with it—as in Canadian sentencing circles (Green 1997). Most often, RJ tends to be restricted to juvenile offenders, but increasingly courts are using it for adults as well. It is certainly restricted to cases in which the offender has admitted responsibility for the harm suffered by the victim. RJC is not a forum for adjudication of guilt but a means of deciding in a consensual way what can be done to repair an acknowledged harm.

RJ as a Referral from Court

Over the past two decades, Australia and New Zealand have allowed cases to be referred to RJ by the courts. In South Australia and the Australian Capital Territory, for example, where police have the discretion to divert cases from the court to RJ themselves, judges may also decide that cases coming before them would more appropriately be dealt with by RJC alone. Judges can then divert these case completely out of the court system. This is often a response to the tendency of police to err on the side of caution. Rather than diverting cases to RJC themselves, police may prefer judicial direction about what kinds of cases ought to be dealt with in court.

More often than diverting cases, however, Antipodean courts treat RJC as a sentencing option for offenders before them. In some jurisdictions this is now obligatory. Since 2002 in New Zealand, for instance, RJC language has been incorporated into its sentencing provisions. Courts are now required to consider the results of RJC and to take into account "offer, agreement, response or measure to make amends" (New Zealand Sentencing Act 2002, No. 9, S10). In deciding whether and to what extent to take these into account, the court must decide whether the offer was genuine and feasible and whether it has been accepted by the victim as expiation or in mitigation of the wrong.

New Zealand uses the provisions of this act in both juvenile and adult cases. In a serious case of aggravated robbery committed by a 17-year-old youth, for example, the judge in the case noted that the circumstances of the offence would normally require a four-year sentence of imprisonment. In light of the outcome of the RJ conference, however, a custodial sentence of ten months was imposed (*Crown v. Makatuu Folaumoeloa* 2004).

While New Zealand, like other countries, has used RJ mainly in juvenile cases, it has been at the forefront in piloting the use of RJ for adults. Indeed, the introduction of the 2002 act may have been strongly influenced by a signal case involving a very serious assault by an adult, *R. v. Clotworthy* (unreported, T971545, Auckland District Court, 24 April 1998). Mr. Clotworthy was drunk when he stabbed his victim six times, an assault for which he could give no explanation. The victim wanted cosmetic surgery for the ugly scarring that he suffered. He was reported in court as saying "The deed is done and I would now like it undone." The judge suggested an RJ conference where the offender's apology was accepted and the offender agreed to pay for the surgery. The victim was definite that he did not want Mr. Clotworthy to go to prison, in part because he knew that the loan involved to pay the reparation could only be obtained if Mr. Clotworthy kept his job. The judge ordered the necessary reparation, substantial community service, and a two-year suspended sentence.

The court's judgment, however, was then appealed. The appeals court agreed that the offending was too serious for such a short sentence. It then required the offender to serve time in prison, even though the court agreed that Mr. Clotworthy was unlikely to offend again. His custodial sentence meant that the victim was unable to receive the funds necessary for the surgery he wanted, a factor that may have contributed to the victim's suicide several years later. It is unlikely that this appeal and prison sentence would have been successful after the introduction of the 2002 legislation.

While rigorous evaluations have been carried out for some of these RJ practices, more commonly RJ has been introduced as policy without testing its effects. In New Zealand, for example, RJ for youths became policy overnight with the introduction of the governing legislation in 1989, leaving no opportunity for assessing its consequences.

RJ Conferences Imposed by a Sentence

Figure 9.2 reports on one RCT that imposed an RJC conference as part of a probation sentence, subject to whether the probation officers randomly assigned the case to undertake such a conference. The experiment was limited to cases in which the victim was willing, and the offender was risk-assessed as being not unduly resistant to the idea. While the sample size was not large and the statistical power was not adequate for a statistically significant result, the RJC-assigned offenders had 55 percent fewer reconvictions than the offenders who were not assigned to RJC. This is a striking result that strongly merits attempts at replication.

RJ Conferences and Re-Entry after Prison

Figure 9.2 also reports on one RCT with prisoners sentenced for violent crimes who were within six months of scheduled release back to the community. The 94 cases in this experiment produced a result that showed 33 percent fewer reconvictions for the RJC group than for the group not assigned to RJC. All of these comparisons from figure 9.2 are at some risk of being based on preexisting differences in frequency rates of convictions, especially for the smaller samples. Random assignment is less reliable at controlling for preexisting differences with small samples than it is with larger samples, and a before-after comparison would have been better for eliminating the theory that the large difference was merely a preexisting condition rather than a result of RJC. Yet it is the best evidence available, and better evidence than from non-randomized designs.

RJ Conferences and Life or Death Sentences

A brief word is required on RJC for life sentences and execution cases. The benefits for victims, which we have not presented in detail here, can potentially justify substantial investments in homicide cases to assist the survivors of these crimes. In a number of cases, RJC has been conducted with the family of murder victims. Qualitative reports from these conferences are generally favorable, with victims reporting feelings of forgiveness that allow them to release their anger (Umbreit and Vos 2000). Whether these feelings can promote better physical health and longevity is a reasonable question for research. As long as the United States continues to execute 30 or more people per annum, this question remains relevant to public policy. Even in Europe, where the death penalty is considered a violation of human rights, there are many murder survivors who could benefit from RJC. Given the public health care systems in these countries, RJC may even be cost-effective at reducing medical costs from post-traumatic stress symptoms among families of murder victims.

IV. RJ as a Social Movement

The research reported above shows strong evidence that using an RJC is comparatively more effective than conventional justice without using an RJC, across a wide range of stages and decisions in sentencing and corrections. Yet evidence alone is not enough. The history of innovations suggests that widespread adoption of new practices is driven more by social movements than by the rational evidence for the changes the movements advocate. This conclusion emerges from the history of many major reforms, from the abolition of slavery (Hochschild 2005) and the creation of juvenile justice (Platt 1969) to the criminalization of alcohol consumption (Gusfield 1986) and increased punishment for drunken driving (Fell and Voas 2006).

Reforms that did not happen quickly, in contrast, have included those for which the technical evidence was very strong, but no social movement arose to promote them. The first U.S. state to require automobile users to wear seat belts, New York, passed its law in 1983. This law was enacted 82 years after the first mass production of automobiles, and 24 years after the invention of the shoulder belt by Volvo's first safety engineer, Nils Bohlin, during which countless people died preventable deaths in auto accidents. In another famous example, the value of limes and lemons to prevent scurvy on British navy ships was first demonstrated by a naval surgeon 42 years before the Navy finally agreed to supply its sailors with citrus fruits (Chalmers 2003).

By some accounts, restorative justice already has the kind of social movement that eluded lime juice and seat belts. The *Wikipedia* (2010) entry on restorative justice, for example, claims that restorative justice is a "a growing social movement to institutionalize peaceful approaches to harm." Yet there is little evidence that the use of RJCs in response to criminal charges has been growing in sentencing and corrections. By the standards of over 3 million members (Hamilton 2000) of the Mothers Against Drunk Driving (MADD)—about one in every 100 Americans—or the one million members of the Royal Society for the Protection of Birds (2010)—about one in every 60 Britons—there is no large membership organization to promote the use of RJC in sentencing and corrections in either the United States or the United Kingdom.

Yet the political conditions for advancing restorative justice remain fertile. Despite the lack of movement, the rhetoric in support of victims remains strong. As one Canadian reporter notes (McKnight 2010):

> . . . it's curious that RJ hasn't gone mainstream, since it aims to do exactly what politicians and others claim they want from the justice system. Rarely a day goes by that we don't hear politicians stressing the importance of meeting victims' needs or of ensuring offenders take responsibility, yet the system, focused as it is on punishment, has never been very good at doing either. Restorative justice, on the other hand, aims to make this rhetoric a reality.

The UK election of a coalition government in 2010 led immediately to similar statements of priorities on meeting victims' needs, but with a key difference. The new government explicitly promised to deliver more restorative justice. Whether and how that promise is kept is an important question for the future of RJ internationally. If it succeeds, it will not be without the active campaigning of civil society groups outside of political parties.

Civil Society Organizations for Restorative Justice

At least three civil society organizations, however, have become established to promote restorative justice conferences. The most global is the Prison Fellowship, which hosts a web site at www.restorativejustice.org. The home page displays the picture of a globe under the title "RJ Around the World." Under the globe are links to click for each continent, including "the Pacific." Each continent has further links to reports from specific countries. These reports suggest, once again, more intellectual interest in RJ than grassroots organizing. The Prison Fellowship itself is a Christian charity founded in the United States by Charles Colson, who served prison time on charges related to the 1972 Watergate burglary. It operates a wide array of programs, including rehabilitation services under contracts from governments. One such service is the Sycamore Tree Program, which does not feature face-to-face meetings between specific victims and their offenders, but does entail real crime victims coming into prisons to talk with groups of offenders. Its primacy in holding and operating the main Internet name for restorative justice (since 1996) suggests its early interest in RJ. It also may suggest the absence of a global victim-centric organization of comparable resources.

A national victim-centric organization was established in the United Kingdom in 2008 by crime victims (and at least one offender) who had participated in the Home Office–funded randomized trials that the authors led in 2002–2005. "Why Me?" is found at www.Why-Me.org, where its subtitle is identified as "Victims for Restorative Justice." The organization was launched in a bookstore at the launch of a book (Woolf 2008) by the offender who had attacked one of the founders, Will Riley—the royalties of which he is donating to the charity. The group has been supported by an older, more government-oriented UK charity, which had been founded in 1997 as the Restorative Justice Consortium (RJC). With the advent of "Why Me?" the RJC decided to emphasize the establishment of standards, training and education for professional practice in restorative justice generally, including what this chapter calls "RJCs" (restorative justice conferences).

In addition to what might be called the advocacy organizations, there are also training and consulting services provided to a wide range of clients in the United States, the United Kingdom, and Australia. Examples include Restorative Solutions (www.RestorativeSolutions.org.uk); the International Institute of Restorative Practices, based in Pennsylvania (www.iirp.org); and Proactive Resolutions (www.proactive-resolutions.com), the Australian company affiliated with the former New South Wales police trainer John McDonald, who trained all of the police officers who participated in the randomized trials of restorative justice reviewed above (figure 9.2). Each of these organizations teaches and leads RJ conferences, rather than organizing members. None can be described as a "corporate interest" attempting to lobby governments to increase the volume of RJ services paid for by tax dollars.

In many ways, these organizations can be said to suffer from an excess of altruism, and a lack of selfish passions. In contrast to medical charities or MADD, there is a distinct absence of trauma-motivated volunteerism. With the exception of "Why Me?" there is no visible victim associated with the social movement. The founder of MADD was motivated by drunk driving that caused the death of her daughter (Fell and Voas 2006). The March of Dimes was strongly supported by polio victim Franklin Roosevelt (Smith 1990). Yet RJ groups are largely staffed by former officials of national governments or local criminal justice agencies. Their leadership is committed but not visibly emotional. They neither comprise nor supply the kind of "heroic victim" that has been identified in campaigns to increase punishment (Zaffaroni 2009). Yet they could still succeed in creating a tipping point to expand RJ.

Elements of a Tipping Point

In speculating about the antecedents of a geometric increase in the rate of change he calls a "tipping point," Malcolm Gladwell (2006) identifies three key concepts: leadership, stickiness, and context. Without elaborating on these ideas here, let us simply observe that a Gandhi-like leader of great national fame could make a big difference, perhaps a family member of a victim of a tragic and highly publicized murder. The use of a "sticky" or memorable slogan would also empower such a movement, as long as it tapped deeply felt emotions.

Most important, however, may be the social context of a social movement. The context that could matter most might be the economy. The impact of the global recession on changing the context is evident in both the United States and the United Kingdom. In the United States, many states facing bankruptcy also bear the burden of an increasing prison population. In the United Kingdom, a plan to build more prisons for a rising prison population was abandoned after the 2010 election put a new government into power. In the latter context, a new criminal justice minister offered these comments in his first public address in that role (Herbert 2010): "There are two kinds of people: those who really get it about the importance of crime victims, and those who don't." As long as it is both financially prudent and morally right to help victims as RJC does, then the context may be ripe for a social movement to advance restorative justice.

V. Conclusion

This chapter has shown strong evidence that RJ conferencing is the now most evidence-based strategy in corrections, given a consistent method across a wide range of settings. That evidence is more than sufficient to justify implementing it widely and immediately, even while doing more research to answer further questions about it. The defense that "more research is needed" is not something we would support, especially after our 15 years of research on the topic. While more research would always be helpful, it is not "needed" to be sure enough that RJC will work in the same settings in which it has been tested and shown to be successful. At the very least, offering RJC to victims of burglary and robbery in English Crown Courts, the location of our London research, is about as evidence-based a decision as any sentencing policy has ever enjoyed.

What is hardest to communicate is the idea that RJC works best—and most cost-effectively—for the most serious crimes and criminals. Shapland et al. (2008) in their independent evaluation of our UK research showed that the greatest savings were for the criminals with the most serious records—those in Crown Court. Yet officials in the United Kingdom and elsewhere remain biased in seeing RJC as appropriate for "kid stuff," or trivial crimes. Several have told us that it is completely "inappropriate" for serious crime, regardless of the evidence. This view can be interpreted kindly as deontologism, for which no amount of evidence will be enough. Yet the same officials may contend that "more evidence is needed." Most recently, we can cite Woods's (2009) analysis of the Canberra results for all three juvenile RCTs, showing that RJC had the biggest effect on the offenders with the longest criminal careers and the highest frequency of convictions. How much more evidence is required?

Any reduction in prison populations, by any means, must insure incapacitation of the "worst" or most dangerous offenders. Some opposition to RJC may be based on a fear that it could be used instead of incapacitation in such cases. Developing risk assessment tools based on large sample sizes (Berk et al. 2009), using a crime harm index (Sherman 2011), can help to reduce the risk of using RJC with very dangerous people. But the risk has always been small, given the requirement that police conduct risk assessments before declaring offenders or defendants eligible for RJCs.

The most important point is that whether or not their offenders are imprisoned, victims can benefit from RJCs in ways that no conventional court process appears able to provide without an RJ conference. Direct benefits of RJC to victims can justify its possible mitigation of retribution and general deterrence, eliminating most legal theory against its use. Even easier is to limit RJC to use as a supplement to prison, rather than as an alternative. After years or decades of using RJC in addition to prison, it could be much easier to test the idea of using RJC as an alternative.

Many criminal justice reforms have been seized by activists and implemented without adequate testing. In this case we have just the opposite. Criminologists have tested RJC repeatedly for over 15 years. We do not believe that more research is needed. What we believe is that more action is needed.

NOTES

1. McCold and Wachtel's (1998) Bethlehem, PA, restorative police experiment was excluded from this review on the grounds that random assignment was conducted before victims and offenders were asked to consent to be randomly assigned, resulting in a very high level of non-compliance with random assignment. This non-compliance reduced the integrity of the experiment as a test of RJC to an unacceptably low level, which conforms with the expectation that post-random assignment consent will fail to provide an unbiased estimate of the effect of RJC when consent is requested in advance under non-experimental conditions.

REFERENCES

Ashworth, Andrew. 2002. "Responsibilities, Rights and Restorative Justice." *British Journal of Criminology* 42: 578–595.

BBC Radio 4. Today Show Interview. July 16, 2010.

Beccaria, Cesare. 1764 [1986]. *On Crimes and Punishments: Milan*, trans. David Young. Indianapolis, IN: Hackett Publishing.

Berk, Richard, Lawrence Sherman, Geoffrey Barnes, Ellen Kurtz, and Lindsay Ahlman. 2009. "Forecasting Murder Within a Population of Probationers and Parolees: A High Stakes Application of Statistical Learning." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 172: 191–211.

Braithwaite, John. 1989. *Crime, Shame and Reintegration*. Cambridge: Cambridge University Press.
Braithwaite, John. 1998. "Restorative Justice." In *The Handbook of Crime and Punishment*, ed. Michael Tonry, 323–344. New York: Oxford University Press.
Braithwaite, John. 2002. *Restorative Justice and Responsive Regulation*. New York: Oxford University Press.
Braithwaite, John, and Heather Strang. 2001. "Introduction: Restorative Justice and Civil Society." In *Restorative Justice and Civil Society*, eds. Heather Strang and John Braithwaite, 1–14. Cambridge: Cambridge University Press.
Chalmers, Iain. 2003. "The James Lind Initiative." *Journal of the Royal Society of Medicine* 96: 575–576.
Christie, Nils. 1977. "Conflicts as Property." *British Journal of Criminology* 17(1): 1–15.
Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates.
Collins, Randall. 2004. *Interaction Ritual Chains*. Princeton: Princeton University Press.
Crown v. Makatuu Folaumoeloa. 2004.
de Waal, Frans. 2009. *The Age of Empathy*. New York: Random House.
Diamond, Jared. 2008. "Vengeance Is Ours." *The New Yorker*, 21 April 2008.
Dignan, James. 2003. "Towards a Systemic Model of Restoratve Justice: Reflections on the Concept, Its Context and the Need for Clear Constaints." In *Restorative Justice and Criminal Justice*, eds. A. von Hirsch, A. Roberts and A Bottoms, 135–156, Oxford, Hart Publishing.
Durlauf, Steven F., and Daniel Nagin. 2011. "Imprisonment and Crime: Can Both Be Reduced?" *Criminology and Public Policy* 10: 13–62.
Durlauf, Steven F., and Daniel S. Nagin. 2011. "The Deterrent Effect of Imprisonment." In *Controlling Crime: Strategies and Tradeoffs*, eds. Philip J. Cook, Jens Ludwig, and Justin McCrary. Chicago: University of Chicago Press.
Fell, James C., and Robert Voas. 2006. "Mothers Against Drunk Driving (MADD): The First 25 Years." *Traffic Injury Prevention* 7: 195–212.
Foa, Edna, Barbara Rothbaum, and David Riggs. 1991. "Treatment of Post-Traumatic Stress Disorder in Rape Victims: A Comparison Between Cognitive-Behavioral Procedures and Counseling." *Journal of Consulting and Clinical Psychology* 59(5): 715–723.
Gladwell, Malcolm. 2006. *The Tipping Point*. Boston: Little, Brown.
Green, Ross. 1997. "Aboriginal Community Sentencing and Mediation: Within and Without the Circle." *Manitoba Law Review* 25: 77–126.
Gusfield, Joseph. 1963 [1986]. *Symbolic Crusade: Status Politics and the American Temperance Movement*. Chicago: University of Illinois Press.
Hamilton, Wendy J. 2000. "Mothers Against Drunk Driving—MADD in the USA." *Injury Prevention* 6: 90–91.
Haas, Jennifer, Celia P. Kaplan, Eric P. Gerstenberger, and Karla Kerlikowske. 2004. "Changes in the Use of Postmenopausal Hormone Therapy after the Publication of Clinical Trial Results." *Annals of Internal Medicine* 140: 184–188.
Herbert, Nick. 2010. Speech Presented to the Policy Exchange, 23 June 2010 (spoken, not published, version).
Hochschild, Adam. 2005. *Bury The Chains*. Boston: Houghton Mifflin.
Huxley, Elspeth. 1939 [1999]. *Red Strangers*. London: Penguin Classics.
Kant, Immanuel. 1785 [2005]. *Groundwork of the Metaphysics of Morals*. London, Routledge Classics.

Laub, John, and Robert Sampson. 2003. *Shared Beginnings, Divergent Lives: Delinquent Boys to Age 70*. Cambridge, MA: Harvard University Press.

Lehrer, Jonah. 2010. "The Truth Wears Off." *The New Yorker*, December 13.

McCullough, Michael. 2008. *Beyond Revenge: The Evolution of the Forgiveness Instinct*. San Francisco: Jossey-Bass.

McKnight, Peter. 2010. "Help for the Victims of Crime—and the Offenders." *Vancouver Sun*, July 10, 2010. Online. http://www.vancouversun.com/news/Help+victims+crime+offenders/3260197/story.html#ixzz0tOSMOTFK (accessed July 11, 2010).

Malvestuto, Robert. 2009. Personal communication with the Chief Probation Officer, Adult Probation and Parole Department, First Judicial District of Pennsylvania.

Mason, Anthony. 2001. "Restorative Justice: Courts and Civil Society." In *Restorative Justice: From Philosophy to Practice*, eds. Heather Strang and John Braithwaite. Aldershot, Hants: Ashgate.

Moore, Michael. 1987. "The Moral Worth of Retribution." In *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*, ed. Ferdinand Schoeman, 179–218. New York and Cambridge: Cambridge University Press.

Murphy, Jeffrie. 1987. "Does Kant Have a Theory of Punishment?" *Columbia Law Review* 87: 509–532.

Musson, Anthony. 2009. "Wergeld: Crime and the Compensation Culture in Medieval England." Gresham College Lecture at the Museum of London, 10 May. Available at: http://www.gresham.ac.uk/event.asp?PageId=45&EventId=951 (accessed January 17, 2011).

New Zealand Sentencing Act 2002, No. 9, S10

Peto, Richard, M. Pike, P. Armitage, et al. 1976. "Design and Analysis of Randomised Clinical Trials Requiring Prolonged Observation of Each Patient, Part 1—Introduction and Design." *British Journal of Cancer* 34: 585–612.

Platt, Anthony. 1969. *The Child Savers: The Invention of Delinquency*. Chicago: University of Chicago Press.

R. v. Clotworthy (unreported, T971545, Auckland District Court, 24 April 1998.

Royal Society for Protection of Birds, 2010. Online. Available at: www.rspb.org.uk/abouts/facts.asp (accessed June 27, 2010).

Ruth, Henry, and Kevin Reitz. 2003. *The Challenge of Crime: Rethinking Our Response*. Cambridge, MA: Harvard University Press.

Shapland, Joanna, Anne Atkinson, Helen Atkinson, James Dignan, Lucy Edwards, Jeremy Hibbert, Marie Howes, Jennifer Johnstone, Gwen Robinson, and Angela Sorsby. 2008. *Does Restorative Justice Affect Reconviction? The Fourth Report from the Evaluation of Three Schemes*. London: Ministry of Justice.

Schneider, Anne. 1986. "Restitution and Recidivism Rates of Juvenile Offenders: Results from Four Experimental Studies." *Criminology* 24(3): 533–552.

Sherman, Lawrence. 1992. *Policing Domestic Violence*: Experiments and Dilemmas. NY: Free Press.

Sherman, Lawrence. 2003. "Reason for Emotion: Reinventing Justice with Theories, Innovations and Research—the American Society of Criminology Presidential Address." *Criminology* 41(1): 1–38.

Sherman, Lawrence 2006. "Criminogenic vs. Preventive Effects of Restorative Justice by Offender Race: Findings and Next Steps." Paper presented at the annual meeting of the American Society of Criminology, November 2006.

Sherman, Lawrence. 2011. "Al Capone, the Sword of Damocles, and the Police—Corrections Budget Ratio: Afterword to the Special Issue" *Criminology and Public Policy* 10: 195–202.

Sherman, Lawrence, Denise Gottfredson, Doris MacKenzie, John Eck, Peter Reuter, and Shawn D. Bushway. 1997. *Preventing Crime: What Works, What Doesn't, What's Promising*. Washington, DC: US Department of Justice, Office of Justice Programs.

Sherman, Lawrence W., Heather Strang, Caroline Angel, Daniel Woods, Meredith Rossner, Geoffrey C. Barnes, Sarah Bennett, and Nova Inkpen. 2005. "Effects of Face-to-Face Restorative Justice on Victims of Crime in Four Randomized, Controlled Trials." *Journal of Experimental Criminology* 1(3): 367–395.

Sherman, Lawrence, and Heather Strang. 2007. *Restorative Justice: The Evidence*. London: Smith Institute. Available at: www.smith-institute.org.uk/publications.htm.

Sherman, Lawrence, and Heather Strang. 2011. "Empathy for the Devil." In *Emotions, Crime and Justice*, eds. Susanne Karstedt, Ian Loader and Heather Strang. Oxford: Hart Publishing.

Smith, Jane. 1990. *Patenting the Sun: Polio and the Salk Vaccine*. New York: Anchor Books.

Strang, Heather. 2002. *Repair or Revenge: Victims and Restorative Justice*. Oxford: Clarendon Press.

Strang, Heather, and Lawrence Sherman. Forthcoming. *Systematic Review of Effects of Restorative Justice Conferencing on Victims and Repeat Offending*. Campbell Collaboration.

Tyler, Tom. 1990. *Why People Obey the Law*. New Haven: Yale University Press.

Umbreit, Mark, and Betty Vos. 2000. " Homicide Survivors Meet the Offender Prior to Execution: Restorative Justice Through Dialogue." *Homicide Studies* 4(1): 63–87.

von Hirsch, Andrew. 1976. *Doing Justice: The Choice of Punishments*. New York: Farrar, Strauss and Giroux.

von Hirsch, Andrew, and Andrew Ashworth. 2005. *Proportionate Sentencing: Exploring the Principles*. New York and Oxford: Oxford University Press.

Weisburd, David. 1993. "Design Sensitivity in Criminal Justice Experiments." *Crime and Justice* 17: 337–379.

Weisburd, David, Cynthia Lum, and Sue-Ming Yang. 2003. "When Can We Conclude That Treatments or Programs 'Don't Work'?" *The ANNALS of the American Academy of Political and Social Science* 587: 31–48,

Woods, Daniel. 2009. "Unpacking the Impact of Restorative Justice in the RISE Experiments: Facilitators, Offenders, and Conference Non-Delivery." Ph.D. dissertation, University of Pennsylvania.

Woolf, Peter. 2008. *The Damage Done*. London: Bantam Press.

Zaffaroni, Raul. 2009. "Can Criminal Law Really Contribute to the Prevention of Crimes Against Humanity?" *Journal of Scandinavian Studies in Criminology and Crime Prevention* 10: 2–25.

Zimring, Frank, and Gordon Hawkins. 1973. *Deterrence: The Legal Threat in Crime Control*. Chicago: University of Chicago Press.