Damages, Norms, and Punishment

Cass R. Sunstein*

How do people make judgments about appropriate punishment? How do they translate their moral judgments into more tangible penalties? What is the effect of group discussion? And what does all this have to do with social norms?

In this essay I attempt to make some progress on these questions. I do some by outlining some of the key results of a series of experimental studies conducted with Daniel Kahneman and David Schkade, and by elaborating, in my own terms, on the implications of those studies. Among other things, we find that the process of group discussion dramatically changes individual views, among other things by making people move toward higher dollar awards. In other words, groups often go to extremes. The point has large implications for the role of norms in deliberation and the effect of deliberation in altering norms. We also find that people’s judgments about cases, viewed one at a time, are very different from their judgments about cases seen together. Making one-shot decisions, people produce patterns that they themselves regard as arbitrary and senseless. The point has large implications for the aspiration to coherence within the legal system.

More particularly, our principal findings are as follows:

- In making moral judgments about personal injury cases, people’s judgments are both predictable and widely shared. The judgments of one group of six people, or twelve people, nicely predict the judgements of another groups of six people, or twelve people.
- In making punitive damage awards for personal injury cases, people’s judgments are highly unpredictable and far from shared. People do not have a clear sense of the meaning of different points along the dollar scale. Hence dollar judgments of one group of six people, or twelve people, do not well predict the dollar judgments of other groups of six people, or twelve people.
- As compared with the median of predeliberation judgments, the effect of deliberation is to increase dollar awards, often quite substantially. Group discussions have the remarkable effect of raising group members’ judgments about appropriate punishment.
- People care about deterrence, but they do not think in terms of optimal deterrence. People are intuitive retributivists, and they reject some of the most common and central understandings in economic and utilitarian theory.
- People’s judgments about cases in isolations are systematically different from their judgments about cases taken together. The consequence of the system of “one at a time” judgments is to produce a pattern of outcomes that seems incoherent to the very people who make those judgments.

Now for some details. For purposes of the present discussion, I will speak broadly and in qualitative terms; readers interested in numbers and statistical analysis might consult the papers from which I shall draw.

I. Steady Norms, Unsteady Awards

Suppose that people are asked to rank a set of personal injury cases, or libel cases, or cases involving sexual harassment or damage to the environment. Suppose too that people are asked to rate those cases, in terms of appropriate punishment, on a bounded numerical scale – say, 0 to 8, where 0

* Cass R. Sunstein is the James Lowell du Pont Chair in Social Theory at the University of Chicago Law School, and is Professor of Psychology and of Law at the University of Chicago. He is also the Director of the Center for the Study of Law and Social Policy at the University of Chicago Law School. He is the author of numerous books, including The Architect of Liberty: Alpheus W. Pitkin and American Constitutional Law (Princeton University Press, 1994), The Perfect Citizen: Why We Need the Constitution (Princeton University Press, 1999), Legal Reasoning (Cambridge University Press, 2001), and The Ticking Clock: Asymmetry in the Legal Time Machine (Cambridge University Press, 2004). He is the editor of the journal Law and Social Inquiry and the co-author, with Daniel Kahneman, of the book The Budgeting of Time (Princeton University Press, 2006). He is currently working on a book on the nature of legal reasoning, entitled Legal Reasoning, 2006 Edition.
means “punished not at all,” and 4 means, “punished moderately,” and 8 means, “punished extremely severely.” Will people agree? Will the decision of one group of six, or twelve, provide good predictions about what other groups of six or twelve will do? The answer will depend on whether the social norms that govern moral outrage and intended punishment are widely shared. If they are shared, we should not expect sharp divergences in terms of both ranking and rating.

A. Remarkably Shared Judgments

Undertaking a series of studies of citizen judgments, we have found that at least in some domains, the relevant norms are indeed widely shared. In personal injury cases, the judgment of any particular group of six is highly likely to provide a good prediction of the judgment of any other group of six. In this sense, a “moral judgment” jury is indeed able to serve as the conscience of the community.

Indeed we can go further. Members of different demographic groups show considerable agreement about how to rank and rate personal injuries cases. We asked thousands of people to rank and rate cases. We also elicited information about the demographic characteristics of all of those people. As a result, it is possible, with the help of the computer, to put individuals together, so as to assemble all-male juries, all-female juries, all-white juries, all-African-American juries, all-poor juries, all-rich juries, all-educated juries, all less-educated juries, and so forth. Creating “statistical juries” in this way, we found no substantial disagreement, in terms of rating or ranking, within any group. In personal injuries cases, people simply agree.

Subsequent work has broadened this finding, showing that people agree on how to rank tax violations, environmental violations, and occupational safety and health violations. From this evidence, it seems reasonable to hypothesize that in a wide range of domains, people will agree how to rank and rate cases. The moral norms within a heterogeneous culture are, to that extent, widely shared, and strikingly so. Now this does not mean that people will agree on how to rank cases from different categories (a point to which I will return). Nor does it mean that small groups will always agree on how to do the ranking. But it does mean that within category, agreement is the exception, not the rule.

B. Remarkably Erratic Dollar Awards

What about dollars? Do the broadly shared norms also produce regularity in jury verdicts? Our of our central findings is that it does not.

With respect to dollars, both individuals and jury-size groups are all over the map. Even when moral rankings are shared – as they generally are – dollar awards are extremely variable. A group that awards a “5,” for defendant’s misconduct, might give a dollar award of $500,000, or $2 million, or $10 million. A group that awards a “7” might award $1 million, or $10 million, or $100 million. In fact there is so much noise, in the dollar awards, that differences cannot be connected with demographic characteristics. It is not as if one group – whites, for example – give predictably different awards from another – say, African-Americans or Hispanics. We cannot show systematic differences between young and old, men and women, well-educated and less well-educated. The real problem is that dollar awards are quite unruly, from one individual to another and from one small group to another.

What accounts for this? Why do people share moral judgments but diverge on dollar awards? The best answer is that the effort to “map” moral judgments onto dollars is an exercise in “scaling without a modulus.” In psychology, it is well known that serious problems will emerge when people are asked to engage in a rating exercise on a scale that is bounded at the bottom but not at the top, and when they are not given a “modulus” by which to make sense of various points along the scale. For example, when people are asked to rate the brightness of lights, or the loudness of noises, they will not be able to
agree if no modulus is supplied and if the scale lacks an upper bound. But once a modulus is supplied, agreement is substantially improved. Or if the scale is given an upper bound, and if verbal descriptions accompany some of the relevant points, people will come into accord with one another.

The upshot is that much of the observed variability with punitive damage awards – and in all likelihood with other damage awards too – does not come from differences in social norms. It comes from variable, and inevitably somewhat arbitrary, “moduli” selected by individual jurors and judges. If the legal system wants to reduce the problem of different treatment of the similarly situated, it would do well to begin by appreciating this aspect of the problem. The point applies to many legal problems, including criminal sentences, pain and suffering awards, administrative penalties, and damages for libel, sexual harassment, and intentional infliction of emotional distress.

II. What Do Groups Do? The Effects of Deliberation

The discussion thus far has been ambiguous about a key question: whether the studies involved deliberating juries, or mere individuals placed, by computer, into small groups, with individual views being somehow “pooled” to create a verdict. The ambiguity stems from the fact that our initial study did indeed involve individuals and statistical pooling, creating what might be called “statistical juries” -- whose verdict, as we reported it, consisted of the view of the median juror. We chose the median juror on the ground that this seemed to be the best estimate of what the jury itself would do. But in a subsequent study, involving about 3000 people, we tested this hypothesis – and found that it was wrong. What we found does not falsify the findings just described; on the contrary, it reinforces them. But it also says a great deal about the effects of deliberation and the role of social norms in that process.

In brief, we tested the effects of deliberation on both punitive intentions and dollar judgments. To test the effects of deliberation on punitive intentions, we asked people to record their individual judgments privately, on a bounded scale, and then we asked them to join six-member groups to generate unanimous “punishment verdicts.” To test the effects of deliberation on dollar judgments, we also asked people to record their private judgment, pre-deliberation, and then to join six-member groups to produce unanimous dollar awards. Juries produced both punishment verdicts and dollar verdicts; half entered punishment verdicts first, and half entered dollar awards first. Only a small number of the 500 juries “hung.”

Two findings are especially important. First, deliberation made the lower punishment ratings decrease, when compared to the median of pre-deliberation judgments of individuals, while deliberation made the higher punishments ratings increase, when compared to that same median. Second, dollar awards of groups were systematically higher than the median of individual group members – so much so that in 27% of the cases, the dollar verdict was as high as, or higher than, that of the highest individual judgment, pre-deliberation. How can this pattern be explained?

A. Group Polarization

With respect to punishment ratings, the answer lies in the phenomenon of group polarization – a pervasive process by which group members end up in a more extreme position in line with the predeliberation tendencies of group members. It is now well-known that if a group has a defined median position – if, for example, people in the group tend to think that global warming is a serious problem, or that gun control is bad idea – members will shift toward a more extreme version of what they already think. Critics of gun control, talking with one another, will end up more critical of gun control. Those fearful of global warming will, as a result of group discussion, end up more fearful of global warming. The basic finding has been made on many topics and in many nations. [2]
In my view, there are large lessons here about the formation of social norms and attitudes, and in particular about the role of groups in forming the norms and views of group members. A homogeneous group might well lead members in quite extreme directions. After speaking with one another, like-minded people are apt to end up thinking what they thought before, but much more so. A heterogeneous group is far less likely to have this effect.

What explains group polarization? What does deliberation drive low punishment ratings down, and move high punishment ratings up? There appear to be two answers. The first involves the exchange of information within the group. In a group that favors a high punishment rating, group members will make many arguments in that direction, and relatively few the other way. Speaking purely descriptively, the group’s “argument pool” will be skewed in the direction of severity. Group members, listening to the various arguments, will naturally move in that direction. As with punishment ratings, so too with much else: Feminism, global warming, capital punishment, affirmative action, and so forth. The initial dispositions of group members will determine the proportion of arguments in the various directions. And individuals will respond, quite rationally, to what they have heard, thus moving in the direction suggested by the dominant tendency.

The second explanation involves social influences. Most people want to be, and to be perceived in, a certain way. If you are in a group that wants to punish someone severely, you might find it uncomfortable to be urging relative leniency. To protect your reputation, and perhaps your self-conception, you might move, if you move at all, in most favored direction. To be sure, some hardy souls will not move at all, and those who are self-identified contrarians might deliberately move in the opposite direction, rejecting the dominant view just because it is the dominant view. But what we observed, and what is universally observed, is that most of those who move tend to go in the group’s preferred direction – and that as a result, the group will be more extreme than its members before deliberation began.

B. Rhetorical Asymmetry

But none of this seems to an adequate account of what was observed with dollar awards. Here we found a general increase in verdicts. To be sure, groups whose predeliberation median was low tended to see a smaller increase than groups whose predeliberation median was high. Dollar awards increased quite generally. Unlike in the context of punishment ratings, there was no “switchpoint” along which some went up and some went down. Why is this?

Some clue, I think, is provided by two little-noticed studies of the effects of group deliberation. It has been found that as compared with individuals, groups of doctors are more likely to engage in heroic measures to save patients. It has also been found that as compared with individuals, groups are more likely to divide sums of money equally with strangers. Perhaps these results can be understood to exemplify group polarization, with the initial tendency – to save the patient, to be favor -- being amplified in both instances. This is certainly possible. But an alternative explanation would suggest that in these domains, there is a kind of rhetorical asymmetry between the two opposing positions. When a rhetorical asymmetry is in place, one side has an automatic upper-hand in argument. One side is likely to win, simply because in light of existing norms, that side has a rhetorical advantage. Doctors who seek more in the way of heroic measures are more likely to win an argument with those who seek less. Group members who seek more fairness are likely, at least in the relevant settings, to prevail over those who seek more selfishness.

It is important to see that rhetorical asymmetry can operate in many domains, and that all this is a
function of social norms. In one group, those who favor stiffer penalties for drug offenders might have an automatic advantage – so that any discussion will move the group toward stiffer penalties. In another group, the opposite might be true, so that the effect of the discussion will be to produce greater leniency. It is undoubtedly easy to come up with a long list of groups showing a rhetorical asymmetry in one direction or another.

With respect to punitive damage awards, we hypothesize that especially in light of the difficulty of scaling, those who favor higher dollar punishments are in a better rhetorical position than those who favor lower ones. The arguments on behalf of the higher awards are simply more intuitive, in the abstract. In fact we conducted a simple follow-up study, asking participants to list arguments for higher and lower awards (knowing nothing about the particular case), and then asking which set of arguments was easier to make. Far more people said that the higher award was easier to support than the lower one.

There are general implications here about the effect of social norms on group discussion and also about the production of social norms through group discussion. It seems clear that preexisting norms can push people in predictable directions, and also that these very movements can help create new norms, or at least stronger versions of the preexisting ones. In an “iterated polarization game,” or an “iterated rhetorical asymmetry game,” very significant shifts are to be expected. Perhaps the point helps account for religious and ethnic strife, for feuds, even for violent behavior.[5]

III. Retribution and Deterrence

On the economic theory of punishment, the state’s goal, when imposing penalties, is to ensure optimal deterrence. To increase deterrence, the law might increase the severity of punishment, or instead increase the likelihood of punishment. A government that lacks substantial enforcement resources might impose high penalties, thinking that it will produce the right deterrent “signal” in light of the fact that many people will escape punishment altogether. A government that has sufficient resources might impose a lower penalty, but enforce the law against all or almost all violators.

In the context of punitive damages, all this leads to a simple theory: The purpose of such damages is to make up for the shortfall in enforcement. If injured people are 100% likely to receive compensation, there is no need for punitive damages. If injured people are 50% likely to receive compensation, those who bring suit should receive a punitive award that is twice the amount of the compensatory award. The simple exercise in multiplication will ensure optimal deterrence.

But there is a large question whether social norms and the theory of optimal deterrence can fit together. Do people want optimal deterrence? Do they accept or reject the economic theory of punishment?

We attempted to cast light on this question through two experiments. In the first, we gave people cases of wrongdoing, arguably calling for punitive damages, and also provided people with explicit information about the probability of detection. Different people saw the same case, with only one difference: varying probability of detection. People were asked about the amount of punitive damages that they would choose to award. Our goal was to see if people would impose higher punishments when the probability of detection was low.

In the second experiment, we asked people to evaluate judicial and executive decisions to reduce penalties when the probability of detection was high, and to increase penalties when the probability of detection was low. We wanted people to say whether they approved or disapproved of varying the penalty with the probability of detection.
Our findings were simple and straightforward. The first experiment found that varying the probability of detection had no effect on punitive awards. Even when people’s attention was explicitly directed to the probability of detection, people were indifferent to it. People’s decisions about appropriate punishment were unaffected by seeing a high or low probability of detection. The second experiment found that strong majorities of respondents rejected judicial decisions to reduce penalties because of high probability of detection -- and also rejected executive decisions to increase penalties because of low probability of detection. In other words, people did not approve of an approach to punishment that would make the level of punishment vary with the probability of detection. What apparently concerned them was the extent of the wrongdoing, and the right degree of moral outrage – not optimal deterrence.

The most general conclusion is that social norms do not coexist comfortably with optimal deterrence theory. People seem to intuitive retributivists. They come to the social role of juror with moral intuitions inconsistent with the economic theory of deterrence.

IV. Coherence, Categories, and Context

I have suggested that people have a fairly easy time rating and ranking cases within a single category. Hence they share judgments about the outrageousness of a defendant’s conduct in a personal injury case. But do people share judgments about how to compare a personal injury case with a libel case? Can people compare cases across categories? Probably most important: What would they think of the pattern that they produce if, as is usual, they tend to decide cases one at a time?

We do not have full answers to these questions; but suggestive evidence has started to emerge. The simplest point is that when people are trying to rank cases from different categories, they have far more difficulty, in the sense that they are unsure exactly what to do. This lack of certainty translates into a lack of consensus. People agree much more on how to rank cases within a category than how to rank cases across categories. (I put to one side the evident difficulties in deciding what counts as a “category.”) It is easy to design experiments in which people will simply disagree about whether (for example) a comparatively serious tax violation is worse, or less bad, than a lawless act that harms the environment. Hence the social norms that govern cross-category comparisons are not as widely shared as the social norms that govern within-category comparisons.

Perhaps this is not big news. A more striking finding is that people’s judgments about cases, taken one at a time, are very different from their judgments about the same cases, taken in the context of a problem from another category. An example: We asked people to assess a case involving personal injury, on a bounded scale and also on a dollar scale. We also asked people to assess a case involving financial injury, on a bounded scale and also on a dollar scale. When the two cases are judged in isolation, the financial injury case receives a more severe rating and a higher dollar award. But when the two cases are seen together, there is a significant judgment shift, in which people try to ensure that the financial award is not much higher, and for many respondents is lower, than the personal injury award. People’s decisions about the two cases are very different, depending on whether they see the case alone or in the context of a case from another category.

We observed exactly the same kind of shift for judgments about two problems calling for government regulation and expenditures: skin cancer among the elderly and protection of coral reefs. Looking at the two cases in isolation, people will pay more to protect coral reefs, and register more satisfaction from doing that. But looking at the two cases together, people will be quite disturbed at this pattern, and will generally want to pay more to protect elderly people from cancer. Here too there is a significant shift in judgment.
Is this a problem? And what accounts for the switch? Let me offer a preliminary account. When people see a case in isolation, they naturally “normalize” it by comparing it to a set of comparison cases that it readily calls up. If you are asked, is a German Shepherd big or small, you are likely to respond that it is big; if you are asked, is a Volkswagen Bug big or small, you are likely to respond that it is small. But people are well-aware that a German Shepherd is smaller than a Volkswagen bug. People answer as they do because a German Shepherd is compared with dogs, whereas a Volkswagen Bug is compared with cars. So far, so good; in these cases, everyone knows what everyone else means. We easily normalize judgments about size, and the normalization is mutually understood. (John Stockton, who is about six feet tall, is a very small basketball player!)

In the context of legally relevant moral judgments, something similar happens, but it is far from innocuous. When evaluating a case involving financial injury, people apparently “normalize” the defendant’s conduct by comparing it with conduct in other cases from the same category. They do not easily or naturally compare that defendant’s conduct with conduct from other categories. Because of the natural comparison set, people are likely to be quite outraged by the misconduct, if it is far worse than what springs naturally to mind. The same kind of thing happens with the problem of skin cancer among the elderly. People compare that problem with other similar problems – and conclude that it is not so serious, within the category of health-related or cancer-related problems. So too with personal injury cases (normalized against other personal injury cases) and problems involving damage to coral reefs (normalized against other cases of ecological harm).

When a case from another category is introduced, this natural process of comparison is disrupted. Rather than comparing a skin cancer case with other cancers, or other human health risks, people see that it must be compared with ecological problems, which (in most people’s view) have a lesser claim to public resources. Rather than comparing a financial injury case to other cases of business misconduct, people now compare it to a personal injury case, which (in most people’s view) involves more serious wrongdoing. As a result of the wider viewscreen, judgments shift, often dramatically.

I believe that this uncovers a serious problem with current practice in many domains of law. The problem is that when people assess cases in isolation, their viewscreen is narrow, indeed limited to the category to which the case belongs, and that as a result, people produce a pattern of outcomes that makes no sense by their own light. In other words, the overall set of outcomes is one that people would not endorse, if they were only to see it as a whole. Their considered judgments reflect the very pattern that they have produced, because of a predictable feature of human cognition. The result is a form of incoherence. We can find that incoherence not only in jury verdicts, but also in administrative fines and in criminal sentencing, where no serious effort has been made to ensure that the overall pattern of outcomes makes the slightest sense.

V. Conclusion

In this paper I have attempted to cast some light on the relationship among social norms, punitive intuitions, group deliberation, coherence, and several other issues in law and legal theory. We have seen that diverse people rank and rate cases, within a single category, in a similar way; that they produce erratic dollar awards largely because of difficulty of using a dollar scale; that with respect to moral judgments, discussion moves people toward a more extreme point in light with their initial predisposition; that with respect to dollars, discussion increases awards; that existing norms fit poorly with optimal deterrence theory; and that one-shot judgments produce patterns that people would reject, if only they were to see them.

These are descriptive points. It is far from clear what, if anything, should be done by way of
legal reform. But it would be reasonable to conclude that a system of one-shot judgments by
juries, scaling in the dark and offered no comparison cases, is not likely to be a sensible way to produce
civil fines. Arbitrariness and incoherence are almost inevitable. In the abstract, a more bureaucratic
approach, allowing a degree of rationalization, would seem to be far superior. Of course I cannot defend
such an approach in this space; bureaucracies have notorious problems of their own, not least because
they can be out of touch with prevailing social norms. But with an understanding of the problems
discussed here, perhaps we can make better sense of some of the largest movements in twentieth-century
law, which consisted precisely in an effort to replace one-shot jury decisions with institutions that are
accountable, and subject to prevailing norms, but also able to overcome serious cognitive problems
faced by isolated individuals and groups. [6] On this view, the proper response to such problems consists
in better institutional design. It would not be at all surprising if the twenty-first century saw bolder
movements in the same general direction.

* Karl N. Llewellyn Distinguished Service Professor, University of Chicago, Law School and
Department of Political Science. This paper draws on joint work with Daniel Kahneman and David
Schkade, who share the credit for what is worthwhile here, and who deserve no blame for what is not.
Interested readers might consult the papers from which I draw: Daniel Kahneman et al., Shared Outrage
and Unpredictable Awards, J. Risk and Uncertainty (1998); Cass R. Sunstein et al., Assessing Punitive
Damages, Yale LJ (1988); David Schkade et al., Deliberating About Dollars: The Severity Shift, Colum.
L. Rev. (2000); Cass R. Sunstein et al., Do People Want Optimal Deterrence, J. Legal Stud. (2000); Cass

[1] The point is discussed in more detail in Cass R., Sunstein, Deliberative Trouble? Why Groups Go To


[6] For example, see Price Fishback and Sheldon Kantor, A Prelude to the Welfare State: The Rise of
Workers’ Compensation (2000)