

## The Paradoxical Consequences of Revenge

Kevin M. Carlsmith  
Colgate University

Timothy D. Wilson  
University of Virginia

Daniel T. Gilbert  
Harvard University

People expect to reap hedonic rewards when they punish an offender, but in at least some instances, revenge has hedonic consequences that are precisely the opposite of what people expect. Three studies showed that (a) one reason for this is that people who punish continue to ruminate about the offender, whereas those who do not punish “move on” and think less about the offender, and (b) people fail to appreciate the different affective consequences of witnessing and instigating punishment.

*Keywords:* affective forecasting, punishment, revenge, rumination

On April 16, 2007, Seung Hui Cho massacred 32 people at Virginia Tech University before turning the gun on himself. In the aftermath, some lamented that his death had robbed the survivors of the emotional satisfaction of exacting their own revenge. As one woman from the Midwest wrote on an Internet blog, “I don’t think there would be anything temporary about the satisfaction I would feel in being permitted to execute the person who killed my child” (Barnstead-Klos, 2007).

“Sweet is revenge,” wrote Lord Byron (*Don Juan*, Canto I, Stanza 124). Indeed, this belief may be both venerable and widespread, but is it right? Research in psychology and behavioral economics has shown that people often miscalculate their hedonic reactions to future events because they miscalculate how often they will think about those events once the events have ended (Schkade & Kahneman, 1998; Wilson & Gilbert, 2003; Wilson, Wheatley, Meyers, Gilbert, & Axsom, 2000). We hypothesized that acts of revenge would cause people to continue thinking about the transgressor whom they have punished, which would prolong their hedonic reactions to the transgression rather than shorten them. Thus we predict that people punish others, in part, to repair their negative mood and to provide psychological closure to the precipitating event, but that the act of punishment yields precisely the opposite outcome. As the poet John Milton wrote, “Revenge, at

first though sweet, Bitter ere long back on itself recoils” (*Paradise Lost*, Book IX, Line 171).

Our research, then, focuses on three questions: (a) What is the affective consequence of exacting revenge? (b) Do people accurately predict this consequence, and, if not, (c) Why do people make this error of prediction?

Although there is an enormous amount of literature on punishment, it has generally focused on the justification for imposing punishment (Bentham, 1843/1962; Ezorsky, 1972; Kant, 1790/1952) or on the consequences of receiving punishment (Skinner, 1938, 1974; Watson, 1924). More recently, psychologists have begun to explore the motives behind the desire to punish (see Carlsmith & Darley, 2008, for a review), but there has been little empirical research performed on the question of whether instigating punishment against another person has affective consequences for the one who punishes. This absence is both surprising and unfortunate given the ubiquity of punishment across societies (Gintis, 2008) and the self-evident importance of this punishment to both instigator and target. Although this discussion is potentially relevant to the myriad forms of punishment (e.g., societal, legal, organizational, etc.), we focus specifically on interpersonal aggression in response to individual norm-violating behaviors. This sort of aggression is typically referred to as revenge or retaliation by laypeople (McKean, 2005), psychologists (Frijda, 1994), and philosophers (Baier, 1955; Hobbes, 1950; Quinton, 1954).

### Catharsis

People believe in the utility of aggression. There is widespread acceptance of the notion that aggression, either directed against the target of one’s ire or displaced to some other object, relieves the tension, and thus the anger, that had been pent up inside. Evidence for this belief abounds in our language (“get it out of your sys-

---

Kevin M. Carlsmith, Department of Psychology, Colgate University; Timothy D. Wilson, Department of Psychology, University of Virginia; Daniel T. Gilbert, Department of Psychology, Harvard University.

We acknowledge the assistance provided by James Elmers, Alison Potocki, Emily Drummond, Sue LeBarron, and Rachel Smith in conducting the experiments.

Correspondence concerning this article should be addressed to Kevin M. Carlsmith, Department of Psychology, Colgate University, 13 Oak Drive, Hamilton, NY 13346. E-mail: kcarlsmith@colgate.edu

tem”), in our public service advertisements (“punch a wall, not your child”), in plays (*The Spanish Tragedy*, *Hamlet*), novels (*The Count of Monte Cristo*, *Moby Dick*, *The Cask of Amontillado*), and films (*Death Wish I–V*, *Revenge*, *Kill Bill*). Bushman (2002) cited numerous self-help books (e.g., Lee, 1993) that extol the virtues of the cathartic release that comes from punching pillows or other objects that stand in for the true target.

This belief in catharsis is not surprising, perhaps, given psychology’s own deep roots in the theory. Early psychological research on aggression adopted the hydraulic model, in which aggression was perceived to be a cathartic release to the pressure brought about by frustration or another stimulus (Berkowitz, 1989; Dollard, Doob, Miller, Mowrer, & Sears, 1939; Freud, 1922/1961). In this model, the eliciting stimulus creates an internal pressure that must be released, and the most likely form of release is aggression. Once cathartic aggression takes place, subsequent aggression is less likely. Thus, venting one’s aggression on an appropriate object (e.g., a pillow) was seen as an effective way to reduce interpersonal aggression and spirals of conflict.

Although the early research on catharsis was mixed, more recent investigations have come to a clear consensus against the theory. Bushman (2002) demonstrated that cathartic aggression not only failed to reduce subsequent aggression, but in fact increased it. Participants in his experiment were initially angered by a fellow student who provided insulting feedback on an essay and were then given an opportunity to vent their anger on a punching bag. Shortly thereafter, participants played a competitive game with the object of their anger and were given the opportunity to deliver aversive noise blasts. Bushman found that use of the punching bag actually increased subsequent aggression rather than reducing it. Moreover, he found that participants who imagined that the punching bag was the other student (a rumination manipulation) aggressed even more than did those participants in the control condition and those focused on the health benefits of the punching bag (a distraction manipulation). Critical for our purposes, participants who vented their anger and subsequently aggressed against the other student also experienced more anger at the end of the experiment.

Thus, Bushman (2002) found that physical venting did not reduce the desire to aggress and that the combination of venting and aggressing led to increased anger. The focus of Bushman’s experiment, though, was on the effects of cathartic aggression, and thus the aggression against the offending student (e.g., the noise blasts) served as a dependent measure rather than as an independent variable. It is thus unclear whether it was the venting or the aggression that increased the anger and whether the aggression caused the anger or whether the anger caused the aggression.

### Cathartic Aggression as Mood Repair

Other research has shown that participants aggress against offending others in the belief that it will improve their mood (Bushman, Baumeister, & Phillips, 2001). Participants were angered and then given an opportunity to aggress against the offender, and those who believed in the cathartic value of aggression—either through preexisting attitudes or through manipulation—aggressed more severely than did those who did not believe in catharsis. This effect disappeared, however, when students believed they had

ingested a “mood-freezing” pill. Thus, participants were strategically using aggression to repair their own mood.

Bushman et al. (2001) primarily examined people’s beliefs about the hedonic utility of aggression. In one study, however, they examined the actual impact of aggression on subsequent mood. Those who punished most severely (because they believed in the utility of punishment) in fact experienced increased negative affect. Once again, however, the aggression variable was measured rather than manipulated. The authors concluded that these results “may be relevant to the question of whether aggression actually does accomplish mood repair, although our study was not designed to provide definitive evidence on that question” (Bushman et al., 2001, p. 26).

In sum, research shows that cathartic venting does not reduce subsequent aggression against the offending object. However, people have a clear belief in the value of catharsis and regulate aggression in service to mood. But it is unclear whether aggression improves mood or if, indeed, it actually makes the aggressor feel worse. In order to examine this question, we used punishment as an independent variable (people did or did not have the opportunity to punish an offender), with the hypothesis that people would take advantage of the opportunity to punish but that doing so would prolong rumination and negative affect.

### The Accuracy of Affective Forecasting

There is ample evidence that people often make mistakes when forecasting their future affective reactions (Loewenstein, O’Donoghue, & Rabin, 2003; Mellers & McGraw, 2001; Wilson & Gilbert, 2003), and the mistaken beliefs about the utility of aggression described above may well be a case in point.

The most commonly observed error in affective forecasting is the *impact bias*, whereby people overestimate the intensity and duration of their future affective reactions. People have been found to overestimate the hedonic impact of a wide range of events, from romantic breakups and political elections to the loss of a job or receipt of unwanted results of a pregnancy test (for a review, see Wilson & Gilbert, 2003). One cause of the impact bias is *focalism*, the tendency to overestimate how much an event that one is thinking about will occupy one’s thoughts as time goes by. Wilson et al. (2000), for example, found that fans overestimated how happy or sad they would be after their favorite team won or lost a football game, in part because they overestimated how much they would think about the game after it occurred (see also Schkade & Kahneman, 1998). Put differently, people underestimated the extent to which the everyday routine of their lives would occupy their thoughts, crowding out thoughts about the game.

Thoughts about some events, however, get crowded out more easily than others. There is evidence that anger, in particular, is an attention-focusing emotion, making it difficult to think about other things (Fredrickson & Branigan, 2005). Anger thoughts can thus be a vicious cycle; the more people think about them the angrier they get, and the angrier they get, the harder it is to think about anything else (Bushman, 2002; Konecni, 1974; Rusting & Nolen-Hoeksema, 1998). Anger may thus be an exception to previous research on affective forecasting and focalism. Instead of overestimating the intensity and duration of their anger following punishment, people may actually underestimate it. People’s lay theory seems to be that revenge is a cathartic act that will bring them

closure, allowing them to stop thinking about the precipitating event, when in fact that course of action might keep their attention on the event and prevent them from coping in other ways.

Suppose, for example, that people did not have the opportunity to punish the object of their ire. They would be forced to deal with their negative feelings in other ways, perhaps by deciding that in the grand scheme of things, the person's actions were not very important and not worth worrying about. By trivializing the behavior, people could move on and not think about it further (Simon, Greenberg, & Brehm, 1995). Ironically, the act of punishing might increase ruminations about the incident and prevent people from dealing with their negative emotions in other ways, such as thinking about unrelated issues, which has been shown to reduce anger (e.g., Bushman, 2002).

### Study 1

We hypothesized that people believe that punishing an offender will improve their mood and bring about psychological closure, but in fact punishment will increase rumination about the offender and lead to a continuation of negative affect. Indeed, there might be a recursive relationship between affect and rumination such that increased negative affect also leads to increased rumination. To test these hypotheses, we used the "free rider paradigm" in which participants played multiple trials of an interactive game with three other people. Players could earn money if they cooperated, but a player who did not cooperate (a "free rider") could earn more and cause others to earn less. Unbeknownst to the participants, the game was staged so that on several trials one of the players encouraged others to cooperate and then did not cooperate himself or herself. *Punishers* were given the opportunity to financially penalize this free rider at a small cost, whereas *nonpunishers* played the same game but were given no opportunity to penalize the free rider. Both groups reported their hedonic reactions, and punishers also estimated their hedonic reactions had they had no opportunity to penalize the free rider. An additional group of *forecasters* played the game and predicted what their hedonic reactions would be if they penalized the free rider. There were thus three conditions: punishers, nonpunishers, and forecasters. The latter group made prospective predictions of affect, but we also asked the punishers to make retrospective predictions of affect had they been in a different condition.

### Method

*Participants and procedure.* Forty-eight university students were allotted a stake of \$1.00 and told that they could invest in the group or save it. After each round of play the total invested amount from all 4 players—plus a 40% dividend—was distributed to all players equally regardless of individual contribution. This arrangement created a classic prisoner's dilemma: The optimal group outcome was achieved when each individual cooperated, but each individual maximized his or her personal outcome by defecting.

Participants generally convened in groups of 4. When fewer than 4 participants were available, participants were led to believe that a 4th individual showed up after the original members entered their cubicles. Participants played a few verbal practice rounds with the experimenters to ensure they understood the instructions. Once the game began, all of the interactions occurred in private rooms via individual computer terminals.

Computer software simulated the play of the other participants. Two were programmed to cooperate initially and to use a tit-for-tat strategy on subsequent rounds. The 3rd player, always a female, initially bid \$0.50 and sent everyone an instant message urging them to cooperate. She then defected on all subsequent rounds. After each of eight investment decisions, participants received feedback about each group member's investment, a record of each person's total earnings, and a prompt to invest for the following round. The program was convincing; few participants doubted that the other players were fellow students.

Participants were randomly assigned to a punishment, no-punishment, or forecaster condition. At the end of the game, participants in the punishment condition were told that, if they wanted, they could anonymously penalize the other players by subtracting points from these players' total earnings. Each point cost the punisher \$.05 but reduced the punished member's profit by three times that amount. After indicating whether they wanted to punish any of the other players and by how much, participants in the punishment condition completed the dependent measures. Participants in the no-punishment condition completed the dependent measures right after the conclusion of the game with no mention of the opportunity to penalize other players. Participants in the forecaster condition received the same instructions participants in the punishment condition received about penalizing the other players, except that instead of actually being given the opportunity to punish, they were asked to imagine that they could do so. They then completed the dependent measures according to how they thought they would feel if they had had the opportunity to punish.

*Dependent measures.* Participants in the punishment and no-punishment conditions rated their hedonic reactions at the conclusion of the game and 10 min later after completing a filler task (writing a detailed description of their typical day). At each time point they rated the extent to which they were experiencing positive affect (pleased, positive, satisfied) and negative affect (negative, vengeful, irritated) on 7-point scales anchored by *not at all* and *extremely*. Participants also reported the extent to which they were thinking about each of the other players, again on 7-point scales (1 = *not at all*, 7 = *very much*). Participants in the punishment condition were also asked immediately postpunishment to estimate what they thought their hedonic reactions would be (on the same scales) if they had not had the opportunity to penalize the free rider. After the game all participants were probed for suspicion, debriefed, and paid slightly more than they expected to earn in the game.

### Results

*Preliminary analyses.* Three participants failed to play cooperatively and therefore earned more money than each of the other players. Thus, each of these participants was the free rider in the group and had an entirely different psychological experience than did the other participants. We excluded these participants (2 in the punishment condition, 1 in the no-punishment condition) from all analyses reported below, leaving a total of 15 participants in each condition. Participants earned an average of \$2.51 (excluding the cost of punishing), whereas the free rider earned \$5.59,  $t(44) = 10.41$ ,  $p < .01$ . Fourteen of 15 people in the punishment condition chose to spend their money to punish the free rider, with an average outlay of \$0.51. This resulted in an average punishment of

\$1.53. No participant penalized either of the other (cooperative) players in the game.

**Data reduction.** Initial analyses revealed that the positive and negative affect items were highly correlated and that the manipulations had similar effects on each type of item. We thus reverse scored the negative items and averaged them with the positive items. Initial analyses also revealed that although there was a main effect of time on this index,  $F(1, 41) = 28.79, p < .001$ , reflecting the fact that participants' mood improved during the 10-min distraction test, time did not interact with the experimental manipulations,  $F(2, 42) = 0.12, p = .89$ . In order to simplify the analyses, we thus averaged the affect ratings over time, resulting in a highly reliable index ( $\alpha = .93$ ). We also report the results of individual affect items.

**Hedonic reactions.** Did punishing the free rider change people's affect? The answer is yes: People in the punishment condition actually felt worse than people in the no-punishment condition. Did people anticipate this effect? The answer is no; those in the forecaster condition, who rated how they would feel if they could punish, predicted that they would feel better than people who actually were allowed to punish. The main effect of condition on affect was significant,  $F(2, 42) = 6.83, p < .01$ , and planned contrasts revealed that people in the punishment condition ( $M = 3.52, SD = .95$ ) reported a significantly worse mood than did people in both the no-punishment ( $M = 4.79, SD = 1.22$ ) and forecaster ( $M = 4.53, SD = .77$ ) conditions,  $ps < .01$ . The no-punishment and forecaster conditions were not significantly different from each other ( $p = .47$ ).

A detailed analysis of each affect item confirms the findings described above. Each negatively valenced item was higher for the punishers and lower for the nonpunishers. Similarly, each positively valenced item was lower for the punishers and higher for the nonpunishers. Each of these differences was significant by Tukey's least significant difference (LSD) at  $p = .05$ . The largest mean differences between punishers and nonpunishers were for the items vengeful (2.07), irritated (1.67), negative (1.27), and positive (-1.20). Thus it is primarily (but not exclusively) an increase in negative affect that results from administering punishment.

**Mean differences in rumination.** There were no differences between the punishment and no-punishment conditions at Time 1 in how much people reported thinking about the free rider ( $M = 6.07$  vs.  $5.73, F < 1.0$ ). However, people in the punishment condition reported thinking about the free rider significantly more at Time 2 than did people in the no-punishment condition ( $M = 4.33$  vs.  $2.67$ ),  $F(1, 28) = 4.46, p = .04$ . The Time  $\times$  Condition (punishers vs. nonpunishers) interaction was significant,  $F(1, 28) = 4.25, p < .05$ , indicating that although all participants initially focused attention on the free rider, the act of punishment kept the free rider at the forefront of the participants' thoughts. Nonpunishers, by contrast, were able to focus attention elsewhere.

**Rumination as mediator between punishment and affect.** Figure 1 reveals that rumination about the free rider at Time 2 was related to both punishment ( $b = .43$ ) and affect ( $b = -.42$ ) as required for a mediated effect, and that the direct effect of punishment on affect was reduced when rumination was included in the equation. The beta coefficient dropped from  $b = -.49$  to  $-.31$ , which was marginally significant by the Sobel test ( $z = 1.77, p = .08$ ). The same analysis does not hold for rumination at Time 1, suggesting that it is the continued rumination that leads to negative affect rather than initial rumination.

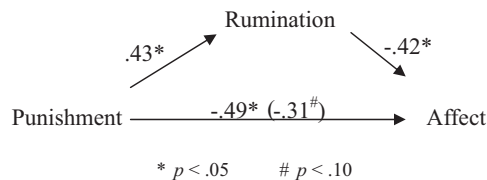


Figure 1. Study 1: The negative relationship between punishment of the perpetrator and personal affect is mediated by rumination about the perpetrator.

**Affect as mediator between punishment and rumination.** We anticipated, and found, a recursive relationship between affect and rumination such that each variable served as a partial mediator for the effect of punishment on the other. Thus, affect at Time 2 was related to both punishment ( $b = -.49$ ) and rumination ( $b = -.40$ ), and the direct effect of punishment on rumination was reduced when affect was included in the equation. The beta coefficient dropped from  $b = .43$  to  $.22$ , which was also marginally significant by the Sobel test ( $z = 1.77, p = .08$ ).

**Affective forecasts.** As noted above, forecasters predicted that if they could punish the free rider they would feel significantly better than punishers reported feeling. That is, they failed to anticipate the negative hedonic effects of punishment. This was also true at the individual level: The correlation between forecasters' predicted affect and predicted level of punishment was  $r(15) = .48, p = .07$ . This is in stark contrast to the negative correlation in the actual punishment condition,  $r(15) = -.50, p = .06$ . The difference between these correlations is significant ( $z = 2.63, p < .01$ ), showing that actual and predicted experience ran in opposite directions. People expected that higher punishment would lead to a better mood, but in fact higher punishment was associated with a worse mood.

However, forecasters did not predict that they would feel better than people in the no-punishment condition reported feeling, which on the face of it is inconsistent with our hypothesis that people believe punishment improves their affective states. In the absence of a no-punishment-forecaster condition, however, in which people predicted how they would feel without the opportunity to punish, this finding is difficult to interpret. That is, our forecasters were predicting how they would feel in a situation that the nonpunishers did not experience, making the comparison problematic (we return to this issue in Study 2). More informative are punishers' beliefs about how they would feel if they had not been given the opportunity to punish. When asked, all participants in the punishment condition indicated that they would have been less happy than they actually were had they not been allowed to punish,  $M_{\text{actual}} = 4.00$  vs.  $M_{\text{counterfactual}} = 5.02, t(14) = -5.35, p < .001$ .

**Rumination forecasts.** Finally, forecasters also failed to perceive the relationship between punishment and rumination. Although the actual punishers revealed a strong positive correlation,  $r(15) = .40$ , between time spent ruminating about the free rider and the severity of the punishment, forecasters thought the relationship ran in the other direction,  $r(15) = -.19$ . A test of the difference of these two correlations yielded  $z = 1.48, p = .07$ . Thus, people thought that punishing the free rider made them feel better, and perhaps even allowed them to "move on" and forget about the free rider. In reality, the opposite occurred: They con-

tinued to think about the free rider and continued to experience negative affect.

### Discussion

We created a real situation in which people were angry at a free rider and then manipulated whether or not they could retaliate against the free rider by means of financial penalties. The result of the punishment was that both the participants and the free rider lost money, and the punishers were less happy than were those who did not have the opportunity to punish. Ten minutes after the game was over, punishers were thinking about the free rider significantly more, and this increased rumination partially mediated the persistent negative affect. Although the actual affective consequences of punishment were clear, people in the punishment condition did not seem to understand these consequences. In fact, each and every participant in the punishment condition reported that their affect would have been lower had they not been allowed to punish. This is a remarkable finding, given that these participants were at the time experiencing more negative affect as a result of the punishment. Further, forecasters failed to predict the negative affective consequences of revenge, in part because (a) the act of punishment led people to ruminate about the perpetrator more, (b) this rumination prolonged the negative affect, (c) the negative affect prolonged rumination, and (d) people failed to predict the effects of punishment on rumination and affect.

Another reason that people mispredicted the effects of punishment may be that they thought only about the pleasure they anticipated from the free rider's imminent misfortune and failed to anticipate the negative affect that would accrue as a result of instigating the punishment. That is, the negative impact of punishment may stem from the fact that participants had, so to speak, to thrust the knife themselves, which may have violated their sense that they were good people who do not harm others. Further, people had to spend some of their own money in order to punish the free rider. Perhaps people could avoid the unpleasant effects of revenge if they were not the agent responsible for it. If they learned that someone else punished the free rider, they would see her get her comeuppance while avoiding the psychological and financial costs of delivering the punishment themselves.

In Study 2 we tested the hypotheses that people fail to appreciate the distinction between instigating and witnessing a negative event befalling the free rider, namely, that people would expect to feel better after the free rider received a financial punishment but that they would be indifferent to the source of that punishment. Another purpose of Study 2 was to compare directly people's forecasts about how they would feel if they could versus if they could not exact revenge. As mentioned earlier, in Study 1 it was difficult to compare people's forecasts about how they would feel if they could punish with the experiences of people who could not punish in the absence of a condition in which people forecasted how they would feel if they could not punish. In Study 2 we asked people to provide both forecasts to better test the prediction that people believe punishing would improve their affective states.

## Study 2

### Method

**Participants.** Ninety-eight undergraduates participated in partial fulfillment of a course requirement. Seventy-three of the par-

ticipants were from the same university the participants were from in Study 1; 25 were from a different, though academically comparable, university. Seventy-seven participants were women, and 21 were men.

**Materials and procedure.** Participants read a detailed description of the basic paradigm used in Study 1 and were asked to imagine that they were one of the participants. They were told that they and the 2 other cooperative players had invested \$3.87 and received \$1.85 profit at the end of the game. The free rider, however, had only invested \$1.00 yet received \$3.43 profit at the end. (Earnings and investments were based on the actual average results from a pilot version of the free rider game.) Participants then forecasted how they would feel at that point in the game on the same affect measures used in the previous study. These measures are the equivalent of the no-punish control condition of Study 1. Participants then learned about the punishment options and were asked to report how they would feel if they had decided to punish the free rider, and also how they would feel if one of the other players had punished the free rider (yielding a within-subjects design). In one condition they were told, "In the game that you just finished, *you* decide to spend \$0.25 on penalties, a move that will reduce the winnings of [the free rider] by \$0.75." In the other condition they were told that "*one of the other players*" spent the \$0.25. Thus there were three within-subject conditions: a no-punish control condition; a punisher condition, in which the participant imagines instigating the punishment; and a witness condition, in which the participant imagines witnessing another player punishing the free rider. The order of presentation of the punisher and witness conditions was counterbalanced across participants.

In addition to affect measures, participants reported the extent to which they would be thinking about the free rider (i.e., the rumination question from Study 1) after each set of affect measures. At the conclusion of the survey, participants completed manipulation checks regarding details of the game. Participants completed this survey in one of two formats: as a stand-alone web survey or as a paper-and-pencil survey appended to the end of an existing survey.

### Results

**Preliminary analyses.** There were no main effects or interactions involving the order of question presentation, the format of the survey, or university affiliation and so these variables are not discussed further. As before, the positive and negative affect items were highly correlated and produced reliable affect scales. We thus collapsed across valence for each of the three conditions and created unitary affect scales as in Study 1 ( $\alpha = .78, .81, \text{ and } .77$ , respectively).

**Predictions of affect.** Overall, people expected that they would feel quite negative after playing the game in which the free rider reaped profits at their expense. The mean affect score in the control condition (prior to any mention of punishment) was 2.36 (on a scale of 1 to 7) with a standard deviation of .84. However, there was also a strong belief that the opportunity to punish would have salubrious effects on mood,  $F(2, 190) = 98.71, p < .001$ . Participants predicted a mean affect of 4.33 ( $SD = 1.05$ ) if they instigated the punishment and a mean affect of 4.76 ( $SD = .99$ ) if they witnessed someone else instigating the punishment. The no-punishment condition was significantly different from the two punishment conditions ( $ps < .001$ ), and the two punishment conditions did not differ from each other ( $p = .16$ ).

A detailed analysis of each affect item confirms the findings described above. Each negatively valenced item was predicted to be higher in the no-punishment condition and lower in the punishment conditions. Similarly, each positively valenced item was predicted to be lower in the no-punishment condition and higher in the punishment conditions. Each of these differences was significant by Tukey's LSD at  $p = .05$ . The largest mean differences between the punish and no-punish conditions were for the items pleased ( $-2.80$ ), irritated ( $2.49$ ), and negative ( $2.33$ ).

*Predictions of rumination.* In Study 1 we found that punishers ruminated more about the free rider and that this increased rumination partially explained subsequent negative affect. Participants in the present study, like the forecasters in Study 1, did not predict that punishers would ruminate more; in fact, they predicted the opposite. In the control condition, their agreement with the statement "I am thinking about" the free rider was 4.57 on a 7-point scale. When they imagined instigating the punishment themselves, that number dropped to 4.29, and when they imagined another player punishing the free rider, the number dropped to 3.53. All differences were significant by Tukey's LSD ( $p < .001$ ).

### Discussion

As hypothesized, participants predicted that they would feel much better after being given the opportunity to punish than before. This clarifies the results of Study 1, in which we did not include both a no-punish and punish forecaster condition. Second, participants believed that punishment would reduce rumination rather than increase it, replicating Study 1. Third, participants failed to differentiate between witnessing and instigating the punishment. In Study 3 we tested the hypothesis that this distinction is critical for the affect people actually experience, namely, that people feel worse when they deliver the punishment themselves. That is, Study 3 tested the hypothesis that participants' forecasts in Study 2 were wrong, in the sense that having a hand in the punishment (instigating it) has more negative hedonic consequences than witnessing someone else doing the job.

### Study 3

We performed a replication of Study 1 with the addition of conditions in which the punishment (or anticipated punishment) was carried out by one of the other players and the participant was merely a witness to the punishment. This design permitted a direct replication of Study 1, a replication of Study 2 using actual game players rather than vignette-based predictions, and a test of whether the affective forecasting errors found previously were attributable, in part, to a failure to recognize the affective distinction between instigating and witnessing punishment.

### Method

*Participants.* Seventy-five university students from the same population as in Study 1 participated in the experiment for partial course credit. There were 45 women and 30 men.

*Procedure.* The procedure was identical to that in Study 1 with the following exceptions. In the punishment conditions, participants were informed that 1 player had been randomly selected to be allowed to financially penalize any or all of the other players.

In the punisher condition, the participant learned that he or she had been the one selected. This condition mimics the punishment condition of Study 1. In the witness condition, the participant learned that 1 of the other cooperative players had been selected and had chosen to levy a fine on the free rider of \$1.65. Thus, in the first condition the participant instigated the punishment, and in the second condition the participant witnessed the punishment. There were two corresponding forecaster conditions in which participants played the entire game but were asked to predict how they would feel if they had either instigated or witnessed a \$1.65 punishment levied against the free rider. Finally, there was a no-punishment control condition in which participants were never informed about any punishment options.

*Materials.* The materials were identical to those used in Study 1 except where noted. New to this study was a single item at the end of the experiment asking participants to predict how they would have felt had they been in one of the other conditions. Thus, the instigators (both actual and forecasters) were asked, "Imagine that it had been *another player* who did the punishing rather than you. Do you think you would be feeling better, worse, or about the same right now?" Participants in the witness conditions and the no-punish control condition, by contrast, were asked how they would be feeling had they punished the free rider. The question used a 7-point scale ranging from 1 (*worse than now*) to 7 (*better than now*), with a neutral midpoint at 4 (*about the same*).

### Results

Participants earned an average of \$1.89, and the free rider earned \$4.41. Eleven of 15 people in the instigator condition penalized the free rider by spending an average of \$0.43, resulting in an average punishment of \$1.28. Nobody penalized either of the cooperative confederates in the game. As before, we checked to see whether we could collapse the affect ratings across time and valence. The 12-item affect scale had acceptable reliability ( $\alpha = .81$ ), and subsequent tests revealed no interactions with time or valence relevant to the hypotheses. There was a significant main effect for time such that people's mood improved during the 10-min distracter task,  $F(1, 27) = 7.10, p < .05$ , and there was a marginal Valence  $\times$  Condition interaction, revealing that the replication of Study 1 (see below) was somewhat more pronounced for negative affect,  $F(1, 27) = 3.61, p = .05$ .

Replicating Study 1, participants in the punisher condition ( $M = 3.74, SD = .86$ ) experienced significantly worse mood than did the no-punish controls ( $M = 4.58, SD = .58, t(70) = 2.98, p < .01$ ).<sup>1</sup> This replicates the finding that the act of punishing the free rider lowered participants' mood. Also as predicted, there was no affective cost to seeing someone else punish the free rider: Participants who merely witnessed the punishment of the free rider ( $M = 4.33, SD = .75$ ) were in no worse a mood than were the no-punish controls ( $M = 4.58, SD = .58, t(70) = 0.86, p = .39$ ), and were in a significantly better mood than were the punishers,  $t(70) = 2.10, p < .04$ .

As seen in Table 1, forecasters did not predict this affective difference between punishing and witnessing. Those asked how they would feel if they instigated the punishment predicted that they would feel better than the punishers actually felt,  $t(70) =$

<sup>1</sup> All predictions in Study 3 were tested with planned comparisons.

2.58,  $p = .02$ , replicating the finding in Study 1 that people do not anticipate the negative affective consequences of punishment. Also as predicted, there was no significant difference between the forecasts of people in the punisher and witness conditions,  $t(70) = 0.14$ ,  $p = .89$ , suggesting that forecasters did not appreciate the different effects of exacting revenge oneself versus witnessing someone else doing it. A contrast testing the prediction that the mean in the actual punishers condition was lower than the means in the actual witness, forecaster punisher, and forecaster witness conditions was significant,  $t(70) = 2.87$ ,  $p < .01$ , and the residual from this contrast was not significant.<sup>2</sup>

As in Study 1, the means in the forecaster conditions were not significantly higher than the mean in the no-punish control condition,  $t(70) = 0.68$ ,  $p = .50$ . However, when we asked people directly at the end of the study how they would have felt had they been in the other condition, a different picture emerged. The punishers reported that they would feel slightly, though not significantly, worse if they had not been allowed to punish. Their average response was 3.76 ( $SD = 1.26$ ), slightly below the midpoint of the scale ( $4 = \textit{about the same}$ ). Both the witnesses ( $M = 5.0$ ,  $SD = 1.57$ ) and the nonpunishers ( $M = 5.29$ ,  $SD = 1.38$ ), however, reported that they would feel significantly better had they been permitted to punish the free rider. Both groups were higher than the punishers ( $p < .001$ ), as well as the neutral midpoint of the scale,  $t(27) = 3.43$ ,  $p < .01$ ;  $t(13) = 3.48$ ,  $p < .01$ . Thus, not only did these nonpunishers fail to recognize the hedonic cost associated with punishment, they actually predicted hedonic gains from the act of punishment.

*Rumination about the free rider.* As in Study 1, there were no differences between conditions at Time 1 in how much people reported thinking about the free rider,  $F(4, 70) = 0.67$ ,  $p = .61$ . Also replicating Study 1, punishers thought about the free rider at Time 2 more than the nonpunishers did ( $M = 4.2$  vs.  $2.5$ ),  $t(70) = 2.29$ ,  $p = .03$ , and more than punisher forecasters predicted they would ( $M = 2.8$ ),  $t(70) = 1.95$ ,  $p = .06$ . These differences disappear when the amount of punishment is statistically controlled,  $F(2, 39) = 1.59$ ,  $p = .22$ , suggesting that punishment mediates this relationship. And, as before, rumination about the free rider was related to both punishment ( $b = .42$ ) and mood ( $b = -.42$ ) at Time 2, and the direct effect of punishment on mood was reduced from  $b = -.40$  ( $p < .05$ ) to  $-.22$  ( $ns$ ) when rumination was included in the equation. This reduction was marginally significant by the Sobel test ( $z = 1.65$ ,  $p = .10$ ). As predicted, punishers reported thinking about the free rider more than did witnesses, but this difference was not significant,  $t(70) = 0.85$ ,  $p = .40$  (see means in Table 2).

Not only did rumination partially mediate the effect of punishment on affect, but once again affect appeared to partially mediate

Table 1  
Study 3: Mean Participant Affect (With Standard Deviation in Parentheses) by Condition (7-Point Scale)

Condition	Experiencer	Forecaster
Punisher	3.74 <sup>a</sup> (.86)	4.43 <sup>b</sup> (.79)
Witness	4.33 <sup>b</sup> (.75)	4.39 <sup>b</sup> (.73)
No-punish control	4.58 <sup>b</sup> (.58)	

Note. The higher the number, the more positive the reported affect. Means with different superscripts are significantly different,  $p < .05$ .

Table 2  
Study 3: Mean Participant Rumination (With Standard Deviation in Parentheses) by Condition (7-Point Scale)

Condition	Experiencer	Forecaster
Punisher	4.20 <sup>a</sup> (2.18)	2.82 <sup>b</sup> (1.74)
Witness	3.57 (2.21)	3.33 (2.35)
No-punish control	2.50 <sup>b</sup> (1.35)	

Note. The higher the number, the more people reported thinking about the free rider. Means with different superscripts are significantly different,  $p < .06$ . Means without subscripts do not differ from any other means.

the effect of punishment on rumination. Affect was related to punishment ( $b = -.40$ ) and rumination ( $b = -.42$ ) at Time 2, and the direct effect of punishment on rumination was reduced from  $b = .42$  ( $p < .05$ ) to  $.26$  ( $ns$ ) when mood was included in the equation. This reduction was marginally significant by the Sobel test ( $z = 1.61$ ,  $p = .11$ ).

Further, forecasters failed to predict the negative affective consequences of revenge, in part because (a) the act of punishment led people to ruminate about the perpetrator more, (b) this rumination prolonged the negative affect, (c) the negative affect prolonged rumination, and (d) people failed to predict the effects of punishment on rumination and affect.

## Discussion

Study 3 replicated Study 1: People expected the act of punishment to make them feel better, but they were wrong, it actually made them feel worse. As in Study 1, the results suggest that one mechanism by which this occurs is that punishment causes people to think about the offender more, and this in turn maintains their negative affect. We also have evidence that negative affect promotes rumination. Study 3 extended these findings by showing that punishing the free rider themselves made people feel worse than witnessing someone else delivering the punishment, a distinction that forecasters did not anticipate. Thus, the present studies implicate two reasons why people mispredict the affective costs of punishment: They underestimate the extent to which punishment will make them ruminate about the free rider, and they fail to realize that this is especially true if they instigate the punishment, as opposed to seeing someone else do it.

One interesting finding in Study 3 was that people in the witness condition did not report feeling better than people in the no-punishment control condition. That is, seeing someone else punish the free rider did not have the affective *cost* of punishing her oneself (as we predicted), but neither did it confer an affective *benefit*. On the face of it this finding is inconsistent with the literature on *schadenfreude*, the idea that people take pleasure in the misfortune of others (Hareli & Weiner, 2002; Leach, Spears,

<sup>2</sup> The results on the individual affect items were consistent with those reported here. Each negatively valenced item was higher for the punishers than for people in the other conditions. Similarly, each positively valenced item was lower for the punishers. The effect was more pronounced for the negative affect items negative (1.52), vengeful (1.45), and irritated (0.59). Thus, as in Study 1, it is primarily an increase in negative affect that results from administering punishment.

Branscombe, & Doosje, 2003; Smith et al., 1996). One difference between the present experiments and work on *schadenfreude* is that our participants had been directly harmed by the free rider, as opposed to being passive bystanders to an event. In the Smith et al. (1996) studies, for example, participants read a news account of a target individual who failed to be accepted to medical school, and in the Leach et al. (2003) studies, participants learned that a rival soccer team lost to the home team. In both cases the misfortune was quite distant from the participant. It may be that when people are directly harmed by a transgressor and experience anger, the pleasure of seeing that person suffer is mitigated by the rumination effects we found in the current studies. That is, learning that someone else punished the free rider may have some affective benefit, but it may also keep one's attention on the transgression, making it difficult to move on and think about something else. The important point for our purposes is that people who delivered the punishment themselves ruminated the most and experienced the most negative affect, contrary to people's forecasts.

### General Discussion

The present studies addressed three questions: (a) What is the affective consequence of exacting revenge? (b) Do people accurately predict this consequence, and, if not, (c) Why do people make this error of prediction? In answer to the first question, Studies 1 and 3 found that most people given the opportunity to punish a free rider did so, but they ended up feeling worse than did people who did not have the opportunity to punish. In answer to the second question, Studies 1–3 found that people did not accurately predict the negative affective consequences of exacting revenge; in fact, they predicted that it would make them feel better. In answer to the third question, Studies 1–3 found that people believed that exacting revenge would bring closure, in the sense that they would think less about the free rider, when in fact it had the opposite effect—punishing the free rider made people think about her more, which in turn made them feel worse. Further, Study 3 found that the punishment → rumination → negative affect sequence occurred only when people punished the free rider themselves and not when they witnessed someone else delivering the punishment. And yet, forecasters did not anticipate that delivering the punishment themselves would make them feel worse than seeing someone else do it.

#### *Affective Forecasting*

The present results add to the growing body of research showing that forecasters do not simply exaggerate the impact of an emotional event, they sometimes misunderstand which of two sets of conditions will make them the happiest (Gilbert & Ebert, 2002; Wilson, Centerbar, Kermer, & Gilbert, 2005). This finding is particularly striking in the domain of anger, because across human history there have been countless acts of revenge, many motivated by the belief that “to witness suffering does one good, to inflict it even more so” (Nietzsche, 1887/1996, p. 48)—an affective forecast that the present studies show is wrong, at least under some circumstances.

The results also suggest an exception to the focalism hypothesis (Wilson et al., 2000), which argues that people typically overestimate how much they will think about a future emotional event. We found the opposite effect in the present studies, in that people in the pun-

ishment conditions underestimated how much they would think about the free rider. Although it will take further research to identify the precise conditions under which people under- or overestimate how much they will think about an event, we suspect that the present results are specific to forecasts about anger, which is an emotion that makes it difficult to think about anything else (Fredrickson & Branigan, 2005). Anger led to punishment which led to increased rumination about the free rider, and this rumination may have prevented people from engaging other strategies (e.g., trivialization) that would have allowed them to move on and think about something else. Participants failed to anticipate these effects, possibly due to their lay theories that punishment would improve their mood (e.g., Bushman et al., 2001).

It is beyond the scope of this article to identify why these theories are wrong, especially when most people have experience in situations like the ones created in our experiments. Many people have retaliated against a driver who was rude to them, for example, or in some other way punished someone who angered them. Why have they not learned that such retaliation can make them feel worse? One reason may be that seeing a transgressor receive his or her comeuppance is pleasurable under some conditions, as found in the literature on *schadenfreude*. People may overgeneralize from such situations, failing to appreciate the conditions (such as those identified in the present studies) that turn punishment into an affectively costly action.

#### *The Neural Basis of Punishment*

Our findings are particularly interesting in light of recent reports of the neural bases of punishment. De Quervain and colleagues (2004) found that the dorsal striatum, the brain region that is closely related to pleasure, was activated when participants decided to punish a fellow participant who had violated a social norm. In other words, that punishment was associated with positive affect, not negative affect. These measurements, however, were taken during the 1-min period prior to making the punishment decision, precisely the same time frame represented by the forecasters in the present experiment. In both studies, then, people showed clear expectations that the punishment would lead to pleasure. The present studies, however, also measured hedonic reactions after the punishment was conducted and found that punishers were substantially less happy than the nonpunishers after both a 1- and 10-min delay. Thus, our results corroborate de Quervain et al.'s findings but add the crucial finding that predicted and experienced emotion are substantially different.

#### *Motives for Punishment*

Our findings support a functional account of punishment—people use punishment to strategically repair their negative mood. Clearly, though, this is not the only function of punishment, and it is left to future research to integrate the different explanations of punishment and to delineate the circumstances under which one particular motive or function becomes dominant.

Carlsmith and his colleagues (Carlsmith, 2006; Carlsmith & Darley, 2008; Carlsmith, Darley, & Robinson, 2002; Carlsmith, Monahan, & Evans, 2007) have shown that citizens' intuitive theories of punishment are consistent with Immanuel Kant's theory of retributive punishment. When people punish, they seek to



give a perpetrator what he or she deserves and to make the punishment proportional to the offense. The function of the punishment in these experiments appears to be one of satisfying the “justice motive” rather than a more instrumental goal. Specifically, Carlsmith found little support for the hypothesis that punishers seek to use punishment for the utilitarian goals of deterrence (Carlsmith, 2006; Carlsmith et al., 2002) or incapacitation (Darley, Carlsmith, & Robinson, 2000).

The critical difference between studies that have found support for retributive punishment and the current hedonic regulation accounts appears to be the participants’ level of involvement. Specifically, the justice motive is dominant in third-party observations of injustice—such as one would experience as a juror. By contrast, when the respondent has been directly harmed by the perpetrator (as in the present experiments), additional motives may be activated, including the desire to repair one’s mood.

In sum, our results suggest that, in some instances at least, revenge has hedonic consequences that are the opposite of what people expect. Revenge can prolong peoples’ hedonic reactions to a transgression because punishing others can cause people to continue to think about (rather than to forget) those whom they have punished. As Sir Frances Bacon noted more than three centuries ago, “A man that studieth revenge, keeps his own wounds green, which otherwise would heal, and do well” (Spedding, Ellis, & Heath, 1858).

## References

- Baier, K. (1955). Is punishment retributive? *Analysis*, *16*, 25–32.
- Barnstead-Klos, A. (2007, May 6). Re: Virginia Tech Massacre thoughts. Message posted to <http://tinyurl.com/5fy2n2>
- Bentham, J. (1962). Principles of penal law. In J. Bowring (Ed.), *The works of Jeremy Bentham* (p. 396). Edinburgh, Scotland: W. Tait. (Original work published 1843)
- Berkowitz, L. (1989). Frustration-aggression hypothesis: Examination and reformulation. *Psychological Bulletin*, *106*, 59–73.
- Bushman, B. J. (2002). Does venting anger feed or extinguish the flame? Catharsis, rumination, distraction, anger, and aggressive responding. *Personality and Social Psychology Bulletin*, *28*, 724–731.
- Bushman, B. J., Baumeister, R. F., & Phillips, C. M. (2001). Do people aggress to improve their mood? Catharsis beliefs, affect regulation opportunity, and aggressive responding. *Journal of Personality and Social Psychology*, *81*, 17–32.
- Carlsmith, K. M. (2006). The roles of retribution and utility in determining punishment. *Journal of Experimental Social Psychology*, *42*, 437–451.
- Carlsmith, K. M., & Darley, J. M. (2008). Psychological aspects of retributive justice. In M. P. Zanna (Ed.), *Advances in experimental social psychology*, (Vol. 40, pp. 193–236). San Diego, CA: Elsevier.
- Carlsmith, K. M., Darley, J. M., & Robinson, P. H. (2002). Why do we punish? Deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology*, *83*, 284–299.
- Carlsmith, K. M., Monahan, J., & Evans, A. (2007). The function of punishment in the “civil” commitment of sexually violent predators. *Behavioral Sciences & the Law*, *25*, 437–448.
- Darley, J. M., Carlsmith, K. M., & Robinson, P. H. (2000). Incapacitation and just deserts as motives for punishment. *Law and Human Behavior*, *24*, 659–683.
- De Quervain, D. J.-F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., & Fehr, E. (2004). The neural basis of altruistic punishment. *Science*, *305*, 1254–1258.
- Dollard, J., Doob, L., Miller, N., Mowrer, O. H., & Sears, R. (1939). *Frustration and aggression*. New Haven, CT: Yale University Press.
- Ezorsky, G. (1972). *Philosophical perspectives on punishment*. Albany, NY: SUNY Press.
- Fredrickson, B. L., & Branigan, C. (2005). Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition and Emotion*, *19*, 313–332.
- Freud, S. (1961). *Civilization and its discontents*. New York: Norton.
- Frijda, N. H. (1994). The lex talionis: On vengeance. In S. H. M. van Goozen, N. E. Van De Poll, & J. A. Sergeant (Eds.), *Emotions: Essays on emotion theory* (pp. 263–289). Hillsdale, NJ: Erlbaum.
- Gilbert, D. T., & Ebert, J. E. J. (2002). Decisions and revisions: The affective forecasting of changeable outcomes. *Journal of Personality and Social Psychology*, *82*, 503–514.
- Gintis, H. (2008). Punishment and cooperation. *Science*, *319*, 1345–1346.
- Hareli, S., & Weiner, B. (2002). Dislike and envy as antecedents of pleasure at another’s misfortune. *Motivation and Emotion*, *26*, 257–277.
- Hobbes, T. (1950). *Leviathan*. New York: E. P. Dutton.
- Kant, I. (1952). The science of right (W. Hastie, Trans.). In R. Hutchins (Ed.), *Great books of the Western world* (pp. 397–446). Edinburgh, Scotland: T. & T. Clark. (Original work published 1790)
- Konecni, V. (1974). Self-arousal, dissipation of anger, and aggression. *Personality and Social Psychology Bulletin*, *1*, 192–194.
- Leach, C. W., Spears, R., Branscombe, N. R., & Doosje, B. (2003). Malicious pleasure: Schadenfreude at the suffering of another group. *Journal of Personality and Social Psychology*, *84*, 932–943.
- Lee, J. (1993). *Facing the fire: Experiencing and expressing anger appropriately*. New York: Bantam Books.
- Loewenstein, G., O’Donoghue, T., & Rabin, M. (2003). Projection bias in predicting future utility. *Quarterly Journal of Economics*, *118*, 1209–1248.
- McKean, E. (Ed.). (2005). *The new Oxford American dictionary* (2nd ed.). Oxford, England: Oxford University Press.
- Mellers, B. A., & McGraw, A. P. (2001). Anticipated emotions as guides to choice. *Current Directions in Psychological Science*, *10*, 210–214.
- Nietzsche, F. (1996). *On the genealogy of morality* (D. Smith, Trans.). Oxford, England: Oxford University Press. (Original work published 1887)
- Quinton, A. M. (1954). On punishment. *Analysis*, *14*, 512–517.
- Rusting, C. L., & Nolen-Hoeksema, S. (1998). Regulating responses to anger: Effects of rumination and distraction on angry mood. *Journal of Personality and Social Psychology*, *74*, 790–803.
- Schkade, D. A., & Kahneman, D. (1998). Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychological Science*, *9*, 340–346.
- Simon, L., Greenberg, J., & Brehm, J. (1995). Trivialization: The forgotten mode of dissonance reduction. *Journal of Personality and Social Psychology*, *68*, 247–260.
- Skinner, B. F. (1938). *The behavior of organisms*. New York: Appleton-Century-Crofts.
- Skinner, B. F. (1974). *About behaviorism*. New York: Knopf.
- Smith, R. H., Turner, T. J., Garonzik, R., Leach, C. W., Urch-Druskat, V., & Weston, C. M. (1996). Envy and Schadenfreude. *Personality and Social Psychology Bulletin*, *22*, 158–168.
- Spedding, J., Ellis, R. L., & Heath, D. D. (Eds.). (1858). *The works of Francis Bacon* (Vol. 6). London: Longman.
- Watson, J. B. (1924). *Behaviorism*. New York: Norton.
- Wilson, T. D., Centerbar, D. B., Kermer, D. A., & Gilbert, D. T. (2005). The pleasures of uncertainty: Prolonging positive moods in ways people do not anticipate. *Journal of Personality and Social Psychology*, *88*, 5–21.
- Wilson, T. D., & Gilbert, D. T. (2003). Affective forecasting. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 35, pp. 345–411). San Diego, CA: Elsevier.
- Wilson, T. D., Wheatley, T. P., Meyers, J., Gilbert, D. T., & Axsom, D. (2000). Focalism: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology*, *78*, 821–836.

Received July 10, 2006

Revision received March 10, 2008

Accepted March 12, 2008 ■