

Effects of face-to-face restorative justice on victims of crime in four randomized, controlled trials

LAWRENCE W. SHERMAN, HEATHER STRANG*, CAROLINE ANGEL,
DANIEL WOODS, GEOFFREY C. BARNES, SARAH BENNETT and
NOVA INKPEN

Jerry Lee Program on Randomized Controlled Trials in Restorative Justice, Jerry Center of Criminology, University of Pennsylvania, Philadelphia, Pennsylvania, USA and Centre for Restorative Justice, RegNet, Australian National University, Canberra, Australia

**corresponding author: E-mail: heather.strang@anu.edu.au*

Abstract. The growing use of restorative justice provides a major opportunity for experimental criminology and evidence-based policy. Face-to-face meetings led by police officers between crime victims and their offenders are predicted to reduce the harm to victims caused by the crime. This prediction is derived not only from the social movement for restorative justice, but also from psychological and sociological theories. Four randomized, controlled trials of this hypothesis in London and Canberra, with point estimates disaggregated by gender, tested the prediction with measures of both successful interaction (apologies received and their perceived sincerity) and the hypothesized benefits of the ritual (on forgiveness of, and reduced desire for violent revenge against, offenders, and victim self-blame for the crime). The meta-analyses of the eight point estimates suggest success (as victims define it) of restorative justice as an interaction ritual, and some benefits as a policy for reducing harm to victims.

Across the world, a growing social movement advocates restorative justice for those affected by crime—victims and offenders alike (Braithwaite 2002). Embracing a wide range of procedures recommended for a wide range of criminal justice settings, the concept of restorative justice is associated with two major hypotheses, both of which are eminently testable (Ruth and Reitz 2003). One is that restorative justice (RJ) will do better than conventional justice (CJ) at reducing repeat offending. The other hypothesis is that RJ will do better than CJ at repairing the harm that crime causes to victims. This article provides new evidence on the second hypothesis, drawn exclusively from randomized controlled trials (RCTs) led by the two senior authors.

Our evidence bears upon one specific approach to restorative justice: face-to-face meetings among offenders, their victims, and their respective family and friends as ‘supporters.’ All of our evidence is based on such meetings being conducted by specially trained police officers who have studied the facts of the case and arranged for all of the participants to attend the meeting. The four RCTs vary by offense type (with both violent and property crimes) and location in the criminal justice process (pre-court diversion vs. pre-sentence), as well as by the physical location of the meetings (inside prisons or in private rooms in police stations) and by nation (Australia vs. United Kingdom). Thus we examine the

further hypothesis that the effects of the meetings on victims will remain consistent across these differences in social context and legal setting, given a consistent approach to the meetings themselves. We also test the hypothesis that the effects vary little by gender of the victims, with mostly male offenders committing crimes against both male and female victims.

These hypotheses are tested with randomly assigned RJ events structured on a standard format, all of which were led by police officers in London (UK) and Canberra (Australia) trained by the same trainers. The format consists of seating all participants in a circle, then allowing (and encouraging) everyone present to express views and ask questions about three focal issues: 1) what happened in the course of the crime that led to the RJ conference? 2) who was affected by what happened, and how were they affected? 3) what should happen now to repair the harm that the crime has caused? In all RJ cases, the police asked the offender to sign, in front of the assembled group, an 'outcome agreement' promising to complete the actions agreed to in the RJ conference. While control group conditions varied by specific trial, all control groups differed from the RJ group by their consistent lack of any supervised face-to-face interaction between victims and offenders.

Theoretical background

While the RJ hypothesis on repeat offending is derived from such criminological theories as reintegrative shaming (Braithwaite 1989), procedural justice (Tyler 1990) and defiance theory (Sherman 1993), the RJ hypothesis on victim benefits has been almost a theoretical afterthought. Criminology has been far more concerned with the effects of sanction content and style on future crime than with a pluralistic conception of 'justice' in the eyes of multiple parties. While, for example, the procedural justice framework applies to crime victims as well as offenders (Strang 2002), in practice victims have been largely ignored in criminological theory.

Nonetheless, two prominent theories outside of criminology both predict that victims should benefit from face-to-face meetings with their offenders. In psychology, a well-developed literature on Cognitive Behavioral Therapy (CBT) suggests that victims can reduce the harm they have suffered by 'normalizing' the context and players in the stressful event of the crime, and reduce the common tendency to blame themselves for having somehow caused the crime. In sociology, a recently-elaborated theory of 'interaction ritual' predicts that crime victims will have their emotional energy and their identity restored, and their commitment to group values reaffirmed, by a successful RJ meeting.

Cognitive behavioral therapy

Although the concept of emotional harm resulting from victim trauma has been known for over a century, post-traumatic stress disorder (PTSD) has only

recently been identified as a recognized medical condition (Foa and Meadows 1997). Reviews of rigorous studies conducted over the past two decades indicate that CBT is one of the most promising interventions in the treatment of PTSD. A particularly effective form of CBT in reducing symptoms of PTSD appears to be therapy that consists of purposively prolonged exposure of victims (in safe settings) to anxiety-provoking stimuli. This typically involves helping victims to face their fears through either repeated reliving of the trauma or planned confrontations with people or things that the victim associates with the traumatic event (Rothbaum and Foa 1999). For example, Foa and her colleagues ask rape victims to record an audiotaped description of the rape in their own words, and then listen to their own voice repeatedly in safe settings. After weeks of this treatment, supplemented by conversations with therapists about the specific elements of the offense (location, time of day, weather, appearance of the rapist), the fear evoked by encountering those elements tends to dissipate. Thus a rape victim can see a young male, or walk into her own bedroom, without recalling the fear associated with the crime.

Foa's CBT treatment program is based on conditioning theory about the way that fear is acquired and extinguished. Foa and Kozak (1986) theorized that fear is a cognitive structure and suggested that exposure to fear is an opportunity for extinguishing it through de-conditioning the memories associated with the traumatic event. Their hypothesis is that the presentation of new, low-risk or 'safe' information about the associations with the event can alter – and dissipate – the emotional reaction to the victim's recollection of the trauma. By associating the details of the offense with a 'safe' setting, such as a therapist's office, the details of the crime themselves become neutralized, and only the actual crime remains a threat.

This leads to a second element of CBT that converges with RJ conferences: understanding the causation of the event. Both CBT and RJ address the common concern of crime victims with the thought that the victims somehow caused the crime through their own negligence or incompetence. In CBT, the therapist talks the victim through an analysis of the events leading up to the crime, repeatedly demonstrating how the victim did nothing 'wrong' or 'stupid.' This reassurance appears in Foa's clinical trial evidence to be an important part of reducing the level of post-traumatic stress symptoms (PTSS). Similarly, in face-to-face RJ conferences, one of the first questions crime victims put to the offenders is "why me?" Victims almost always seem reassured when the offenders say they did not 'target' the victim for any particular reason, when the crime occurred as an almost random intersection of offender and victim in time and space.

A comparison of the CBT literature to the RJ process in our experiments suggests that a face-to-face RJ conference has all the elements that are necessary for de-conditioning the fearful associations in victims' recalling of the event. By bringing the single most fearsome association with a crime – the offender – into direct contact with the victim, an RJ conference provides a

safe setting for de-conditioning even fear of the offender himself. This, in turn, should theoretically promote fear reduction and hence reduction of PTSS. Meeting the offender in a safe place reminds them that associations with the event are not dangerous *per se*; that remembering the events is different from experiencing them again; that feeling anxious does not necessarily lead to loss of control (Rothbaum and Foa 1999). The natural anxiety involved in confronting one's offender can be discharged safely in the police station or prison location of the RJ conference as a safe setting, reducing the PTSS that victims of serious crime experience.

Quasi-experimental evidence so far supports the PTSS reduction hypothesis, such as Strang's (2002: 99) finding that Canberra victims reported less than half as much fear of the offender after an RJ conference as they had experienced before the conference (single retrospective measure only). A stronger test of the hypothesis is also underway in relation to our two London RCTs (Angel 2004), but is not yet available for purposes of the present analysis.

The present analysis focuses on blame, in two ways: self-blame and desire for violent revenge against the offender. Self-blame is the second major component of Foa's CBT model: whether RJ can reduce victims' tendency to blame themselves for the crime. Whether or not victims have a factual basis for blaming themselves, of course, may vary by the specific characteristics of the crime. The victim of an assault in a pub fight, for example, may in fact have contributed to the causation of the assault by using provocative speech or gestures. In most of the cases in the present analysis, however, the crime was not preceded by any interaction between victim and offender. Rather, the crimes of robbery, burglary, auto theft and other personal property crime are predatory in character, with offenders striking at victims without warning. This was equally true for control and experimental groups. Thus we may fairly ask whether RJ reduced victim self-blame for the crime.

The self-blame measure, to be sure, is not an idea indicator of whether victims have had their identity as 'competent' people restored by the RJ ritual. It is unclear whether victims blame themselves for something they did or something about themselves as a person that caused the offense (characterological self-blame). It is this second type of self-blaming "I am an incompetent person – I deserved the crime" that is most relevant to test Foa's theory, of which we were unaware when designing the four experiments. Some might not expect a conference to modify self-blame in terms of things a person did. A person might realize that she was not the target of the offense through talking to the offender (which may make her feel less vulnerable), but nevertheless will still blame herself for leaving the door unlocked. We present this measure with full awareness of their limitations in order to spark further research and discussion.

The self-blame test of RJ can be supplemented by a test for the level of anger at the offender. A prediction for that test, however, is more appropriately based upon our second theoretical framework.

Interaction rituals

Collins (2004: 111) conceptualizes RJ conferences as an example of a larger category of social events that he (after Durkheim and Goffman) formally defines as ‘interaction rituals.’ Collins defines interaction ritual (IR) as social encounters with four distinct features: 1) people are physically together so that they are influenced by each others’ bodily presence; 2) the boundaries of interaction membership are clearly defined so that everyone knows who is participating and who is not; 3) participants focus on a common purpose and know that all are focusing on it; and 4) participants share a common, if dynamic, emotional mood or ‘entrainment’ experience. Feedback between these elements via ‘mutual entrainment of emotion and attention’ (Collins 2004: 48) produces a shared experience at both the emotional and cognitive level. The Durkheimian concept of the feeling of ‘collective effervescence’ that ensues from a successful – but not from a failed – IR gives participants a sense of social solidarity with each other in the moment. It then causes longer-term effects in emotional energy that include feelings of confidence, elation and initiative in action, and *recommitment to the group’s standards of morality*.

The empirically falsifiable claim about any IR event is that it has succeeded. Collins (2004: 51) specifies the criteria for IR success or failure as follows, referring to both the ingredients and the outcomes of rituals as indicators of a continuum of the degree to which IR may fail:

“... a low level of collective effervescence... no shared entrainment... [and] on the output side: little or no feelings of group solidarity; no sense of one’s identity as affirmed or changed... no heightened emotional energy – either a flat feeling unaffected by the ritual, or worse yet, a sense of a drag... These imply a continuum of just how badly rituals fail...”

Collins specifically cites RJ conferences to illustrate his theory that both negative and positive emotions can drive an IR event towards success. His claim is that it is the intensity of emotions, rather than their content, that increases the chances of IR success. Anger, friendliness, fear or many other emotions may dominate the initial mood; emotional contagion may then take place nonetheless because everyone present is focusing intensely on the same feeling. This contagion can allow ritual to transform emotions from bad to good, binding together the participants in their commitment to shared morality and righteous anger at violations. Since RJ conferences may thus enlist the offenders in condemning their own conduct, they open the door to the “hate-the-sin-but-love-the-sinner” re-integration model suggested by Braithwaite (1989).

Collins suggests that RJ conferences are likely to be successful IR events because they are likely to feature high intensity of emotions, and are structured to include all the ingredients the theory posits. This should include the major symbol of group solidarity in RJ events, which is an apology by the offender to the victim.

Whether this symbol is successfully achieved, however, may be further indicated by the victim's post-IR assessment of whether the apology was sincere.

The claim that successful IR also reaffirms identity may suggest that crime victims should come away from RJ events more likely to see themselves as competent people, and less likely to blame themselves for somehow having caused the crime. While this linkage is not as clear from Collins' framework as the central role of a sincere apology, it is a plausible link between the CBT framework in clinical psychology and the microsociology of IR.

The recommitment to group morality that Collins hypothesizes should also predict less victim desire for violent revenge against the offender. The offender's violation of group morality is no moral excuse for the victim to do the same ("two wrongs do not make a right"). If the victim responds to the interaction ritual by suppressing a desire to break norms, this should indicate a successful outcome effect as Collins' theory predicts.

Hypotheses and criterion measures

The present analysis is intended as only a partial examination of the two main hypotheses derived from the foregoing theoretical analysis. One hypothesis concerns the interaction ritual *ingredients* of RJ conferences. The other concerns the *consequences* of those ingredients in RJ for crime victims' subsequent emotions in the weeks and months after the RJ event. These main hypotheses are as follows:

- H1: RJ conferences are more likely than conventional justice processes without RJ to produce a successful interaction ritual fostering social solidarity among offenders and victims that reaffirms their mutual group morality.
- H2: Victims randomly assigned to RJ conferences are more likely to show future psychological benefits from the justice process than victims not assigned to RJ.

A wide range of criterion measures can, in theory, be employed to test these hypotheses. The Canberra experiments, for example, featured systematic observations of RJ conferences as well as CJ processes, as well as interviews of both offenders and victims after the events (Harris 2002; Strang 2002). The authors were not funded to undertake systematic observations in the London RCTs, but an independent evaluation by the University of Sheffield gathered such data and is expected to report on them in future years (see Shapland et al. 2004). Conversely, the London experiments include victim measures of post-traumatic stress symptoms employing the same measures as Foa and her colleagues (Angel 2004), but such data were not collected in Canberra.

The present analysis focuses primarily on criterion measures for the two main hypotheses that can yield multiple estimates across offense types and victim genders. This approach, from which we can derive eight point estimates on most

criteria, has the advantage of increasing the reliability and external validity of the conclusions—in contrast to reporting each RCT or subgroup estimate one at a time. By providing the full CONSORT statement (Moher et al. 2001) description of each of the four RCTs for which we report these common criterion measures, the analysis simultaneously accomplishes elements of both primary and meta-analysis. The limitation of this approach is that it confines the analysis to a small selection of all possible measures unique to each site. Thus the analysis should be seen as preliminary, to be informed by further results presented in a similar format for other criterion measures of the two main hypotheses.

Two criterion measures are available for testing H1. One is based on the highly symbolic value of an apology (Strang 2002: 20–23) to victims, representing an important symbol of group solidarity and shared morality (Collins 2004: 48). The criterion here is whether RJ conferences result in more apologies. Given Collins' (2004: 53) discussion of 'forced ritual' as potentially unsuccessful IR, we add a second criterion as a check on the first: whether the victim perceived an offender's apology to have been sincere.

Two criterion measures are available for testing H2 across all eight point estimates: whether the victims blame themselves for having caused the crime, and whether they desire violent revenge for the crime against their offenders. The first measure directly taps one of the key predictors of clinical success in reducing PTSS among rape victims, and may reflect a victim's sense of identity as a competent person. A related measure of the degree to which the victim forgives the offender is not available in all four RCTs, but does help to illuminate the psychological consequences, if any, of RJ conferences. The analysis can focus in particular on the relationship between self-blame and forgiveness, the latter of which has been found in other research to predict better mental and physical health.

Victims' desire for violent revenge is an indicator of harm done by the crime (or potentially cured by RJ), given the enormous risks of actually carrying out a revenge crime. Vengeance may cause a large share of all crime (Black 1982), putting avengers at risk of arrest and imprisonment, as well as potential retaliation by the original offenders or their friends. Reducing the stated victim desire for such revenge – to the (unknown) extent that it actually predicts violent behavior – is both a desirable policy benefit and a test of Collins' predicted recommitment by victims to shared group morality.

Research designs

This article reports on the results to date from two completed randomized controlled trials (RCTs) and two that are still in progress. The two completed are the violence and personal property RCTs in the Reintegrative Shaming Experiments (RISE) conducted in Canberra, in which RJ was organized as a diversion from prosecution and resulted in no criminal record (see generally Strang 2002). The two in progress are the robbery and burglary RCTs reaching completion in

London, in which RJ occurs after a guilty plea but before sentencing in Crown Court (Sherman and Strang 2004a, b).

This Section presents core data for these RCTs in compliance with the requirements of the CONSORT Statement (Moher et al. 2001). This statement, the *Consolidated Standards of Reporting Trials*, has been adopted by such leading publishers of experimental work as the *Journal of the American Medical Association*, *The Lancet*, and the National Library of Medicine. Its checklist of items that every randomized trial should report is available at www.consort-statement.org, and is the basis for the selection of items reported in Table 1: the relevant facts on each CONSORT item for each of the four RCTs.

These standards have been developed in reaction to the large numbers of randomized trials that have been published without adequate descriptions of how the research was conducted. While RCTs are often described as a 'gold standard,' that gold may quickly turn to dross from any number of threats to internal validity.

The following discussion elaborates on the content of Table 1. While Table 1 presents the data for the overall RCTs in London as of 31 October 2004, the hypothesis tests in London are based on varying subsets of the full RCT for which victim interviews had been completed as of November. These subsets are described in the discussion below.

Settings: Pre-Court Diversion in Canberra, Pre-Sentence Meetings in London

The first two RCTs were conducted in Canberra, Australia, from July of 1995 through June of 2000, as part of the Reintegrative Shaming Experiments (RISE), and were conducted in cooperation with the Australian Federal Police. The Juvenile Personal Property crime RCT (JPP) and the Juvenile Violent Crime (JVC) RCT both operated in the same way. Arrestees for eligible offenses who admitted their crimes were randomly assigned either to go to court in the usual way or to be diverted to a face-to-face restorative justice conference with the victim of their offense, as well as family and friends of both the victims and the offenders in each case.

The second two RCTs began in London in July of 2002, in cooperation with the London Metropolitan Police: the London Burglary (LOB) RCT and the London Robbery (LOR) RCT. The offenses in these experiments were serious enough to be dealt with in the Crown Court, and most of the offenders recruited for the trials had been remanded into prison prior to sentencing. Cases in which offenders pleaded guilty to their offenses in court were randomly assigned either to proceed to sentence in the usual way or to participate in a face-to-face restorative justice conference prior to sentence. In the latter case, the sentencing judge could take the RJ conference and outcome agreement into account as potential 'mitigation' for the offense, which could reduce sentence length. Despite case law created during the RCTs (in 2003) indicating that such reductions would be appropriate even for offenders merely volunteering for RJ conferences (regardless of whether one actually occurred), our findings to date suggest no statistically significant

Table 1. CONSORT elements up to results for four randomized controlled trials of RJ vs. CJ.

<i>CONSORT element</i>	<i>Canberra violence (VIC)</i>	<i>Canberra property (JPP)</i>	<i>London robbery (LOR)</i>	<i>London burglary (LOB)</i>
Eligibility criteria	Offenders <30, offenses of medium seriousness, normally prosecuted in court	Offenders <18; offenses of medium seriousness, normally prosecuted in court	Adult offenders, serious offenses prosecuted in Crown Court	Adult offenders; serious offenses prosecuted in Crown Court
Interventions	Diversion from prosecution to an RJ meeting and agreements; control group was CJ prosecution	Diversion from prosecution to an RJ meeting and agreements; control group was CJ prosecution	Post-guilty plea but pre-sentence RJ meetings; control group had no meetings	Post-guilty plea but pre-sentence RJ meetings; control group had no meetings
Hypotheses	RJ victims would be better off in short-term than CJ victims	Same	Same	Same
Outcome measures	1. Apologies to victims 2. Sincerity of apology 3. Self-blame for crime	Same	Same	Same
Sample size	Power-based target = 300 cases; Actual $N = 100$ cases	Power-based target = 300 cases; Actual $N = 173$ cases	Power-based target = 100 cases; current interim report $N = 125$ cases	Power-based target = 100 cases; current interim report $N = 216$ cases
RA formulae	See text	See text	See text	See text
Allocation concealment method	Police obtained offender consent with allocation concealed; victims consented after RA	Police obtained offender consent with allocation concealed; victims consented after RA	Police obtained offender and victim consent to RA with allocation concealed	Police obtained offender and victim consent to RA with allocation concealed
Randomization implementation	Police called local 24-h research hotline; sealed envelopes opened by research staff	Police called local 24-h research hotline; sealed envelopes opened by research staff	Police called US 11-h hotline; programmed computer-generated allocation by research staff	Police called US 11-h hotline; programmed computer-generated allocation by research staff
Blinding	No blinding	No blinding	No blinding	No blinding
Statistical methods	Intention-to-treat, overall and by gender	Same	Same except no interview when victim dropped out of RJ	Same except no interview when victim dropped out of RJ

difference in sentence length between the randomly assigned RJ and control groups. All sentences, were decided after the completion of the RJ conference, if any.

Participant eligibility and characteristics

The two Canberra experiments included a broad range of middle-level offenses committed by young offenders. The cases all involved offenses serious enough to be dealt with in the Magistrates' Court. Domestic violence and sexual offenses were excluded from the study. Eligible offenders who made full admissions to the police were offered the opportunity to participate in RJ and were told that if they were willing to meet with their victims and accept responsibility for their crime they might not be prosecuted. No offender in these two experiments declined to participate. Uniformed, operational police performed the initial screening for eligibility and then called a 24 h research hotline. Eligibility was then confirmed by a member of the research staff before opening a sealed envelope to reveal the random assignment, which was relayed to the police officer who processed the offenders accordingly. Random assignment occurred prior to obtaining victim consent, largely due to police need to know immediately how to proceed at the time of offender arrest and interview. Very few victims declined to participate. In a small number of cases victims could not be contacted, while in others there was no identifiable victim (see Strang 2002: 74–81).

Offenses in the property experiment (JPP) included theft, burglary, car theft, shoplifting in owner-operated shops (not big stores), criminal damage, fraud, vehicle break-in and receiving or possession of stolen goods. The experiment was restricted to juvenile offenders (mean age 15.6 years); of the 249 offenders in the 173 cases, 16% were female and 9% Aboriginal. The 151 property crime victims who were interviewed (88% of all who could be approached) were 38 years of age on average; 41% were female and 2% were Aboriginal.

Offenses in the violence experiment (JVC) included mostly assault occasioning actual bodily harm, common assault, fighting and arson. All offenders were under age 30 (mean age 17.8 years). Of the 121 violent offenders in 100 cases, 18% were female and 12% Aboriginal. The 81 victims of violence who were interviewed (91% of all who could be approached) were 25 years of age on average; 32% were female and 6% were Aboriginal.

The two London experiments were limited to adults who pled guilty to certain categories of two types of offenses, burglary and robbery, serious enough to be dealt with in Crown Court. All such cases in which judges ordered the Probation Service to prepare a pre-sentence report were identified by our research team on a daily to weekly basis on the National Probation Service – London's computerized 'Probation Tracker.' The period of adjournment before sentencing in order to prepare the report provided a window of about 4 weeks during which time restorative justice could be offered and completed. Offenders, usually in prison, were approached by Scotland Yard police officers trained as restorative justice facilitators who had checked the cases for eligibility and 'risk assessment' (in

relation to potential violence in a conference, which eliminated less than ten cases). The primary screening was to insure that there had been a guilty plea and not a trial, since the RJ treatment required offender acceptance of responsibility prior to meeting with the victim. This was re-checked when police met with the offenders to discuss the case. If the offenders confirmed their acceptance of responsibility for having caused the crime and were otherwise eligible, the police officers asked the offenders if they wished to participate in the RCT. The offenders were told that if they agreed, and if their victims agreed to participate as well, the case would have a 50% chance of each of two alternatives: proceeding to sentence in the usual way, or participating in a face-to-face meeting with their victim prior to sentence.

In the burglary experiment, 78% of offenders who were invited consented to participate. The victims of consenting offenders were then asked to consent themselves; 59% of those victims consented, for a total of 46% of the eligible cases adjourned for pre-sentence reports after guilty pleas in the Crown Courts. These London Courts included those participating in the RCT at any point in time (the *N* of courts gradually expanded from the original 2 to reach all 12 in London by mid-2003). In the robbery experiment 82% of offenders police approached gave their consent, as did 52% of the victims of those offenders, for a total of 43% of potentially eligible cases. As soon as the facilitator obtained the consent of the victim he or she called the random assignment hotline number (staffed in the US) immediately, so that the victims knew right away whether they would be participating in a conference. The offender was advised as soon as possible thereafter.

The burglary experiment (LOB) included both burglary and aggravated burglary. The latter is defined as the offender either carrying a weapon at the time of the offense or the offender threatening to assault or actually assaulting the victim in the course of the burglary. In the cases used for the present analyses, 27 of 145 (19%) of all burglary victims had seen their offender at the time of the incident. Offenses in this experiment included both dwelling and non-dwelling burglaries, but mostly the former. As of 31 October 2004, 216 cases had been enrolled in the experiment involving 216 offenders. Only cases with single offenders were eligible for LOB and LOR. (In contrast, the Australian JVC and JPP trials took all co-offenders who were arrested in each case and treated them simultaneously in a single RJ conference if the case was randomly assigned to RJ). The average age of LOB offenders in the RCT was 31 years; 9% were female and 37% were non-white. There were 268 victims involved in these cases: their average age was 36 years; 50% were female and 13% were non-white.

The robbery experiment (LOR) included both robbery and 'street crime' (mainly handbag-snatching without force), with and without weapons. At 31 October 2004 there were 125 cases enrolled in this experiment involving 125 offenders (again, co-defendant cases were excluded). Their average age was 28 years; 6% were female and 57% were non-white. There were 138 victims involved in these cases: their average age was 31 years; 51% were female and 29% were non-white.

For the subset of victims who were interviewed for the present analyses, additional criteria were established. Victims were not approached for an interview

unless they were (a) adults who could (b) read, speak, and understand English. Participants were excluded from the interviews if they were (c) unable to be located following random assignment or participation in a conference, (d) withdrew from the RCT prior to attending a conference or (e) their case was subsequently cancelled or disqualified after initial screening. While we attempted to create an Intention-To-Treat (ITT) approach to this analysis, ethical considerations led police to prefer that we not contact the less than 5% of those who had withdrawn after random assignment – who were typically quite distraught.

For administrative reasons in the management of the RCTs, the interviews did not begin until six months after the start of the RCTs. Hence a final eligibility criterion for the present cases is that the cases were all recruited from January 2003 through July of 2004 (185 LOB cases and 115 LOR cases). The interviews reported here constitute 88% of the burglary victims randomly assigned in that 18-month period who were approached for interviews under criteria “a” through “e” listed above (146 interviewed out of 166 approached). They also constitute 77% of the robbery victims randomly assigned in that 18-month period who were approached for interviews (69 interviewed out of 90 approached). The percent of interviewed victims in the burglary experiment (LOB) who were female was 56% in the RJ (39 of 70) and 61% in the CJ group (4 of 76); in the robbery experiment (LOR) the percent of interviewed victims who were female was 44% in the RJ group (5 of 34) and 57% in the CJ group (20 of 35). (Given our use of gender-specific estimates in the analysis, the imbalance of gender in the LOR sample is not a confound.) The only other demographic characteristic gathered in the interview was age, which was also quite similar across treatments. The percent of burglary victims interviewed who were older than 30 years of age was 81% in the RJ group (57 of 70) and 78% in the control group (58 of 74). The percent of robbery victims interviewed who were older than 30 years of age was 59% for the RJ group (20 of 34 who responded) and 50% for the control group (17 of 34 responses).

The mean number of days from random assignment to interview for the subset of London robbery cases reported below was 28 for victims assigned to RJ and 24 for victims assigned to conventional justice. The mean number of days from random assignment to interview for the subset of London burglary cases reported below was 29 for victims assigned to RJ and 24 for victims assigned to CJ. The mean number of days between the RJ conference and the interview was 14 for robbery victims and 16 for burglary victims. All victims in Figures 1–4 are reported as independent subjects, although random assignment was done at the level of the case. The mean number of victims per case for the robbery subset was 1.12; the mean number of victims per case for the burglary subset was 1.24.

Interventions

The treatment tested in these RCTs varied only slightly. In all four of these experiments the facilitators who conducted the face-to-face meetings were police

officers, all of whom had received the same four-day training course from the same team of Australian trainers. Training consisted of both restorative justice theory and role-play practice at conducting the sessions. A great deal of emphasis was placed on adequate preparation of both offenders and victims prior to the meetings – although this was given more attention in the London experiments, where facilitators were able to build on the experience gained in the Canberra RISE project. Whereas in RISE facilitators sometimes only talked briefly to participants by phone, in London they met every offender and victim at least once prior to the conference, and sometimes more often; they frequently supplemented these meetings with multiple telephone conversations. In all the experiments victims and offenders were urged to bring friends and family to the conference.

In London, but not in Canberra, facilitators frequently spoke to these supporters to explain their roles and responsibilities. A good deal of preparation work involved transport arrangements, particularly in London with its transportation difficulties, to ensure that everyone was able to attend. The result was that, in general, the London conferences had fewer, but better-prepared participants than RJ meetings in Canberra; Canberra had more participants, especially offender supporters, but they were not as well prepared as in London.

The locations of the RJ conferences were almost always secured by the presence of other criminal justice officials besides the police officer leading the conference. In Canberra (JPP and JVC) the RJ conferences all took place in police stations; the London conferences (LOB and LOR) mostly took place within prisons, because the majority of offenders were remanded in custody (and most were later given custodial sentences).

The most important factor from an experimental standpoint is that the format, timing, and interaction structure of the conferences were highly consistent. In all four experiments the RJ conferences averaged 1.5 to 2 h duration, during which time participants sat in a circle and took turns speaking. The topics for discussion were 1) what had happened at the time of the offense, 2) what the consequences had been for everyone affected – victims, offenders and their supporters – and 3) what should happen now to repair the harm caused. Conferences were often extremely emotional encounters, especially where the offense involved violence. Facilitators were trained to manage these events in a way that permitted the expression of anger while retaining enough civility to allow discussion to move towards some kind of resolution. This took the form of an ‘outcome agreement’ reached at the end of the conference, to which all participants contributed. The agreement usually consisted of promises by the offender to undertake either reparations to the victim or community, or rehabilitation of themselves, or both. Agreements included financial restitution to the victim, community service (if the offender was not given a prison sentence), participation in drug treatment, literacy education or other programs designed to reduce the risk of reoffending. Compliance with outcome agreements was monitored by the facilitators; our current estimate is that across all four experiments offenders completed around 75% of their undertakings.

In the case of the Canberra experiments, offenders assigned to the control group were dealt with in the Magistrates' courts: those aged under 18 went to a special Magistrates' Children's Court, a closed court where only members of offenders' families could observe (although special permission was given to our research team, conditional on a veto by offenders that was exercised only twice, to attend as well and collect observational data on the way the case was dealt with). The majority of offenders were sentenced to post 'bonds' to be 'of good behavior' for a period of 12 months. In addition, some received rehabilitative orders and a few were fined; not one received a custodial sentence (incarceration) for the presenting offense that brought them into the experiment.

In the London experiments, offenders in both groups proceeded to sentence after their guilty plea and their pre-sentence report. The outcome agreements of offenders in the experimental group were placed before the sentencing judge to acquaint him or her with the undertakings made in the conference. In LOB only, from case 111 onwards (effective August, 2003), a two to three page account of the content of the conference (written by a staff criminologist who was present but did not participate in the RJ discussion) was provided to the judge for each of the burglary cases that had received an RJ conference. The latter element had been added at the request of the judges after the High Court had ruled that Judges should reduce time in prison when offenders had undertaken RJ conferences (R. v. Collins 2003).

Outcome measures

This article reports on five outcome measures for crime victims, two for the first hypothesis and three for the second. Measures of post-traumatic stress symptom levels are the subject of another report in preparation. Four of our outcome measures are available, and hence reported, across all four RCTs, all derived from Likert-scaled questions (except number 1 below) asked in the victim post-treatment interview:

1. Has the offender apologized to you?
2. (If apology) do you feel the apology was sincere?
3. Do you blame yourself or place any responsibility on yourself for the crime? (London) or Do you sometimes think that the incident might have been prevented if you had been more careful or less provoking? (Canberra).
4. Would you do some harm to the offender yourself if you had the chance?" (Canberra) or "Do you wish you could physically retaliate against the offender now?" (London).

The fifth outcome, available for controlled comparison only for London, was the question "Have you forgiven the offender?" We can also report this measure in Canberra for the RJ group only (as it was not asked of the control group).

In JPP and JVC all victim interviews were conducted face-to-face by the second author. In LOB and LOR all interviews were conducted by telephone by one of the

other authors. Both interviewers were female; in London the interviewer had an American accent, and in Australian the interviewer was Australian.

Sample size

In all four RCTs, sample size targets were set on the basis of statistical power calculations about the projected reduction in repeat offending among offenders. Because the base rate of repeat offending was predicted to be a smaller proportion of each treatment group than the various categories of victim response, we did not conduct separate power estimates for victim responses. Actual sample sizes achieved were larger than predicted in London, but smaller than predicted in Canberra. The RCTs in Canberra were conducted over a 5 year period, with a declining rate of referral over time from the operational police. The RCTs in London lasted slightly over 2.5 years with a steady rate of randomly assigned cases identified from probation records by the research team.

Random Sequence Formulae

A simple randomization, generated with a computer random number function, was used to split cases into predetermined blocks. Each block was constrained to split equally into experimental and control groups. The Canberra personal property trial began with a block of 150 cases. Later, this RCT had five additional 10-case blocks appended onto the end of its original sequence. A total of 173 of these 200 assignments were used. The Canberra violence experiment consisted of a single block of 100 assignments, all of which were used. Both London RCTs employed a single block of 300 case assignments which will not be used completely.

Allocation concealment

One of the greatest dangers of RCTs is that the allocation is actually biased by a preference for treatment choice by a member of operating staff. All four present RCTs invested great energy in preventing such bias. In all four experiments the treatment was concealed until after cases were enrolled and prior to random assignment; in none of them was treatment concealed thereafter. In the Canberra experiments, the research team agreed with the Australian Federal Police that they could not know the random assignment in advance of deciding whether or not a case was eligible – although some officers did ask on occasion when calling the research hotline. Police were asked to refer to the experiments those cases in which the offender had admitted to a property or violent offense that was “serious enough to be prosecuted in court but not so serious that it must be prosecuted in court.” This judgment was highly discretionary for each police officer. The aggregate result of their decisions ultimately excluded the majority of arrests for eligible

offenses from the RCTs, but prior to random assignment. Daily audits of the envelopes opened suggested that they all had case details entered at the time of opening the envelope, despite five late-night mistakes (out of 1,290 cases) as to which of the four RCT sequences of envelopes to open.

A further opportunity for eligibility screening prior to random assignment occurred when operational police telephoned the RISE 24 h hotline to refer a case to research staff. Checks were made to ensure the exclusion of age-ineligible offenders, those who had not made full admissions about (or declined to deny) their offense, and any case with domestic violence or sexual aspects. By the time the call was made the police had already sought the offender's consent to participation in the research and had explained the random assignment process. Victims' consent in Canberra was obtained only after random assignment to the experimental group: victims whose cases went to court in the usual way were not contacted until the post-treatment interview. Each case assignment was kept in an individual sealed envelope by research staff and only opened when the referring police officer had satisfactorily answered all screening questions. Because it was impossible for police or research staff (on pain of dismissal) to know in advance what the assignment would be before entering a case, the decision about whether or not to refer the case was thus made entirely independently of the treatment to be assigned.

In the London experiments the entire case referral and random assignment process was much more in the hands of the research staff. It was research staff who identified potentially eligible offenders from the Probation Service 'Tracker' and referred them to the police facilitators, with whom they shared office space. These police officers checked for offense eligibility by reference to the Metropolitan Police Crime Reporting Information System (CRIS) data base, where details of each incident are recorded. Grounds for exclusion included age of the offender (adults only), no guilty plea, no personal victim in the case, co-defendants in the case (excluded because of the difficulties in obtaining agreement to participate from more than one offender in such a tight time frame before the sentencing date) and domestic violence or sexual aspects in the case. When eligibility was established, facilitators contacted offenders to discuss the program and the research with them and to seek their consent to participation. After offender consent, the facilitator checked with research staff that all requirements had been met short of victim consent. If all was in order, and before approaching the victim, the research staff gave the facilitator a 'Go Code' number to quote if and when random assignment was sought. As soon as victim consent was given, the facilitator called a member of the research staff (located in Pennsylvania), gave the 'Go Code' number and was then given the assignment which had been automatically generated from the quasi-random number sequence held on a secure computer. This assignment was communicated separately to London research staff, who were thus able to verify that the facilitator was proceeding with the case according to the random assignment. (Facilitators were conscious of the often strong feelings of victims and offenders about the prospect of a conference; they sometimes reported considerable anguish about communicating an assignment that did not correspond to participants' wishes and about dealing with that disappointment.)

Implementing randomization

In the Canberra experiments, after the arresting officer had called and obtained random assignment, the offender (but not the victim) was informed immediately about which way the case would be dealt with. Those in the control group were charged and proceeded to court in the usual way. Those in the experimental group were informed that they would be contacted by a police RJ facilitator and asked to attend a conference. Following participation in the conference, including an outcome agreement signed by offenders, victims and police, the legal disposition would be a 'caution' – a record without a criminal conviction. It was the responsibility of the police facilitator to check on offender compliance with these undertakings.

In London, victims were told about random assignment before offenders. As soon as the assignment was given by phone to police officers in the presence of the victim, police immediately told victims whether or not there would be a conference. Offenders were advised as soon as possible thereafter. Facilitators had only a very limited period of time in which to set up a conference, as the sentencing date was already set for all participants. The date for the conference needed to be suitable not only for the offender and victim but also for the prison authorities (if the offender was held in custody): issues to be settled included the availability of space in which to hold the conference, prison staff availability and internal security arrangements.

In the case of the burglary experiment, a member of the research team attended conferences in order to prepare a report for the sentencing judge on what had transpired. In addition, an observer from the team of external evaluators from the University of Sheffield (appointed by the UK Home Office, the program funders) also attended each conference (Shapland et al. 2004). (Virtually all conferences had at least one observer who did not participate, sometimes including senior officials in government.)

All conferences concluded with a written agreement about what the offender would do to repair the harm caused by the offense. The range of options was sometimes limited by the prospect of a prison sentence. If this was likely, the agreements took incarceration into account by including promises such as the offenders seeking out drug treatment and employment programs; the agreements might also include community reparation to be undertaken if a non-custodial sentence was the outcome.

Blinding

While the ideal of a double-blind trial provides perhaps the highest degree of protection against bias in any aspect of a research design, it was not possible to achieve post-random assignment blinding in these experiments. In all four experiments it was plainly necessary for police to know whether each case was in the experimental or control group, once assignment had taken place, so blinding the police was not possible. In JPP and JVC diversion to conferencing was plain for

everyone to see, victims and offenders alike, because they were all directly involved in the process. For those assigned to the control group, no one other than the offender knew that the case was involved in RISE unless the offender chose to disclose it. In LOB and LOR, where restorative justice was in addition to court (not instead of it), virtually no blinding occurred. Offenders, victims and their supporters, and defense solicitors all knew about the conference, as did the sentencing judge who received a copy of the outcome agreement (and the report prepared by the research observer, in the case of the burglary offenses). But for cases assigned to the control group only, the victims, the offenders and the offenders' solicitors knew that they were part of an experiment. Legal representatives in court frequently alluded to offenders' agreement to participate in the experiments as evidence of 'mitigation' of the harm of the offense by the offender that could justify a less severe sentence. (The Court of Appeal agreed in *R v Barci* [October 2003] that an agreement to participate in restorative justice, even if it did not take place, was grounds for consideration in sentence determination; it had already decided in April 2003 in *R v Collins* that actual participation in restorative justice should be taken into account in mitigation).

Blinding was also impossible on the research team. It was necessary for all those following up on victims' and offenders' welfare and attitudes post-treatment to be aware of the treatment assignment in order to assess the impact of treatment and to compare the two groups.

Statistical methods

All staff were told from the outset of all four RCTs that we would compare victims on the basis of intention-to-treat (ITT) analysis. The plan was to use difference-of-means and difference-of-proportion tests, as well as odds ratios. These comparisons were planned for main effects only. When it became apparent that we had fairly even distributions of male and female victims in each RCT, we decided to disaggregate the outcome measures for the meta-analysis. The meta-analysis uses the standard approach and "forest graphs" generated by REVMAN software provided by the Cochrane Collaboration (www.cochrane.org). This article employs the inverse variance method of meta-analysis for analyzing the treatment effects across all studies. Each study is assumed to estimate exactly the same quantity, which justifies using a fixed effect meta-analysis. Finally, the standard error of the overall estimate is used to compute a *z*-test statistic testing the null hypothesis of no treatment effect.

RCT-specific differences

Table 2 sets out various elements relating to the four experiments in the analysis. Both the Canberra Violence (JVC) and Canberra Personal victim Property (JPP) experiments were finalized in July, 2000 after 5 years of data collection. The London Robbery and London Burglary experiments were still underway after 2

Table 2. CONSORT elements on cases collected and analyzed.

<i>CONSORT element</i>	<i>Canberra violence</i>	<i>Canberra property</i>	<i>London robbery</i>	<i>London burglary</i>
Participant flows	Victim interviews = 81 Victim treatment as intended = 84% of RJ (experimental) group (38 of 45 interviews)	Victim interviews = 151 Victim treatment as intended = 72% of RJ (experimental) group (51 of 71 interviews)	Victim interviews = 69 Victim treatment as intended = 100% of RJ (experimental) group (34 of 34 interviews)	Victim interviews = 145 Victim treatment as intended = 100% of RJ (experimental) group (70 of 70 interviews)
Recruitment dates	July 1995 to July 2000	July 1995 to July 2000	July 2002 to January 2005	July 2002 to January 2005
Numbers analyzed	81 victims interviewed E & C	151 victims interviewed E & C	69 victims interviewed E & C	145 victims interviewed E & C
Outcome estimation	See figures	See figures	See figures	See figures
Gender analyses	Yes	Yes	Yes	Yes
Adverse events	None	None	None	1 heart attack during an RJ conference

years, with planned termination date in early 2005; a number of cases were still to be finalized and many victims still to be interviewed as of this writing. The victim interview response rate for Canberra was 91% for Violence and 88% for Property; for London to date it has been 82% across all cases. There have been no statistically significant differences in victim response rates by treatment groups. Thus while attrition may have hidden some subset of victims who responded very differently to the randomly assigned conditions, it is less likely that attrition poses a threat to the internal validity of observed RJ vs. CJ differences among those victims who responded to the survey. With small samples it is always possible that slightly differential attrition may reduce baseline equivalence of the groups, but the use of meta-analysis across multiple point estimates reduces the impact of any one biased point estimate on the overall conclusion.

A higher proportion of the victims in the London cases have been treated as intended – somewhat higher than that achieved in Canberra. This resulted partly from building on the experience of Canberra in overcoming some of the administrative impediments to successful treatment, and partly from the fact that a high proportion of the London offenders were being held in custody pending their sentence in court at the time of their restorative justice conference, rather than on bail and hence less likely to attend RJ events. Moreover, none of the interviewed victims in London had been denied an RJ conference after one had been offered to them, as some had been in Australia. The mean level of victim satisfaction in London was higher because there was no subset of victims angry at losing their chance to meet with the offender (Strang 2002).

Table 3 shows baseline demographic data for victims in each treatment group in each of the four experiments. The percentage of women in the sample varied from a low of 27% for Canberra Violence to a high of 55% for London Robbery. There was a marked difference between the two geographic sites on race, reflecting the very different racial composition of the sites. Within the London experiments there was a higher percentage of nonwhite victims than in Canberra, and more nonwhite victims in LOR than in LOB; the Robbery victims also tended to be younger and

Table 3. CONSORT elements: Baseline data.

<i>Baseline data</i>					
<i>Demographic</i>	<i>Percent female</i>	<i>Percent non-White</i>	<i>Percent over 30</i>	<i>Percent married</i>	<i>Percent unemployed</i>
Canberra violence RJ	27	4	24	27	9
Canberra violence CJ	39	8	19	14	11
Canberra property RJ	48	3	61	61	0
Canberra property CJ	36	1	80	71	1
London robbery RJ	44	33	58	NA	NA
London robbery CJ	55	30	53	NA	NA
London burglary RJ	49	19	81	NA	NA
London burglary CJ	52	17	78	NA	NA

more often nonwhite than the Burglary victims. In Canberra, while the percentage of non-white victims was negligible (Aboriginal people make up less than 2% of the Canberra population), again the Violence victims were noticeably younger than Property victims, and much less likely to be married or employed.

Results

The results in Figures 1–5 are all presented as forest graphs, plotting the point estimates and confidence intervals for each test with each sample. The logic of a forest graph is that it displays an overall pattern of all available tests of a single hypothesis, and then estimates the probability of that pattern of results occurring by chance. While the interpretation of such graphs can be highly complex (Sherman and Strang 2004), in this article the results are relatively straightforward. Each graph shows, in a diamond at the bottom of the graph, the weighted average odds ratio of experimental vs. control results, with the odds ratios calculated on the basis of such simple methods as percentage of one randomly assigned group divided by the percentage of the other. Point estimates are located in relation to the distance between an odds ratio of zero and the estimated odds ratio in each test. The ‘tree’ lines around each point estimate in the forest graph indicate the confidence intervals around the point estimate. When a horizontal line crosses the sole vertical line in the middle of the graph, the 95% two-tailed confidence interval crosses zero, and the point estimate of the difference between experimental and control groups is judged to be non-significant. The same is true for the weighted average diamond at the bottom of the graph. Where none of the trees or the diamond touch the center vertical line, the differences are conventionally judged to be statistically significant ($P = 0.05$ or less).

The results on both criterion measures for Hypothesis 1 (“RJ conferences are more likely than conventional justice processes without RJ to produce a successful interaction ritual fostering social solidarity among offenders and victims that reaffirms their mutual group morality”) are consistent with the hypothesis. Figures 1 and 2 suggest that RJ conferences are significantly more likely to produce offender

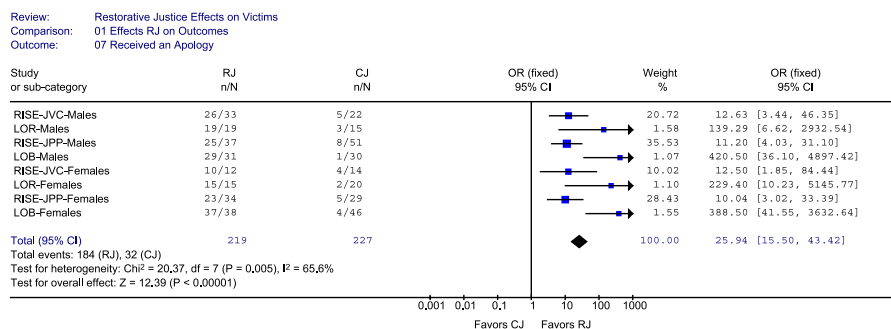


Figure 1. Victims received apology.

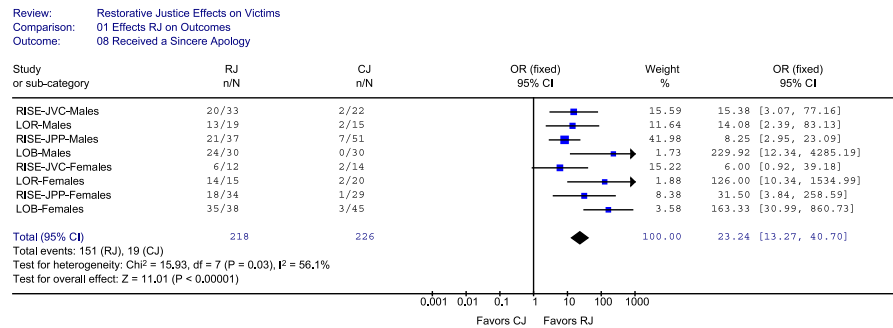


Figure 2. Received a sincere apology.

apologies to victims, who then perceive the apologies as sincere – affirming the offenders' commitment to group morality.

Apologies are rated as extremely important by victims for the repair of the emotional harm they have experienced (Strang 2002). In every experiment and for both sexes, victims were far more likely to receive an apology from their offenders if they were assigned to RJ. This is partly because few victims attend court. All cases required full admissions for entry to the experiments, and for the cases that went to court all involved a guilty plea. Control group victims were rarely in court as they were not required as witnesses and hence were not often notified that the case was to be heard. In addition, victims were prohibited from attending court proceedings in Canberra if an offender was a juvenile. Almost all apologies in the control groups were offered outside the court setting.

There were differences between the Canberra and London experiments. Apologies were transacted far more often (almost 100%) in London conferences than in Canberra conferences. This may account for the difference in levels of forgiveness felt by victims across all experiments (see Figure 3). Moreover, all London conference victims who received an apology did so in the course of the conference; in Canberra some apologies were transacted outside the conference setting, at the following percentages: JPP Male victims 8% outside conference, JPP Female victims 9% outside conference, JVC Male victims 12% outside conference, and JVC Female victims 0% outside conference.

These differences, however, do not seem large enough to invalidate the test of the hypothesis across all eight point estimates. The consistently larger effect sizes for the London experiments for apologies in Figure 1 are not matched by similar differences in the effect sizes for a sincere apology in Figure 2. Thus from the standpoint of IR theory, the RJ rituals in both Canberra and London succeeded in producing an outcome judged by the victims to be a successful recommitment to group morality – between 10 and 100 times more likely with RJ than without it.

The large odds ratios for sincerity of apologies (Figure 2) seem to be tied to the RJ process, rather than to apologies offered in any way. Not only was restorative justice in both sites and for both sexes more likely to result in apologies, but victims in all experiments generally rated apologies given by offenders randomly

assigned to a conference as sincere. Overall, female RJ victims showed the highest ratings of the sincerity of apology, in both cities, relative to male victims. London victims tended more often than Canberra victims to rate the apology as sincere, though London male robbery victims were more skeptical than others and London male burglary victims the least skeptical.

The sincerity of apology, however, was not enough to produce much forgiveness of offenders, making the evidence for Hypothesis 2 less consistent (“Victims randomly assigned to RJ conferences are more likely to show future psychological benefits from the justice process than victims not assigned to RJ”). The London-only evidence on victim forgiveness in Figure 3 for Hypothesis 2 is much weaker and less consistent than the large effect sizes in Figures 1 and 2, with the weighted average in Figure 3 barely significant. Burglary victims, but not robbery victims, assigned to RJ are more likely to forgive the offender than control group victims, and females significantly so. (We have no comparable data for Canberra). But random assignment to restorative conferences does not result in victims saying significantly less often than controls that they blame themselves for the crime, in any of the eight tests – for which the summary effect of RJ is zero.

Figure 3 contains experimental vs. control data only for the London experiments. RJ group victims (only) in Canberra’s JPP and JVC experiments were asked “Since the conference, in thinking about the offender(s), you have felt: Very unforgiving; Unforgiving; Neither forgiving nor unforgiving; Forgiving; Very forgiving.” The results show a much lower level of forgiveness among victims offered RJ in Canberra (ITT) than among victims given RJ in London. The percentages of male and female Canberra victims offered RJ who said they were either ‘forgiving’ or ‘very forgiving’ were as follows: JVC-Males: 41%; JVC-Females: 40%; JPP-Males: 28%; JPP-Females: 44%. This average of 38% among Canberra victims offered RJ compares to an average forgiveness level of 75% among victims offered RJ in London. It is all the more striking that London victims much more often forgave their offenders than Canberra victims did when we consider the much greater seriousness of the offenses in London.

Hypothesis 2 suffers most in relation to the evidence in Figure 4, which directly contradicts the hypothesis. As the location of the weighted mean average right on the center line indicates, there are no consistent differences in self-blame across treatment groups, genders, cities or RCTs. (This analysis is based on proportions of

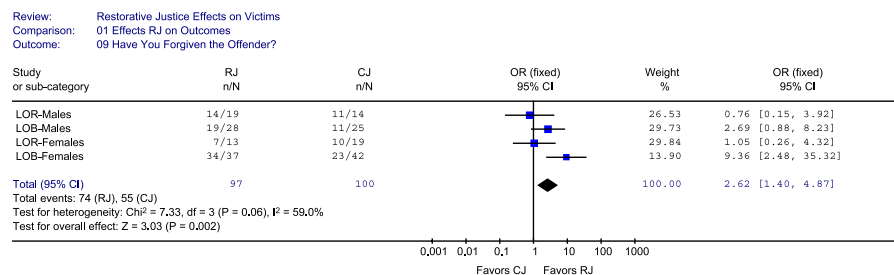


Figure 3. Forgiveness (UK experiments).

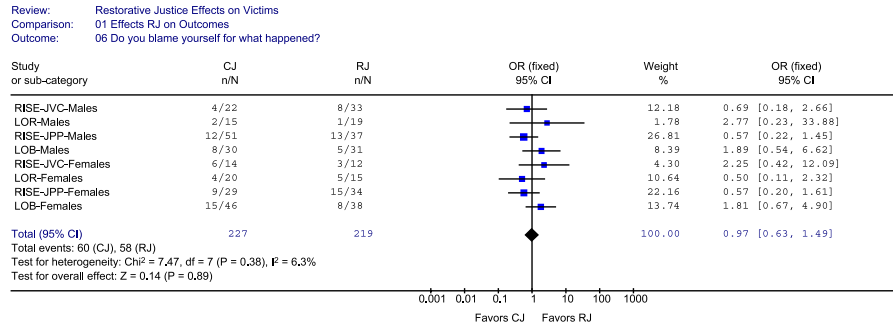


Figure 4. Victim self-blame for the crime.

victims who ‘agreed’ or ‘strongly agreed’ with a self-blame statement.) Despite our clinical impressions that victims were greatly relieved to hear from offenders that the crime had occurred at random, the effect of that information did not seem to reduce self-blame. The analogue between RJ and the CBT methods of Foa and her colleagues therefore seems to fail on this one measure. While experimental comparisons show clear benefits for reducing PTSS in victims in early analyses of the London data (Angel 2004), that benefit does not seem to hinge on a reduction in self-blame for the crime at the initial interview.

We can also eliminate level of satisfaction with the RJ conferences as an issue in interpretation of the self-blame data. The victims assigned to RJ in London were far more likely to be ‘satisfied’ with the conference (97%) than the victims assigned to RJ in Canberra (61%). They were also more likely to be “pleased with the outcome” in London (92%) than in Canberra (62%). But in neither city did RJ reduce victim self-blame.

In both cities, however, the evidence strongly supports Hypothesis 2 with what may be the measure with the greatest construct validity. If the victim’s own commitment to shared morality is best indicated by the level of the victim’s willingness to obey the law, then RJ clearly increases that commitment. Victim desire for violent revenge against the offender is consistently and strongly reduced by random assignment to RJ (Figure 5). In eight out of eight gender-specific tests,

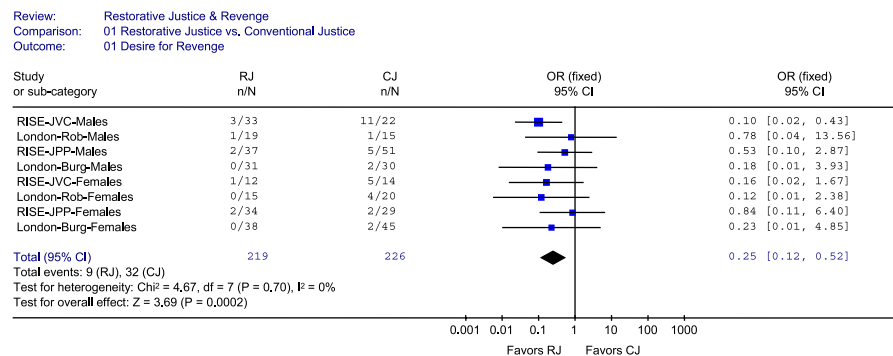


Figure 5. Post-treatment victim desire for revenge

victims assigned to RJ were less likely than those not assigned to say they either 'agreed' or 'strongly agreed' with the statement that they would "harm the offender if [they] had the chance" (Canberra) or "wished [they] could physically retaliate against the offender now" (London). While only one of the eight estimates is statistically significant in itself, the overall pattern of point estimates is extremely unlikely to be due to chance. The substantial differences in control group levels of desire for revenge across experiments and cities (see Sherman 2006) yields no indication of an interaction effect, as the *F*-squared test result indicates, because the effect sizes are so similar regardless of the base levels of desire for revenge.

Discussion

These results show that from a crime victim's perspective, restorative justice conferences create a successful interaction ritual for renewing commitment to group morality. Offenders in RJ are many times more likely to admit that they breached their moral obligations, and by apologizing reaffirm their commitment to those obligations, than similar, willing, offenders who are not allowed to engage in RJ. The apologies offered in RJ are perceived by victims as sincere, as a further indication of a successful interaction ritual. These conferences also succeed in 'normalizing' victim contact with an offender, as required by Cognitive Behavioral Therapy, in order to make discussion of the crime and the nature of the criminal a topic less threatening by virtue of becoming more familiar.

The interaction ritual of reconciliation in which victim and offender participate in a successful restorative justice encounter, where apologies are offered and accepted (and sometimes forgiveness offered in return), meets the conditions proposed by Collins (2004) for mutual entrainment leading to a successful interaction ritual. The broken bond that the offense represents, and that is the source of shame for victim and offender alike (Scheff 1990), is transformed by the emotional energy released by the conference. The interaction ritual can thereby help to symbolize the effort to repair the harm, consistent with other evidence on how victims describe their experience with RJ (Strang 2002; Strang and Sherman forthcoming).

The results of these rituals are not as consistent as the elements of the ritual itself, at least with the minimal available measures. Victims do not consistently forgive offenders simply because they have received a sincere apology. RJ, moreover, has no effect on the tendency of crime victims to blame themselves – in contradiction of the goal (and frequent result) of standard Cognitive Behavioral Therapy.

These measures are by no means ideal, let alone exhaustive, tests of the two theoretical frameworks described at the outset. The debatable nature of their construct validity reflects the larger absence of theoretical development in restorative justice theory of victim effects. Such theory can have many dimensions, including mental and physical health of crime victims. Future analyses of the

London RJ experiments, for example, will examine post-traumatic stress symptoms of victims who do and do not receive restorative justice. The value of such experiments, to the extent that multiple measures can be drawn from them, is that they can test a wide range of theoretical perspectives, many of which have strong policy relevance. If substance abuse or physical health of victims were to be substantially improved by RJ, saving substantial sums in medical costs for cancer or heart disease over a victim's lifetime, the policy implications for investing in the expansion of RJ services could be enormous. Further measures can still be gathered with the subjects in the present experiments, with many more such questions to be addressed.

As a test of the two theoretical frameworks employed in the present analysis, the measures are not as central to the constructs to be tested as other measures which can be used in the future. The cognitive behavioral therapy theory will be better tested in a forthcoming analysis of PTSS, while the interaction ritual theory was not even published when the present experiments were designed.

Nonetheless, all of these measures offer small building blocks for inductive theory about victim reactions to RJ. Forgiveness of offenders and self-blame for crime, whatever they say about any one specific theory, will remain important for understanding the emotions and cognitive perspectives victims carry away from their experiences with RJ. These facts do not complete the puzzle of the two theories tested, but they do stand firmly established in a larger arena of evidence on a leading innovation in justice.

What the measures may lack in construct validity may be compensated for in part by what they offer in range: the fact that these data are drawn from four experiments conducted in different nations, different kinds of offenders and offenses, at different points in the criminal justice process. When the measures are consistent – as most of them are – they reveal important effects of restorative justice on victims, and its capacity to deliver apologies far more often than conventional justice, regardless of gender or punishment.

From a justice policy perspective, the most favorable indication of the effect of RJ on victims is also the most important: the substantial and consistent reduction in the stated desire of victims for violent revenge against offenders. While RJ began with a primary focus on the goal of reducing repeat offending by offenders, these results suggest that it may be possible that RJ reduces the risk of future crime by crime victims. Many offenders are themselves crime victims, in one way or another. Future research could assess the extent to which official crime data are consistent with the stated intentions of victims. In the meantime, the victim interview data provide support for the claim that RJ benefits victims, if only in reducing their level of anger at offenders – and in reaffirming their own commitment to conventional group morality (i.e., obeying the law).

A further finding of great policy relevance is what this analysis does not show: any evidence of harm to victims for participation in face-to-face conferences with criminals who have hurt them. Given the level of seriousness and even violence in the crimes included in these samples, the lack of harm to victims actually falsifies the concern often expressed that crime victims may be 're-victimized' by the RJ

process. Their forgiveness levels, for example, could (have been) lower for the RJ group than for the CJ group. Or their desire to effect physical revenge against offenders could just as logically have been increased by an RJ conference, rather than the consistent reduction we observe. The measures are indeed suitable as partial tests of the victim harm hypothesis, which they falsify.

The results in the present paper are incomplete in several respects. There are more outcome measures to be analyzed and reported within these four tests. There are more tests in progress, both by the present authors and by other investigators. This paper does not begin to constitute a systematic review of the effects of restorative justice on victims. Such a review is underway with an approved protocol from the Campbell Collaboration Crime and Justice Group. Any overall judgment on the effectiveness of RJ on providing victim benefits, or effects on repeat offending, should await the completion of that review (see www.aic.gov.au/campbellcj/). Even more important, such conclusions should be continually modified by new findings added to that review. Seen as an evidence-based innovation for reducing harm, restorative justice will always remain a work in progress, open to its own improvement through better knowledge of its consequences.

Acknowledgements

We wish to acknowledge the financial support of the Home Office for England and Wales, the Esmee Fairbairn Foundation, the Smith Richardson Foundation, the National Institute of Justice, and the participating agencies that made the research possible. We offer special thanks to the Jerry Lee Foundation which has made the multi-site RCT program possible.

References

- Angel, C. (2004). Crime Victims Meet Their Offenders: Testing the Impact of Restorative Justice Conferences on Victims' Post-Traumatic Stress Symptoms. Paper presented to the American Society of Criminology, Nashville, Tennessee, 17.
- Black, D. (1982). Crime as social control. *American Sociological Review* 48, 34–45.
- Braithwaite, J. B. (1989). *Crime, shame and reintegration*. Cambridge: Cambridge University Press.
- Braithwaite, J. B. (2002). *Restorative justice and responsive regulation*. Oxford: Oxford University Press.
- Collins, R. (2004). *Interaction ritual chains*. Princeton, New Jersey: Princeton University Press.
- Foa, E. B. & Kozak, M. J. (1986). Emotional processing of fear: Exposure to corrective information. *Psychological Bulletin* 99, 20–35.
- Foa, E. B. & Meadows, E. A. (1997). Psychosocial treatments for posttraumatic stress disorder: A critical review. *American Review of Psychology* 48, 449–480.

- Harris, N. (2002). Shaming and shame: Regulating drink-driving. In E. Ahmed, N. Harris, J. Braithwaite & V. Braithwaite (Eds.), *Shame management through reintegration*. Cambridge: Cambridge University Press.
- Moher, D., Schulz K. F. & Altman, D. (2001). The CONSORT Statement: Revised recommendations for improving the quality of reports of parallel-group randomised trials. *Journal of the American Medical Association* 285, 1987–1991.
- Rothbaum, B. O. & Foa, E. B. (1999). Exposure Therapy for PTSD, *PTSD Research Quarterly*, The National Center for Post-Traumatic Stress Disorder, White River Junction, Vt, vol 10, 2:1–8.
- Ruth, H. & Reitz, K. R. (2003). *The challenge of crime*. Cambridge, Mass.: Harvard University Press.
- Scheff, T. J. (1990). *Micro-sociology: Discourse, emotion and social structure*. Chicago: University of Chicago.
- Shapland, J., Atkinson, A., Colledge, E., Dignan, J., Howes, M., Johnstone, J., Pennant, R., Robinson, G. & Sorsby, A. (2004). *Implementing restorative justice schemes (Crime Reduction Programme): A report on the first year*. Home Office Online Report 32/04. London: Home Office (<http://www.homeoffice.gov.uk/rds/pdfs04/rdsolr3204.pdf>).
- Sherman, L. W. (1993). Defiance, deterrence and irrelevance. *Journal of Research in Crime and Delinquency* 30, 445–473.
- Sherman, L. W. (2006). Revenge may be sweet, but restorative justice can suppress the appetite. In S. Karstedt, H. Strang, & I. Loader (Eds.), *Emotions, Crime and Justice: A Volume of the International Institute for the Sociology of Law, Onati, Spain*. Oxford: Hart Publishing.
- Sherman, L. W. & Strang, H. (2004a). Verdicts or inventions? Interpreting results from randomized controlled experiments in criminology. *American Behavioral Scientist* 47, 575–607.
- Sherman, L. W. & Strang, H., in collaboration with Daniel J. Woods, Caroline M. Angel, Geoffrey C. Barnes, Nova Inkpen, Dorothy Newbury-Birch and Sarah Bennett (2004b). *Restorative justice: What we know and how we know it*. Philadelphia: University of Pennsylvania, Jerry Lee Center of Criminology, <http://www.sas.upenn.edu/jerrylee/rjWorkingPaper1.pdf>.
- Strang, H. (2002). *Repair or revenge: Victims and restorative justice*. Oxford: Oxford University Press.
- Strang, H. & Sherman, L. W. (2006). Victim evaluations of face-to-face restorative justice experiences: A quasi-experimental analysis. *Journal of Social Issues*.
- Tyler, T. (1990). *Why people obey the law*. New Haven: Yale University Press.

About the authors

Lawrence W. Sherman is Director of the Jerry Lee Center of Criminology and Chair of the Department of Criminology at the University of Pennsylvania, where he is the Albert Monroe Greenfield Professor of Human Relations. He has designed and directed over 25 randomized controlled trials in crime and justice, including the Jerry Lee Program of Randomized Controlled Trials in Restorative Justice. He is a Fellow, Sutherland Awardee and Past President of the American Society of Criminology, and was the founding President of the Academy of Experimental Criminology.

Heather Strang is Director of the Centre for Restorative Justice at the RegNet, Research School of Social Sciences, Australian National University, and co-director of the Jerry Lee Program on Randomized Controlled Trials in Restorative Justice. She is a Fellow of the Academy of Experimental Criminology and the author of *Repair or Revenge: Victims and Restorative Justice*.

Caroline Angel was awarded the PhD in nursing and criminology in 2005 from the University of Pennsylvania, where she is a member of the Jerry Lee Program on Randomized Controlled Trials in Restorative Justice at the Jerry Lee Center of Criminology. Her dissertation measured post-traumatic stress symptoms of robbery and burglary victims randomly assigned to restorative or conventional justice.

Daniel J. Woods is a PhD candidate in criminology at the University of Pennsylvania, where he is a research assistant in the Jerry Lee Program on Randomized Controlled Trials in Restorative Justice at the Jerry Lee Center of Criminology.

Geoffrey C. Barnes is Director of Data Analysis for the Jerry Lee Program on Randomized Controlled Trials in Restorative Justice at the Jerry Lee Center of Criminology, University of Pennsylvania. His personal research interests include tests of procedural justice in the Canberra RISE project.

Sarah Bennett is assistant director of the Justice Research Consortium – UK for the Jerry Lee Program on Randomized Controlled Trials in Restorative Justice at the Jerry Lee Center of Criminology, and a candidate for the PhD in criminology from Cambridge University.

Nova Inkpen served from 2001 to 2005 as assistant director of the Justice Research Consortium – UK for the Jerry Lee Program on Randomized Controlled Trials in Restorative Justice at the Centre for Restorative Justice at the RegNet, Research School of Social Sciences, Australian National University.