In-kind vs. out-of-kind penalties: preference and valuation

Jane Beattie
Laboratory of Experimental Psychology
University of Sussex
Brighton, BN1 9QG
Sussex, England
beattie@epunix.susx.ac.uk

Jonathan Baron
Department of Psychology
University of Pennsylvania
Philadelphia, PA 19104
USA
baron@cattell.psych.upenn.edu

September 20, 2003

Running head: In-kind penalties

Abstract

Utilitarian and economic theories of deterrence hold that the relation of the penalty to the misdeed should be irrelevant. In five experiments using hypothetical cases, judgments of penalties depended on whether this relation was in-kind (IK) or out-of-kind (OK). When victims were identifiable, IK penalties were higher than OK and preferred to OK. Subjects seemed to confuse penalties and compensation. When there were no identifiable victims (e.g., environmental damage), IK penalties were preferred, but OK penalties were higher. Here, OK judgments were more uncertain, and subjects preferred to err on the side of overpunishing. The results can be explained in terms of overgeneralization of usually-useful heuristics. Alternative hypotheses concerning aggression and scale compatibility are rejected. The findings have implications for the setting of penalties in legal cases (e.g., the Exxon Valdez) and for lay theories of punishment in non-legal settings.

In-kind vs. out-of-kind penalties: preference and valuation

> *But if there is serious injury, you are to take life for life, Eye for eye, tooth for tooth, hand for hand, foot for foot, Burn for burn, wound for wound, bruise for bruise.* (Exodus 21, 23-25, New International Version)

Utilitarian and economic theories of punishment and penalties imply that the expected utility of a penalty should be just great enough to deter the commission of harm, presumably just a bit greater than the expected utility gain that results from committing the harm (Bentham, 1948; Posner, 1986). In this way the expected utility of a harmful act will be, on balance, negative, and people will be sufficiently deterred from harming others. Excessive punishment, beyond this point, can only reduce overall utility.

These theories assume that utility sources are interchangeable, i.e., the outcomes giving rise to the utility (or wealth) are irrelevant beyond the magnitude of utility that they contribute. Thus the sources of the expected gains (benefits to the injurer) and expected losses (penalties) are allegedly irrelevant to the suitability of the penalty to the harmful act committed. Yet folk conceptions of justice often include the notion that somehow the penalty should "fit" the crime, as suggested by the Biblical adage quoted above. We present a series of experiments investigating whether people view utility as interchangeable, or whether the type of harm influences the preferred type or magnitude of penalty (when utility levels are held constant).

Such conceptions of justice may affect the outcome of public issues as well as private disputes. When the Exxon Valdez spilled oil in Prince William Sound, Alaska, the Exxon company responded, in part, by trying to undo the spill, often using extremely inefficient means, such as wiping oil from individual rocks by hand. Some actions may even have caused more harm than good, such as blasting oiled beaches with boiling water (Rauber, 1992). Perhaps both Exxon officials and the public saw this as a more appropriate (self-)punishment than spending the same amount of money on other projects that might have

done more good. Ultimately, Exxon agreed to pay $900,000 into a trust fund administered by appointees of the State of Alaska and the U.S. government, for "restoring, replacing, enhancing, rehabilitating or acquiring the equivalent of natural resources injured" by the spill. Initial proposals for the fund focused on development of tourism in Prince William Sound and research on the effects of the spill. As it has become apparent that better uses of the money could be found, the Sierra Club and others successfully advocated spending some of the money on other environmental projects, such as buying Alaskan forests to prevent clear-cutting (Rauber, 1992; Kenworthy, 1994). Such uses of the money are "equivalent" in kind as well as in value. Courts have recognized the concept of "indirect environmental restitution" in which "paid into public or private trust funds used to clean up past or future contaminated sites, or even to acquire environmentally sensitive areas for preservation" (Zornow & Reed, 1992).

Similar attempts to punish people "in kind" are seen in the criminal justice system. For example, Judge Ted Poe sentenced three young men who stole antique comic books (including Batman #1) to work at the public library and read books to children (Taylor, 1991). Likewise, when one country harms another through trade restrictions or diplomatic expulsions, the preferred response is to retaliate in kind, even though this response harms *both* countries. In private disputes, the same attitude may influence the punishments imposed on naughty children or unfaithful lovers. We seem to favor in-kind penalties, despite the theoretical claim that the deterrent effect does not depend on this, so long as other things are equal.

The goal of the research reported in this paper was to investigate lay beliefs about the appropriate type and level of penalties. In particular we wished to investigate whether compatibility between the harm and the penalty will affect these judgments. All the examples in our experiments were hypothetical legal cases, inspired loosely by the Exxon Valdez case, although our main interest was in penalties in general, rather than in legal

settings. We used legal cases because they allowed us to manipulate the factors of interest in ways that were plausible to our subjects, although in many cases they would not be plausible to experts in Anglo-American law. In addition, legal judgments are a major application of the principles behind our results.

We asked subjects two kinds of questions concerning the relation between penalty and harm. First we ask a simple *preference* question: do subjects prefer that penalties be related to harms ("in kind", abbreviated IK) or unrelated ("out of kind", abbreviated OK). For example, if a company damages a beach, do subjects prefer that the company pay for the restoration of similar beach elsewhere or for the otherwise equivalent restoration of forest? Second, in the *valuation* question, we ask subjects to judge how much of each type of penalty, in kind (IK) or out of kind (OK), is appropriate: i.e., how much beach if the penalty is beach restoration only, and how much forest if the penalty is forest restoration only. The utilitarian theory acts as the null hypothesis, from which we look for systematic departures. It holds that the choice and valuation of the penalty should depend only on the magnitude of damage, not on the kind of damage.

We begin with the hypothesis that subjects will not adhere to the utilitarian model stating that harm/penalty compatibility should not affect preference and valuation of penalties. Instead, we hypothesize that subjects' responses in our tasks will be determined by use of heuristics which often have utilitarian justification, but which have been applied to situations for which the justification is no longer present. For example, IK penalties are usually easier to decide on than OK penalties when the penalty also compensates the victim, so people will favor these penalties even when compensation is outside of the decision. Such overuse of heuristics has been demonstrated in many cases (for a review, see Baron, 1994). For example, Baron & Ritov (1993) found that subjects (including retired judges) assigned penalties without regard to their deterrent effect (which was presumably at least part of the original consequentialist justification for the penalty). Use of such

heuristics by those who decide on penalties will tend to make penalties not conform to the utilitarian model.

We considered two types of situation in which people might use such heuristics in evaluating penalties. First, and probably most frequently, are cases in which an identifiable victim exists and in which compensation and damages are coextensive: a single fine (penalty) may be imposed which is then given to the victim in compensation. When an injurer compensates a victim directly, restitution in kind (IK) is best because it is likely to be most efficient for maximizing utility: if the victim is subject to loss aversion, more compensation will be required for OK than for IK restitution, assuming that victims will integrate IK compensation (but not OK compensation) with the loss sustained (in the sense of Thaler, 1985). In addition, by reducing ambiguity about the appropriate penalty, estimation of appropriate penalties is easier, and conflict between the parties is minimized. For example, if someone dents your car, she must pay the amount that is necessary to fix the damage. It is convenient for the penalty and damage to be expressed in the same currency (money) because in this way we can be sure that she has exactly compensated you for the damage. If she were to pay you in terms of time (e.g., by cleaning your house for some specified period of time), it would be unclear whether she had compensated you too little or too much. Each of you might distort your judgments in his or her favor, and you would disagree (see Thompson & Loewenstein, 1992). Our heuristics hypothesis predicts that subjects will favor IK penalties even when these justifications are absent, in particular, when compensation is decided independently from penalties.

To examine intuitions about penalties independent of compensation, we constructed cases in which there is an identifiable victim, but the victim has already been compensated, as often happens when victims are compensated by their own insurance. In order to examine the deterrent effect only, we told the subject that the penalty would not be given to the victim (as done by Baron & Ritov, 1993). Although this is unrealistic in many legal

systems, New Zealand compensates most victims through social insurance but still allows *criminal* prosecution of injurers (thus relegating deterrence to the criminal law; Brown, 1985; Mastromatteo, 1993). Here, we propose that subjects will not necessarily judge that IK penalties should equal OK penalties, because they will think about penalties as if they were thinking about compensation of victims. What we shall call our *heuristics hypothesis* predicts that subjects want the IK award (penalty) to be greater because they would want this if the IK award will help the injured party and the OK award will not.

For example, suppose that a drug company has caused epilepsy to some children through neglecting to follow safety procedures. Compensation has already been given, but the drug company is ordered to contribute to the building of a school for (other) children. In one condition the children to attend the school also have epilepsy (IK condition), and in another condition they are blind children (OK condition). Our compensation-heuristic hypothesis suggests that subjects will *prefer* the award to go to the school for epileptic children (IK). They may also decide on a *greater* penalty if the school is to be built for epileptic (IK) than blind (OK) children. To summarize, our hypothesis states that, in cases where there are clear victims, subjects will tend to both *prefer* IK judgments and set *higher* IK penalties than OK penalties. These predictions arise from the subject using heuristics based on compensation, even when compensation is not at issue.[1]

The second part of our heuristics hypothesis concerns cases in which there are no clear victims to be compensated (e.g., pollution damage to remote sites of environmental significance). In this case a compensation heuristic would not be evoked. Here, in the preference task, we hypothesize that subjects will still apply the general preference heuristic that IK penalties are better than OK penalties. This heuristic is easily justified as a general rule, and this is why it is used. But it is more difficult to justify in the cases we present. In the valuation task, there is more uncertainty in matching the OK penalty to the damage than in matching the IK penalty, and subjects may prefer to err on the side of larger penalties.

Hence, subjects would assign more OK penalty than IK penalty.[2] Note that this heuristic could also be used when a victim is present, but it may be overwhelmed by the opposing heuristic described previously. To summarize, we predict a shift from setting higher IK penalties when there are victims, to setting higher OK penalties when there are not. The predictions of the hypotheses we will consider in this paper are summarized in Table 1.

(Insert Table 1 near here)

The experiments below test whether subjects' evaluations of penalties follow the utilitarian model, or whether they are guided by the hypothesized heuristics. However, two other psychological theories appear to make alternative predictions for these tasks: Tversky, Sattath & Slovic's (1988) scale compatibility theory; and Foa's (1971) aggression theory. While neither theory was designed to explain penalty judgments *per se*, each can be extended to make a prediction of how subjects might evaluate penalties. We now briefly review the predictions of these two theories (also summarized in Table 1).

Tversky et al. (1988) showed that subjects evaluating multiattribute stimuli tend to overweigh stimulus attributes that are expressed in the same scale as the response mode (scale compatibility). For example, subjects were asked to evaluate gambles that varied in probability of winning and amount to be won. Differences in amount to win were more important, relative to differences in probability, when judgments were expressed as willingness to pay (compatible with amount to win), than when they are expressed as desirability ratings. Tversky et al. provide two reasons why the weight of the compatible attribute might be higher in the decision. First, the compatible attribute is thought to be more salient and hence receive more attention (leading to greater decision weight). Second, evaluating the incompatible attribute is assumed to be more effortful and errorful. Hence it is deemphasized by the subject.

The analogy between Tversky et al.'s experiments and our experiments on judgments of appropriate penalties is a loose one, because our evaluation tasks do not explicitly mention

two attributes (harm that is in-kind or harm that is out of kind relative to the penalty): penalties are unidimensional, with the second dimension value assumed to be zero in each case. However, if we assume that subjects perceive the two dimensions implicitly and that more attention to the harm dimension related to the valuation response leads to larger judged penalties, then we would expect *more* penalty when the penalty is IK (regardless of whether there are victims). We would also expect preference for the IK penalty (as this will receive more weight in the decision).

An alternative view was offered by U. Foa (1971), based on E. Foa's (1970) study of preferred resource allocation in aggression. She found that subjects who suffered an aggressive act preferred to seek retribution (punishment) of a similar kind. For example, if an experimental stooge had criticized the subject's performance, the subject later tended to retaliate by removing social status from the stooge, rather than by hiding information from him or her. The reverse was found if the stooge had withheld information. Thus this hypothesis again predicts that subjects will prefer IK penalties, although for a different reason from either the heuristics or scale compatibility hypotheses.

For valuation tasks, Foa (1971) proposed that "the intensity of the response will be higher, the larger the distance between the preferred and available resource." This hypothesis, the opposite of the scale compatibility hypothesis, suggests that incompatible penalties will be larger (more intense) than compatible penalties. Foa's justification for this hypothesis was that the penalty is less and less satisfactory (or valuable), the farther it is in kind from the original harm. For example, a person who has love removed by another (e.g., in a divorce) might most like to have the love returned (IK compensation), or to remove love from their ex-spouse (IK penalty). If these avenues are not open, the person may be left with only financial compensation/penalty (OK) in the divorce settlement. As this type of good cannot satisfy his or her desire for love, the person demands excessive amounts of money in settlement, to both hurt the ex-spouse (OK penalty) and

to try (unsuccessfully) to satisfy the need for love (OK compensation).

The five studies presented below investigated harm-penalty compatibility effects in both choice and valuation tasks, and also elicited justifications for subjects' responses. The design of the first three experiments was similar: we presented questionnaires describing two kinds of harms (each expressed on a different scale), and two kinds of penalty (each expressed on the same scale as one of the harms). For example, the first study described environmental damage to either a beach or a forest, and it described penalties that either clean up a beach or a forest. The harms and penalties were presented in factorial combination to yield four scenarios. Two scenarios were IK (i.e., the harm and penalty were on the same scale, such as cleaning up a beach when a beach is damaged), and the other two were OK (e.g., cleaning up a beach when a forest was damaged).

For all penalties, it was stated clearly that the action (e.g., cleaning the beach) will be taken anyway by the government and that the cost of the two penalties is the same. Thus there is no reason based on consequences for the subjects to prefer the IK penalty. Subjects were asked to either choose between the penalties (IK or OK) for a given harm (Experiments 1, 2, and 4), or to set amounts for each penalty (e.g., how many miles of beach should be cleaned), given a particular harm (Experiments 3–5). In this way we can examine, for preference and valuations, whether the kind of harm affects the value of the penalty. Experiments 1–4 made up a 2x2 factorial design in which we crossed (i) presence/absence of victims, with (ii) valuation/preference task, as shown below (see Table 2). This allowed us to pit the predictions of the various models against one another (see Table 1).

(Insert Table 2 near here.)

In Experiments 4 and 5 we also examine the effect of errors in judgment on preferences and valuations for IK and OK penalties.

We often use the term "punitive damages" in the descriptions of our scenarios, because

we thought that this term would be clear to naive subjects, even though no law may justify such payment in our cases.

Experiment 1: Forest/Beach Study (Choices)

Method:

Subjects: Subjects were 36 students at University of Pennsylvania, recruited by advertisements around the campus. All were paid.

Procedure: Subjects completed the following questionnaire. The order of harms and penalties were each counterbalanced (4 orders) in a fully within-subject design.

Questionnaire.

> Harm 1: (forest)
> A fire occurs in a chemical factory near a forest. The fire spreads to the forest. It destroys 100 square miles of old-growth forest. Investigation into the cause of the fire finds that the company failed to follow the rules for safety. The government must punish the company for violating these rules.
>
> Consider the following punishments:
>
> Punishment 1: (forest)
> The company must pay the cost of planting 100 square miles of new trees in a different state. The trees were going to be planted anyway as part of a government program to restore forests. The company will pay for the trees and the government will save the money.
>
> Punishment 2: (beach)
> The company must purchase 100 miles of beach land in an unpopulated area so that the beach will become a national park and will not be developed. The government was going to purchase the land anyway for a park. The company will pay for the land and the government will save the money.
>
> The cost of the two punishments is the same. Which punishment should be used? Or does it matter? Explain .... [Subjects gave written justifications for their choice.]
>
> [Subjects then repeated the procedure with the following scenario.]
>
> Harm 2: (beach)
> A chemical company spills chemicals into the ocean. The spill contaminates 100 miles of beach in an unpopulated area. The beach cannot be used for

> recreation for 10 years. Investigation in to the cause of the spill finds that the company failed to follow rules for safety. The government must punish the company for violating these rules.

Results

(Insert Table 3 near here)

The results of Experiment 1 are shown in Table 3. There was a general preference for "forest" penalties (32 responses) over "beach" penalties (17 responses) (sign test, $p = .046$, two-tailed). This is not of relevance to the experimental hypothesis: it merely indicates that our subjects tend to value forests more highly than beaches. The issue of interest is whether subjects tend to prefer IK penalties (as suggested by our heuristics hypothesis, the compatibility hypothesis, and Foa's aggression hypothesis), or are indifferent (supporting the normative model that utility is totally substitutable). Subjects preferred penalties to be IK; that is, the forest penalty was more preferred when the forest was damaged than when the beach was damaged (U test, $p = .029$, two tailed). Justifications pointed strongly to the desirability of the compatibility of harm and penalties as driving preferences. Consequentialist justifications were notably lacking.

> "I think that the punishment SHOULD BE RELATED to the crime." (Subject's emphasis)
> "They ruined the beach, so they should provide a new beach ...."
> "It really does not matter ... but if I had to choose I would choose the ... beach. It seems like a more appropriate matching between the crime and punishment."

The final subject is particularly interesting as s/he (and other similar subjects) recognizes the economic/utilitarian argument that "it really doesn't matter," but still feels that the IK penalty is better. Such subjects were counted (for the purpose of Table 3) as viewing the penalties as equally desirable, suggesting that the preference for IK penalties is even stronger than that we recorded.

Discussion of Experiment 1

The results of Experiment 1 are inconsistent with the normative model that utility from different sources is interchangeable. People prefer the harm and penalty to be IK, even if some recognize that, in some sense, this should make no difference. These results are consistent with the heuristic, aggression and compatibility hypotheses. However, these three theories are based on different motivations, so subjects' justifications should allow us to distinguish between them. In Experiment 2 we have a sufficiently large sample to explore in detail the types of justifications given and their relative proportions.

Experiment 2: Blind/Epileptic Study (Choices)

Experiment 2 examined the generality of the IK preference by moving to a new scenario: one which involves damage by drugs to children (victims), rather than victimless environmental damage. The structure of the questionnaire is similar to that used in Experiment 1: two types of damage (causing epilepsy or blindness) are crossed with two types of penalty (contributing to the cost of a school for (a) blind or (b) epileptic children). As before, the design yields two IK and two OK scenarios, and we can evaluate whether there is a general preference for IK penalties as in Experiment 1. Again, we remove any consequentialist reason for subjects to prefer one penalty over the other: the schools are to be built anyway, and the children who were damaged will be unable to attend the school, as it is to be built in a different state. All the hypotheses except the normative model predict that the subjects will prefer the IK penalty, although for different reasons. In this experiment we collect a sufficiently large number of observations to support quantitative analysis of the justifications.

Method

Subjects: Subjects were 150 students at University of Pennsylvania, solicited and paid as before.

Procedure: Subjects completed the following questionnaire. The order of damage and penalty pairings were counterbalanced (4 orders) in a within-subject design.

Questionnaire.

> Damage 1: (blind)
> A drug company manufactured a drug which caused irreversible blindness in 10 children. Investigations showed that the company did not carry out proper safety tests before marketing the drug. The government must punish the drug company for these actions.
>
> Damage 2: (epilepsy)
> As problem 1, but the drug caused severe epilepsy in 10 children.
>
> Consider the following punishments:
>
> 1) The drug company must pay part of the cost of building a new residential school for blind children in another state. The school was to be built anyway. The company will pay for part of the building and the government will save money.
>
> 2) As above, but the school was for epileptic children in another state.
>
> The cost of the two punishments to the company is the same. Both programs come from the same government budget.
>
> Which program should be used? Or does it matter? Explain....

Results

We classified subjects' responses to the first question that they answered, based on their preferred penalty given the damage, and also based on their justifications. All but seven subjects gave the same type of response for both types of damage (e.g., if they preferred the IK penalty for the epilepsy damage, they also preferred it for the blindness damage). We divided subjects' responses to the first question that they answered into four mutually-exclusive and exhaustive categories. Inter-rater reliability, based on a randomly selected sample of 22 subjects' responses, was 100%.

Unqualified preference for IK. 68 subjects.
Definition: Subject merely asserts that the IK penalty is better. No economic/utilitarian analysis of consequences.

"Some element of the punishment fitting the crime must be included." "Effect of drugs on some children caused epilepsy. Therefore company should compensate those they INDIRECTLY affected." (Our emphasis).

Qualified preference for IK. 38 subjects.
Definition: Subject acknowledges that there is no objective reason to prefer one penalty to the other, but expresses a preference for the IK penalty.

"The [IK] punishment because the company hurt these children. . . . it doesn't really matter, if both schools are to be built . . . ."

No preference. 39 subjects.
Definition: Subject asserts that it makes no difference which punishment is chosen. May note that there is no consequentialist reason to prefer either penalty. Gives no indication that one penalty is better for any reason. May say that others might prefer IK.

"Since the aim is to punish the company and the cost of the two schools is the same, it does not really make any difference . . . ."

Other. 5 subjects.
Definition: Response does not fit into above three categories, e.g., explains why one of the two penalties is always better regardless of the harm, or why the company should be severely punished.

To summarize, 106 (= 68 + 38) of the 150 subjects expressed some preference for type of penalty. All 106 preferred the IK penalty on both scenarios (sign test, $p < .0005$). Experiment 2 thus replicated the results of Experiment 1 in showing that subjects overwhelmingly prefer IK penalties. Again this conflicts with the normative model that utility from different sources is interchangeable.

We were interested in the reasons that subjects gave for their preferences, as this may distinguish between the remaining experimental hypotheses. As in Experiment 1, the overwhelming majority of responses were non-consequential, expressing merely an intuition that the "punishment should fit the crime." This pattern of results strongly supports the heuristics hypothesis – which holds that peole apply principles without thinking of

their consequentialist justifications – rather than the remaining two hypotheses. We also replicated Experiment 2 with a sample of British undergraduates ($N = 23$), as Experiment 3 follows up from Experiment 2 and uses British subjects. The stimuli were slightly altered to describe damage to children in Scotland and a school to be built (far away) in the South of England. The results were substantively similar to the US sample with 12 subjects showing unqualified preference for the IK penalty; 3 showing qualified preference; 4 showing no preference; and 4 were "other." Thus all 15 subjects who expressed a preference favored the IK penalty (sign test, $p < .0005$). Justifications were similar to the US sample.

Experiment 3: Blind/Epileptic Study (Valuation)

Our first two experiments examined subjects' preferences in scenarios with and without victims. We found preference reversals based on the compatibility of the harm and the penalty. In Experiments 3 and 4 we move from a choice task to a valuation task: after all, anyone who decides on a penalty must determine its magnitude, and that is a valuation task ("what should the level of payment be") rather than a choice between penalties of different kinds. The valuation task allows us to more clearly tease apart the predictions of the three experimental hypotheses. Recall that the heuristics hypothesis predicts a higher valuation for IK penalties when there are victims (as in Experiment 3), but lower IK penalties when there are no victims (as will be investigated in Experiment 4).

The remaining two hypotheses make opposing predictions to each other, but these predictions do not depend on the presence or absence of victims. The compatibility hypothesis proposes that higher weight will be assigned to the IK attribute. This suggests that more compensation will be required when the damage is compatible with the penalty, and that valuations should be higher for IK stimuli. The opposite prediction is made by Foa's aggression hypothesis, which states that the intensity of the response (i.e., size of

penalty) will be greater for the less preferred (OK) resource.

In this study we use similar scenarios to Experiment 2. Subjects were asked to consider a single damage (blindness/epilepsy) paired with a single penalty (contribution to blind/epileptic school), and asked to set the amount that the drug company should pay toward the construction of the school. As before, all consequentialist reasons for preferring the IK penalty were removed (the school was to be built anyway and the government would merely save the money awarded). The measure of interest is the presence and type of interaction between the level of award and whether the award is IK or OK.

Method

Subjects. Subjects were 68 University of Sussex undergraduates. They completed the question voluntarily at the end of a class.

Procedure. We used a between-subjects design with 4 conditions (2 damage types * 2 penalty types) (17 per group). Subjects read the relevant damage and penalty scenario and were asked to write down the amount of money that they thought that the company should contribute to the school as a penalty. Subjects were told that the government had capped the penalty at £100 million. This was to reduce the number of unclassifiable responses such as "The company should pay everything it can."

Questionnaire.

> A drug company manufactured a drug which caused irreversible blindness [epilepsy] in 10 children in Scotland. Investigations showed that the company did not carry out proper safety tests before marketing the drug.
>
> The drug company has already compensated the children involved.
>
> In addition, the drug company must pay punitive damages.
>
> The government has decided that the company must pay part of the cost of building a new residential school for blind [epileptic] children in the South of England. The school was to be built anyway. The company will pay for part of the building and the government will save money.
>
> The cost of the building is £100m, but the government has already agreed that this amount is much too large for the punitive damages. They only want the

drug company to pay part of this cost.

How much money should the government ask the drug company to contribute to the school? Remember that they should not be asked to pay the full amount.

The company should pay £_____

Results

(Insert Table 4 near here)

The results of Experiment 3 are shown in Table 4. There were no significant main effects, but there was a significant interaction of penalty type with damage type (2-way ANOVA, $F(1, 64) = 4.25, p < .05$). Subjects awarded higher penalties for IK cases. This result is consistent with the heuristics and compatibility hypotheses, but is the reverse of the prediction of the aggression hypothesis.

Experiment 4: Logging Study

The results of Experiments 1–3 have been consistent with the compatibility and heuristics hypotheses (although the subjects' justifications were more supportive of the heuristics hypothesis). The one case in which the two theories make opposing predictions is the case of valuations where there are no victims. The compatibility hypothesis predicts that greater IK penalties will be set (as in Experiment 3), due to greater attention to the compatible attribute. Conversely, the heuristics hypothesis is consistent with higher OK judgments in cases without clear victims. This result would occur if, first, people make more errors in trying to match the utility of attributes on different scales (i.e., the OK judgments are noisier) (see Figure 1), and, second, people believe that it is worse to under-penalize than to over-penalize (i.e., they think that the errors are asymmetrical). Taken together these two assumptions imply that subjects would set lower values for IK penalties, because they will need a greater amount of OK penalty to be sure that they

have punished sufficiently. We tested these assumptions directly in Experiment 4 by asking subjects how they viewed the severity of the two types of error and by elicting the highest and lowest acceptable amounts of penalty for each type of harm (IK and OK). The heuristics hypothesis just described predicts a much larger band of acceptability (akin to a confidence range of judgment) for OK judgments than IK judgments (consistent with the greater error in the judgments).

(Insert Figure 1 near here.)

We directly test the hypothesis that preference depends on the probability of the level of penalty being appropriate (see Figure 1). Subjects are asked to choose between their best estimates of fair penalties of each type. Then they are asked to choose between these penalties doubled, and between these penalties cut in half. Our hypothesis predicts a shift toward preference for the OK penalty for both the half and double ranges. This is because, if there is more error in the OK judgments, then diversions from the point estimate of the optimal judgment are more likely to be the true optimum than will be the case for a less errorful (IK) judgment. This is shown in Figure 1 by the crossing of the IK function with the OK function, so that the IK function is above the OK function for values close to the optima, but falls below the OK curve as values depart from the optima in either direction.

The design of Experiment 4 is somewhat different from the previous experiments. Subjects considered only one scenario, but they made many judgments about it. The scenario described negligent logging of a protected forest (no victims) and asked for judgments about fair amounts for three kinds of penalties: same forest (SF), different forest (DF) and money (Mo). The penalty for SF is to turn over to the government some amount of forest identical to what was destroyed. DF involves turning over a different kind of forest, and Mo involves payment. The DF–SF contrast compares IK and OK payments expressed in the same units (sq. miles), while the contrasts with Mo represent different units as well.

Method

Subjects. Subjects were 103 students at University of Pennsylvania, solicited and paid as in Experiments 1 and 2.

Procedure. Each subject completed the following questionnaire at his or her own pace. Order of items compared to each other was balanced, so the following is one form only.

Questionnaire.

> **[SF condition]**
> A logging company negligently cut down 100 square miles of protected forest in the U.S. It failed to check to make sure that cutting timber in this forest was legal. The question is how the government should penalize the company.
>
> Suppose that the penalty requires the company to turn over some of its own forest – which it would otherwise use for logging – to add to a national park at a different location. The forest to be turned over is essentially identical to the forest that was destroyed, both in value to the company and in its contents and appearance.
>
> A. What is your best estimate of the number of square miles that would be a fair penalty? _____sq. mi.
>
> B. How confident are you that this estimate is the right amount? Answer by giving the lowest and highest values that you would find acceptable.
> *lowest* _____sq. mi. *highest* _____sq. mi.
>
> C. If the decision were up to you, how much would you penalize the company? _____sq. mi.
>
> **[DF condition]**
> Suppose that the company has no forest identical to what was destroyed. It must turn over forest that is very different in contents and appearance. The forest to be turned over is now similar in value to the company to what was destroyed, and environmentalists also consider it to be equivalent. Nobody knows, though, about the long-run value of the two kinds of forest.
>
> D. What is your best estimate of the number of square miles that would be a fair
> penalty? _____sq. mi.
>
> E. How confident are you that this estimate is the right amount? Answer by giving the lowest and highest values that you would find acceptable.
> *lowest* _____sq. mi. *highest* _____sq. mi.
>
> F. If the decision were up to you, how much would you penalize the company? _____sq. mi.

[Mo condition]
Suppose that the penalty requires the company to pay a fine to the government. The market value of the forest destroyed was $1,000,000. The government was prevented by law from using this money to buy forest to replace what was lost.

G. What is your best estimate of the number of dollars that would be a fair penalty? $_____

H. How confident are you that this estimate is the right amount? Answer by giving the lowest and highest values that you would find acceptable.
lowest $_____ *highest* $_____

I. If the decision were up to you, how much would you penalize the company? $_____

J. Which judgment did you find more difficult to make, case A or case D? (Circle one):
A more difficult *equally difficult* *D more difficult*

K. Which judgment did you find more difficult to make, case A or case G? (Circle one):
A more difficult *equally difficult* *G more difficult*

L. Write down your answer to question A and your answer to question D.
A _____sq. mi. D $_____ (circle one):
Suppose you had to choose between one of these two penalties. Which would you prefer? (Circle one.)
prefer penalty A *both equal* *prefer penalty D*

M. Write down HALF of your answer to A and HALF of your answer to D.
half of answer to A _____sq. mi. half of answer to D $_____
Suppose you had to choose between one of these two penalties. Which would you prefer? (Circle one.)
prefer half of A *both equal* *prefer half of D*

N. Write down TWICE your answer to A and TWICE your answer to D.
twice answer to A _____sq. mi. twice answer to D $_____
Suppose you had to choose between one of these two penalties. Which would you prefer? (Circle one.)
prefer twice A *both equal* *prefer twice D*

O. Write down your answer to question A and your answer to question G.
A _____sq. mi. G $_____ (circle one):
Suppose you had to choose between one of these two penalties. Which would you prefer? (Circle one.)
prefer penalty A *both equal* *prefer penalty G*

P. Write down HALF of your answer to A and HALF of your answer to G.
half of answer to A _____sq. mi. half of answer to G $_____
Suppose you had to choose between one of these two penalties. Which would

you prefer? (Circle one.)
prefer half of A *both equal* *prefer half of G*

Q. Write down TWICE your answer to A and TWICE your answer to G.
twice answer to A _____sq. mi. twice answer to G $_____
Suppose you had to choose between one of these two penalties. Which would you prefer? (Circle one.)
prefer twice A *both equal* *prefer twice G*

R. In general, which is worse, that the company is penalized 5% too much or 5% too little? (circle one):
too much is worse *equally bad* *too little is worse*

The "difficulty" items (J, K) and the "confidence" items (B, E, H) were designed to test whether OK judgments were more difficult and errorful than IK judgments. Items L–Q were designed to test the possibility that IK penalties would be preferred only when they are perceived as accurate. OK penalities might be preferred when they are twice or half of what subjects consider fair, because of the greater uncertainty about what their ideal level is. More generally, these items test the robustness of the IK preference effect.

Results

Valuations

(Insert Table 5 near here)

The mean valuations for Experiment 4 are shown in Table 5. Due to high variability in the data, we performed non-parametric statistical tests. Subjects tended to set higher amounts for DF than SF penalties. The difference between DF and SF was significant for "best" estimates (items A, D, and G) ($p < .0005$, Wilcoxon test; DF exceeded SF for 44 subject vs. 12 for whom SF exceeded DF), and for final estimates (items C, F, and I) ($p < .0005$; DF exceeded SF for 41 vs. 14 subjects). Similar comparisons of SF or DF with Mo are not meaningful due to differences in units of measurement.

To make unit-free comparisons between the certainty of being close to the optimum for different types of penalty, we defined a variable "acceptable range": the difference between

the highest and lowest acceptable figures, divided by the highest. Ranges for SF and DF did not differ significantly, but the range for Mo was significantly higher than both of these ($p = .028$ for DF, $p = .037$ for SF, Wilcoxon tests). We hypothesized that range would be greater for OK penalties because of greater uncertainty about what penalties were appropriate. Although this may be true for Mo, we have no good evidence for it for DF vs. SF. In fact, the correlation coefficient between the range effect (DF vs. SF) and the effect for final penalty (both expressed as the log of the ratio) was only 0.065.

Although the increases between best estimates and final estimates were significant for all three measures (Wilcoxon test, $p < .0005$ for SF and DF; $p = .001$ for Mo overall, but not significant for one of the two orders), the differences between increases for SF, DF, and Mo were not significant. We might have expected greater increases in conditions with greater uncertainty about appropriate penalties, but it is also possible that the uncertainty was already largely taken into account in the "best" estimates given first. This would explain why the best estimate for DF tends to exceed that for SF. Also, the fact that final assessments were higher suggests that subjects did want to err on the side of too much rather than too little (as they later stated directly in item *R*: see below).

Preferences.

Of the 102 subjects who expressed a preference, 68 preferred SF to Mo in choosing between the "best" penalties that they had assigned, and 19 had the opposite preference (sign test, $p < .0005$). The effect holds over the doubled penalties (65 vs 24; $p < .0005$) and the half penalties (58 vs. 34; sign test, $p = .016$). However, the effect for the half penalties was significantly smaller ($p = .002$, sign test) than that for the full penalties. (No other differences were significant). Similarly subjects preferred the "best" estimate for SF to that for DF by 44 to 12 ($p < .0005$) for the full penalties, and by 46 to 17 for the double penalties (sign test, $p < .0005$). This effect was not present for the half penalties (31 vs

31), and the IK preference for the half penalties was significantly lower than for both the full penalties ($p = .008$) and double penalties ($p = .004$).

Note that the preference for IK full penalties was between penalties judged to be best. We used these figures rather than the final figures because we wanted the subjects' best estimates, before any presumed correction for error of estimation. The relative shift of preference to OK for half penalties could result from the hypothesized tendency to accept OK penalties that were far from this best estimate on the grounds that they were more likely to be correct, because of the greater uncertainty about correctness for OK estimates (see Figure 1). The failure to find this effect for double penalties is difficult to explain. (We cannot explain it, for example, in terms of distance from the optimum.) One possibility is that we did not move sufficiently far from the optimum for the IK curve to fall beneath the OK curve. The effects of deviations from the optimum penalty are investigated further in Experiment 5.

Errors and Difficulty of Judgment

SF judgments were generally regarded as easier than Mo judgments (63 vs 14 subjects; sign test, $p < .0005$), and SF judgments were easier than DF judgments (59 vs 11 subjects; sign test, $p < .0005$). Of those who thought that there was a difference in under-punishing vs over-punishing, 79 subjects viewed penalizing the company too much as better, while only 9 thought the reverse (sign test, $p < .0005$). These results are consistent with the heuristics hypothesis.

Discussion

Experiment 4 completed the four experiments looking at each pairing of victims/no victims with preference/valuation. We found that subjects set higher penalties for OK than IK. Experiments 3 and 4, taken together, show that the relative sizes of IK and OK valuations

are dependent on whether or not there were victims to the harm. These results were consistent with the heuristics hypothesis but not with the aggression or compatibility hypotheses, neither of which accounts for the different results in the cases with victims and those without victims (see Table 1). Experiment 4 also lends direct support for the heuristics hypothesis in the justifications, and in the questions concerning errors and difficulty.

Experiment 5: Logging with ratings

The first four experiment have dealt mainly with subjects' evaluations of IK and OK penalties at the perceived optimal level. In Experiment 5 we examine in more detail (following from Experiment 4) how subjects' preferences for IK and OK penalties change as the penalty value moves away from the ideal point. This experiment allows us to contrast various models of how errors might affect preferences and valuations for penalties.

One hypothesis can be derived from the assumption that subjects, in essence, perform the valuation task first. That is, they have some idea of the magnitude of an appropriate penalty of each type. One version of this hypothesis assumes that subjects have something like a single-peaked subjective utility function for magnitudes of each type of penalty.[3] (They could choose the magnitude leading to the peak, or optimal, value of each function as their response in the valuation task.) They would do the preference task by choosing the penalty with the higher subjective utility according to these functions. If we further assume that the function for OK penalties is the same as that for IK, but transposed to the right (higher magnitudes of penalty), then subjects would prefer the IK penalty if they were asked to choose between two equal penalties that are lower than the optimal IK amount, and they would prefer the OK penalty if they were asked to choose between two equal penalties that were both higher than the optimal OK amount. For example, if their optimal amounts are 3 units for IK and 4 units for OK, they will prefer 2 units of IK to 2

units of out-of kind, but they will prefer 5 units of OK to 5 units of IK. This situation is depicted graphically in Figure 2(a). Note that the IK peak is higher than the OK peak. This illustrates the finding from Experiment 4 that the optimum IK penalty tends to be preferred to the optimum OK penalty.

(Insert Figure 2 near here.)

Another subset of hypotheses – derived from our earlier hypothesis about the valuation task – holds that subjects are more uncertain about OK penalties and that they will compare penalties on the basis of the probability of each one being appropriate. By this account, subjects will increasingly prefer OK penalties when both penalties are farther from their ideal in either direction – too high or too low – because the IK penalty is increasingly likely to be inappropriate. This situation is shown in Figure 2(b), in which the OK curve is more spread out than the IK curve, allowing the OK curve to rise above the IK curve at either end. Again the OK optimum is lower than the IK optimum. This hypothesis received partial support in Experiment 4, and is investigated in much more detail below. In order to examine further the relation between valuation and preference with a wider range of values, we asked subjects to assign penalties in the SF and DF conditions and then to rate a number of SF and DF penalties of different magnitudes. The various penalties represent deviations from the subjects' ideal IK and OK penalties. The results of the experiment should enable us to distinguish between the various error models presented above.

Method

Thirty-nine subjects were given a single questionnaire using only the SF and DF scenarios, counterbalanced for order of presentation. For the valuation task, subjects were asked simply, "...how many square miles would you penalize the company?", and they were also asked for highest and lowest acceptable values. For the rating task, subjects were

given a table with the columns labeled "same-forest" and "different-forest" and the rows labeled with different numbers of square miles: 50, 100, 150, 200, 250, 300, 400, 500, 600.[4] They were told: "Suppose that we do not know yet which penalties are possible, so you are asked to rate all the penalties. The penalty imposed will be the one, of those that are possible, with the highest average rating. The final penalty will either be all same-forest or all different-forest, not some of each. Rate the penalties by assigning 100 to the best one and 0 to the worst. Then rate all the others on this scale. Ties are allowed."

Results

As found in Experiment 4, the mean valuation for DF (mean of 232 sq. mi.) was greater than that for SF (199 sq. mi.) ($p = .007$, Wilcoxon test; 20 subjects showing the effect vs 4 showing the opposite effect). The mean ranges for SF (430 sq. mi.) and DF (569 sq. mi.) did not differ significantly, however.

In the rating task, the maximum rating for SF (mean of 100) was greater than that for DF (92) ($p < .0005$, Wilcoxon test). The peak value on the scale of penalties (50 to 600) was also computed by finding for each subject the scale point leading to the maximum. (The scale was treated as consisting of equal steps for statistical purposes.) When more than one step was given the maximum value, the mean of these steps was used. The mean step value for DF (step 5.24, that is, a bit more than 250 sq. mi.) was not significantly greater than that for SF (5.01). However, when the comparison was limited to those subjects who assigned greater penalties to OK than to IK in the initial questions, the step value for DF (5.50) was significantly higher than that for SF (5.00) ($p = .011$, Wilcoxon test). Although the rating task led to variable and idiosyncratic responses (hence to high variance), the peak values are thus roughly consistent with the penalties assigned in the valuation task.

Several analyses were done to look at the off-peak ratings for SF and DF. The general

pattern was that SF ratings were consistently higher than, or equal to, DF ratings at the same distance from the peak. (For example, for subjects with the same peak steps for SF and DF, the sum of off-peak SF ratings was higher than that of off-peak DF ratings: $p = .046$, Wilcoxon test. Analyses that included subjects with different peaks yielded the same trend, but it was not significant.) In sum, we have no evidence that the subjective utility function of penalty sizes takes a different form for SF and DF. It appears that the function is simply shifted toward higher values for DF, for many subjects (see Figure 2a).

General Discussion

The experiments presented demonstrate clearly that people do not regard utility derived from different sources as interchangeable. They do not think that the only relevant aspect of a penalty is the level of (dis)utility it represents. Our experiments allowed us to contrast the predictions of three other psychological models of the evaluation of penalties: the heuristics hypothesis, the aggression hypothesis, and the scale compatibility hypothesis. The pattern of preferences, valuations and justifications from the five experiments supports the heuristics hypothesis more strongly than the other hypotheses.

Experiments 1, 2, 4, and 5 established that subjects prefer IK penalties. The one exception to this general result is that more subjects (but still not a majority) prefer OK penalties when both kinds are half of the amounts considered best. Although this result alone is consistent with subjects being more tolerant of erroneous penalty levels when more uncertainty about the appropriate level is present, we found no other evidence for this general hypothesis.

Nor did we find any evidence for the hypothesis that preferences for IK vs. OK depended on differences between kinds in closeness to the ideal penalty level. This hypothesis cannot explain the preference effects in Experiment 4, where the preferences were between levels considered best. (Or, alternatively, if the "best" responses were underestimates for

OK penalties – as we might think from looking at the "final" answers – then we would have found greater preference for OK penalties when both kinds were twice the levels considered best. This was not found either.) Moreover, Experiment 5 found direct evidence that the best IK penalty was rated higher than the best OK penalty.

The simplest account of the preference results is that subjects are applying a general rule that penalties should be IK, as expressed in Exodus. As noted in the Introduction there are many cases in which there is a good consequentialist justification for preferring compatibility between the harm and the penalty, at least when the penalty is linked to compensation. However, in the scenarios we devised, we removed all such consequentialist justification: e.g., the schools were to be built anyway, and would not benefit the damaged children. The preference results therefore appear to be another example – adding to those presented by Baron (1994) – of the use of a generally good heuristic rule without checking to ensure that the purposes of the rule are served, at least the consequentialist purposes.

The valuation results are more complex because they appear to depend on the presence or absence of victims to the harm. In Experiment 3 (with victims), subjects provided greater penalties when they were IK. In Experiments 4 and 5 (no victims), subjects provided greater OK penalties. The scale compatibility and aggression hypotheses do not distinguish between situations with and without victims and hence were not able to account for all the results. The scale-compatibility hypothesis was ruled out in Experiments 4 and 5, in which OK valuations were higher, while the aggression hypothesis cannot account for the results of Experiment 3, in which IK valuations were higher. The apparently conflicting experimental results can be resolved, however, by our heuristics approach, which explains the reason why these two cases are treated differently. Subjects' justifications for Experiment 3 suggest that they were providing more penalties for IK because this situation was most like the more usual one in which the injurer compensated the victim by making up the loss in kind. Similar carry-over from intuitive rules for compensation

to those for assignment of penalties were found by Baron and Ritov (1993). Experiments 4 and 5 had no particular identifiable victims. The difference there can be explained in terms of greater uncertainty for OK penalty, plus a desire to err on the side of too much rather than too little. Thus, subjects appear to use more than one rule for valuation tasks. Direct evidence for the use of these heuristics came from subjects' justifications, and from questions concerning errors and difficulties in judgment.

The proposed mechanism for the aggression hypothesis was that the less preferred (more distant) resource is less valuable to the aggrieved party, and hence is required in disproportionately large amounts. Some support for this was found in Experiment 4 (the environmental damage valuation study), in which the OK settlements were viewed as less valuable. However, we have suggested an additional mechanism: namely, the effect of uncertainty and differential concern with the two types of errors. The idea that there is more uncertainty about OK judgments was supported by the fact that the range of acceptable penalties was greater for money (OK) than for forests (IK – although this effect was not found for the two different kinds of forests). Also, subjects found OK judgments more difficult, and most subjects preferred to err on the side of too much penalty.

The studies we report have implications for the outcomes of legal and nonlegal cases. The most common valuation scale for penalties, especially in the law, is monetary. Our results suggest that, in cases where there is no obvious victim (e.g., environmental damage), perpetrators will get reduced penalties if the damages are expressed in terms of the units of damage. Conversely, perpetrators may receive excessive penalties when the money they pay cannot be used to "undo" the damage they have done, as in cases that involve loss of life, personal relationships, or health. The heuristics that increase OK penalties in our studies may be part of the thinking of jurors who assess damage awards that appear excessive to others, both in cases of personal injury and environmental damage (Huber, 1988). They may think that they have to err on the side of excess, since the penalty can-

not be easily matched with the magnitude of the harm. Legal experts such as appellate judges may not as subject to these effects as juries, although we have not investigated this possibility. Instructions to juries might also reduce the effects in the courtroom.

In addition, the preference for in-kind penalties may lead to inefficiencies in the structure of environmental law itself. One of the acts that applied to the Exxon Valdez spill, the (U.S.) Comprehensive Environmental Response, Compensation, and Liability Act of 1984 (amended 1986), specifies that payments be used for environmental projects. This requirement is even clearer in the Oil Pollution Liability and Compensation Act of 1990, which was passed in response to the Valdez spill. The deterrent effect depends on the size of the penalty, not on what it is used for. It may be more efficient to use the payments for something else, but our results suggest that this proposal would seem less acceptable to the public. Increased understanding of the deterrent rationale of penalties could thus increase the efficiency in use of resources (Baron, 1994; Baron & Ritov, 1993).

In our experiments we examined only penalties, but we might expect similar effects in judgments involving compensation either IK or OK, even when the "harm" is agreed to. For example, in-kind compensation may be preferred when a hazardous facility is sited in a community – lower electric rates for a nuclear power plant, free garbage collection for a landfill, etc. (Gregory & Kunreuther, 1990). Such effects would be worthy of further investigation.

Another issue which deserves more attention concerns exactly what constitutes IK vs. OK penalties. Experiment 4 showed that this is not a strict dichotomy, but more of a continuum. In this experiment, penalties ranged from replacement of exactly equivalent forest (SF condition), which was clearly IK; through substitution of somewhat different forest (DF condition), which was "somewhat IK"; to a monetary fine, clearly OK. In Experiments 2 and 3 (the drug damage cases), the penalties were not the most strictly IK, in that the health was not be restored to blinded or epileptic children: rather quality of life

to other blind or epileptic children was improved through the donation to the school. Our experiments suggest that compatibility can be a matter of degree. The effect of degree of compatibility on various other psychological processes and in other scenarios seems worthy of further investigation.

References

Baron, J. (1994). Nonconsequentialist decisions (with commentary and reply). *Behavioral and Brain Sciences, 17*, 1–42.

Baron, J. & Ritov, I. (1993). Intuitions about penalties and compensation in the context of tort law. *Journal of Risk and Uncertainty, 7*, 17–33.

Bentham, J. (1948). *An introduction to the principles of morals and legislation.* Oxford: Blackwell Publisher. (Original work published 1843)

Brown, C. (1985). Deterrence in tort and no-fault: the New Zealand experience. *California Law Review, 73*, 976–1002.

Foa, E. B. (1970). Frustration-aggression as exchange of resources. Unpublished Ph.D. thesis, University of Missour, Columbia.

Foa, U. (1971). Interpersonal and economic resources. *Science, 171*, 345–351.

Gregory, R., & Kunreuther, H. (1990). Successful siting incentives. *Civil Engineering, 60*, 73–75.

Huber, P. W. (1988). *Liability: The legal revolution and its consequences.* New York: Basic Books.

Kenworthy, T. (1994). Pristine forests at stake as Alaskans balance priorities in Valdez cleanup. *The Washington Post*, August 23, p. A3.

Mastromatteo, E. (1993). Alternative compensation schemes: New Zealand's universal disability system. *OH&S Canada, 9*, 124–127.

Posner, R. A. (1986). *Economic analysis of law* (3rd edition). Boston: Little, Brown.

Rauber, P. (1992). Slick maneuvers; use of Exxon Valdez settlement. *Sierra, 77* (6), 34.

Shavell, S. (1987). *Economic analysis of accident law.* Cambridge, MA: Harvard University Press.

Taylor, G. (1991). Batman comics thieves get creative punishment. *National Law*

Journal, May 13, p. 43.

Thaler, R. (1985). Mental accounting and consumer choice. *Marketing Science, 4*, 199–214.

Thompson, L. & Loewenstein, G. (1992). Egocentric interpretations of fairness and interpersonal conflict. *Organizational Behavior and Human Decision Processes, 51*, 176–197.

Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review, 95*, 371–384.

Zornow, D. M., & Reed, P. D. (1992). Guidelines don't fit environmental cases. *National Law Journal,* March 2, p. 19.

Author Notes

This research was supported by N.S.F. grants SES91-09763 and SBR92-23015. We also thank Brain Bornstein, Howard Kunreuther, and two anonymous reviewers, for helpful comments on previous drafts of this manuscript, and George Loewenstein for a discussion on a presentation about this work at the Judgment and Decision Making Society meeting, Washington, 1993.

Footnotes

1. The heuristics hypothesis in its general form could also predict the opposite result for the valuation task. Subjects may find reason to set greater penalties in the OK case. We find that this does not happen in cases with identifiable victims, so we do not need to explain it. However, our statistical tests of our hypothesis concerning valuation are two tailed, in recognition that the general form of the hypothesis predicts a departure in either direction.

2. According to utilitarian or economic theory of penalties, the expected utility of the optimal penalty should equal the harm done. There is no reason to expect small errors in one direction to be any worse than small errors in the other direction in their overall consequences. See, for example, Figure 2.1 in Shavell (1987). Total social costs increase only a little if the amount of care taken is on either side of the optimum. So this heuristic is difficult to justify.

3. We call this subjective utility to distinguish it from the utility at issue in utilitarian theories of optimal penalties.

4. These values were representative of the range of values that subjects gave in a pilot study.

Table 1: Summary of predictions of different theories concerning preferences and valuations for OK and IK penalties.

THEORY	PREFERENCE TASK	VALUATION TASK
Normative (utilitarian)	$IK \approx OK$	$IK = OK$
Heuristics hypothesis	$IK \succ OK$	$IK > OK$ (if victims) $OK > IK$ (if no victims)
Compatibility hypothesis (Tversky et al.)	$IK \succ OK$	$IK > OK$
Aggression hypothesis (Foa)	$IK \succ OK$	$OK > IK$

Table 2: Characteristics of Experiments 1–4.

TASK	VICTIMS	NO VICTIMS
PREFERENCE	Experiment 1	Experiment 2
VALUATION	Experiment 4	Experiment 3

Table 3: Numbers of subjects showing preferences for beach or forest penalty, given damage to forest of beach in Experiment 1. IK preferences in bold face. ($N = 36$)

Preferred penalty	**Forest Damage**	**Beach Damage**
Forest	**20**	12
Equal	11	12
Beach	5	**12**

Table 4. Table of mean punitive damage awards (in millions of £) in Experiment 3. (IK values shown in bold face). ($n_1 = n_2 = n_3 = n_4 = 17$)

School	**Blindness**	**Epilepsy**
Blind	**47.35**	27.06
Epilepsy	26.74	**35.65**

Table 5. Mean valuations and standard errors (S.E.) of valuations in each condition (SF, DF, Mo) for Experiment 4.

	Mean	**S.E.**
Best estimate		
SF	304	101
DF	355	145
Mo	$1,969,216	$287,036
Final estimate		
SF	731	485
DF	993	733
Mo	$2,315,505	$316,567
Acceptable range: (highest − lowest) / highest		
SF	0.441	0.024
DF	0.459	0.022
Mo	0.473	0.053

Figure Captions

Figure 1. Graphical representation of model showing more error in OK judgments (dashed) than IK judgments (solid). The vertical axis here represents the probability density that the penalty is the optimal one.

Figure 2. Predictions of two different models of how subjective utility of penalty varies with size of penalty for IK and OK penalties. Figure 2(a) assumes equal seriousness of departures from the optimum, with OK curve shifted to the right. Figure 2(b) assumes much greater tolerance for deviant OK judgments. Both models assume IK optimum has greater utility than OK optimum.