2002

Outrage Leary Lecture

Cass R. Sunstein

Outrage

Cass R. Sunstein[*]

Outrage is a common reaction to perceived wrongdoing. Those who feel outrage are often motivated to injure or punish wrongdoers, whether through violence, law, or otherwise. But how are legal outcomes affected by outrage? Inside and outside of law, how do people translate their outrage into more tangible penalties? How does group discussion affect outrage? Are groups more outraged than individuals or less so?

In this Article I attempt to make some progress on these questions. For a number of years, I have been engaged in a series of experimental studies of punitive damage awards with Daniel Kahneman, David Schleade, and Ilana Ritov.[1] The goal of those studies has been to uncover some of the sources of unpredictability in such awards and also to understand the sources of low and high awards. The resulting work, much of it highly technical, seems to me to have broader implications for a range of issues in both law and politics. Many public officials and juries are charged with the task of "mapping" some factual or moral judgment onto a numerical scale, involving dollars or years. If punishment judgments are typically a function of outrage, the problems found in jury behavior might well have analogies in criminal sentencing and administrative fines. And if punitive awards are erratic, for predictable reasons, might not the same be true for awards involving libel, intentional infliction of emotional distress, sexual harassment, and more?

In any case, outrage, in mild and less mild forms, seems to help produce behavior in many domains. If we know something about the dynamics of outrage, we might be able to understand behavior that is otherwise inexplicable. An especially important issue involves the relationship between individual action and group action. If groups moderate outrage, or increase it, then we might be able to see what produces radicalism, extremism, and even violence and terrorism. In investigating some of these issues, I draw throughout on our empirical findings,

---

[*]Karl N. Llewellyn Distinguished Service Professor of Jurisprudence, Law School and Department of Political Science, University of Chicago. This text is the basis for the Leary Lecture, delivered at the S.J. Quinney College of Law at the University of Utah on September 10, 2002.

[1]*See* Daniel Kahneman et al., *Shared Outrage and Erratic Awards: The Psychology of Punitive Damages*, 16 J. RISK & UNCERTAINTY 49 (1998); David Schkade et al., *Deliberating About Dollars: The Severity Shift*, 100 COLUM. L. REV. 1139 (2000); Cass R. Sunstein et al., *Assessing Punitive Damages (with Notes on Cognition and Valuation in Law)*, 107 YALE L.J. 2071 (1998) [hereinafter Sunstein et al., *Assessing Punitive Damages*]; Cass R. Sunstein et al., *Do People Want Optimal Deterrence?*, 29 J. LEGAL STUD. 237 (2000) [hereinafter Sunstein et al., *Optimal Deterrence?*]; Cass R. Sunstein et al., *Predictably Incoherent Judgments*, 54 STAN. L. REV. 1153 (2002). Many of these papers are collected in CASS R. SUNSTEIN ET AL., PUNITIVE DAMAGES: HOW JURIES DECIDE (2002) [hereinafter SUNSTEIN ET AL., PUNITIVE DAMAGES].

but in a way that involves my own extrapolations, some of them admittedly speculative.

The simplest lessons are that punitive awards are rooted in outrage; that levels of outrage command a degree of agreement among diverse Americans, at least in some domains; but that people have a great deal of difficulty in "mapping" their outrage onto a bounded scale. Among other things, we find that the process of group discussion dramatically changes individual views by making people move toward higher dollar awards. In other words, groups often go to extremes.[2] The point has large implications for the role of outrage in deliberation and the effect of deliberation in altering outrage. 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:[3]

- 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 judgments of another group of six people, or twelve people. These shared judgments cut across demographic differences, so that there is no difference, in the relevant cases, between rich and poor, old and young, white and African-American, poorly educated and well-educated.
- In making punitive damage awards about 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. And when people are outraged, group discussion makes them more outraged still.
- 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 isolation are systematically different from their judgments about cases taken together. The consequence of the system

---

[2]The point is discussed in more detail in ROGER BROWN, SOCIAL PSYCHOLOGY: THE SECOND EDITION 200–45 (1986).

[3]These findings are based on the studies cited *supra* note 1.

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[4]

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 means "punished not at all," 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 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 injury cases. Thousands of people were asked to rank and rate cases. Information was elicited 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 injury 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

[4]For more detailed discussion of the methods and findings presented in Parts I and II, see the experimental studies cited *supra* note 1.

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).[5] Nor does it mean that small groups will always agree on how to do the ranking. But it does mean that within each category, agreement is the rule, not the exception.

B. Remarkably Erratic Dollar Awards

What about dollars? Do the broadly shared norms also produce regularity in jury verdicts? One 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.[6] 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 levels of outrage or social norms. It comes from variable, and

---

[5]*See infra* Part IV.

[6]For a discussion on problems with "mapping," see Kahneman et al., *supra* note 1, at 53–55, 74–75; Sunstein et al., *Assessing Punitive Damages*, *supra* note 1, at 2106–07.

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. In these areas as well, those entrusted with the task of "mapping" lack a modulus with which to discipline their decisions. An empirical study of pain and suffering awards finds that at least forty percent of the variance cannot be explained by differences in case characteristics.[7] A legal system that does not give guidance for "mapping" is bound to create similar problems in other areas. Indeed, the rise of guidelines for criminal sentencing can be understood as responsive, at least in part, to exactly this problem.

II. WHAT DO GROUPS DO? THE EFFECTS OF DELIBERATION

A. The Problem

The findings thus far did not involve deliberating juries. They were based on the judgments of individuals placed by computer into small groups with individual views being "pooled" to create a verdict. The result was to create "statistical juries" whose verdict consisted of the view of the median juror, which seemed to provide a reasonable estimate of what the jury itself would do. But how does group deliberation affect outrage? In a subsequent study, involving several thousand people, we found that the median juror is not, in fact, a good predictor of the ultimate verdict of the jury. What we found does not falsify the findings just described; in a way, it reinforces them. But it also says a great deal about the effects of deliberation on moral outrage.

The study tested the effects of deliberation on both punitive intentions and dollar judgments. The study involved about three thousand jury-eligible citizens; its major purpose was to determine how individuals would be influenced by seeing and discussing the punitive intentions of others.[8] To test the effects of deliberation on punitive intentions, people were asked to record their individual judgments privately, on a bounded scale, and then to join six-member groups to generate unanimous "punishment verdicts." Hence, subjects were asked to record, in advance of deliberation, a "punishment judgment" on a scale of 0 to 8, where 0 indicated that the defendant should not be punished at all and 8 indicated that the defendant should be punished extremely severely. After the individual judgments were recorded, jurors were asked to deliberate to a unanimous

---

[7]David Leebron, *Final Moments: Damages for Pain and Suffering Prior to Death*, 64 N.Y.U. L. REV. 256, 310 (1989).

[8]The study is described in detail in Schkade et al., *supra* note 1.

"punishment verdict." It would be reasonable to predict that the verdicts of juries would be the median of punishment judgments of jurors; but the prediction would be badly wrong.

B. Findings

Two findings are especially important. First, deliberation made the lower punishment ratings decrease when compared to the median of predeliberation judgments of individual jurors, while deliberation made the higher punishment ratings increase when compared to that same median. When the individual jurors favored little punishment, the group showed a "leniency shift," meaning a rating that was systematically lower than the median predeliberation rating of individual members.[9] But when individual jurors favored strong punishment, the group as a whole produced a "severity shift," meaning a rating that was systematically higher than the median predeliberation rating of individual members.[10]

When the median juror judgment was 4 or more on the 8-point scale, the jury's verdict was above the median predeliberation judgment of individuals.[11] Consider, for example, a case involving a man who nearly drowned on a yacht that was defectively constructed.[12] People tended to be outraged by the idea of a defectively built yacht, and groups became far more outraged than their median members.[13] But when the median juror judgment was less than 4, the jury's verdict was below the median judgment of individuals.[14] Consider a case involving a shopper who was injured in a fall when an escalator stopped suddenly.[15] Individual jurors were not greatly bothered by the incident, seeing it as a genuine accident rather than a case of serious wrongdoing; and groups were more lenient than individuals.[16]

The second important finding is that dollar awards of groups were systematically higher than the median of individual group members—so much so that in twenty-seven percent of the cases the dollar verdict was as high as, or higher than, that of the highest individual judgment, predeliberation.[17] The basic result is that deliberation causes awards to increase, and it causes high awards to increase a great deal. The effect of deliberation, in increasing dollar awards, was most pronounced in the case of high awards. For example, the median individual judgment in the case involving the defective yacht was $450,000, whereas the

---

[9] *Id.* at 1152, 1154–56.
[10] *Id.*
[11] *Id.*
[12] *Id.* at 1151–52.
[13] *Id.*
[14] *Id.*
[15] *Id.*
[16] *Id.*
[17] *Id.* at 1140, 1163.

median jury judgment in that same case was $1 million.[18] But awards shifted upwards for low awards as well.[19] Consider three examples of extreme severity shift from the raw data:

- A jury whose predeliberation judgments were $200,000, $300,000, $2 million, $10 million, $10 million, and $10 million reached a verdict of $15 million;
- A jury whose predeliberation judgments were $200,000, $500,000, $2 million, $5 million, and $10 million reached a verdict of $50 million;
- A jury whose predeliberation judgments were $2 million, $2 million, $2.5 million, $50 million, and $100 million reached a verdict of $100 million.[20]

Notably, the degree of dispersion between individual predeliberation judgments did not contribute to greater or lesser shifts as a result of deliberation. In other words, juries whose members were in rough agreement (*i.e.*, had a low standard deviation) about dollars or punishment did not show a different shift from groups whose members were in substantial disagreement about dollars or punishment.

How can this pattern be explained? What is the relationship between outrage and processes of group interaction?

C. Group Polarization

With respect to punishment ratings, the answer lies in the phenomenon of group polarization. This is the 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.[21] Consider some examples of the basic phenomenon, which has been found in over a dozen nations:[22] (a) a group

---

[18] *Id.* at 1152.

[19] *Id.*

[20] *Id.* at 1155.

[21] For general discussion, see Cass R. Sunstein, *Deliberative Trouble? Why Groups Go to Extremes*, 110 YALE L.J. 71 (2000).

[22] *See* BROWN, *supra* note 2, at 222. These nations include the United States, Canada, New Zealand, Germany, and France. *See, e.g.*, Dominic Abrams et al., *Knowing What to Think By Knowing Who You Are: Self-categorization and the Nature of Norm Formation, Conformity and Group Polarization*, 29 BRIT. J. SOC. PSYCHOL. 97, 112–17 (1990) (New Zealand); Johannes Zuber et al., *Choice Shift and Group Polarization: An Analysis of the Status of Argument and Social Decision Schemes*, 62 J. PERSONALITY & SOC. PSYCHOL. 50, 58–60 (1992) (Germany). Of course, it is possible that some cultures would show a greater or lesser tendency toward polarization; this would be an extremely interesting area for empirical study.

of moderately profeminist women will become more strongly profeminist after discussion;[23] (b) after discussion, citizens of France become more critical of the United States and its intentions with respect to economic aid;[24] (c) after discussion, whites predisposed to show racial prejudice offer more negative responses to the question whether white racism is responsible for conditions faced by African Americans in American cities;[25] (d) after discussion, whites predisposed not to show racial prejudice offer more positive responses to the same question.[26] As statistical regularities, it should follow, for example, that those moderately critical of an ongoing war effort will, after discussion, sharply oppose the war; that those who believe that global warming is a serious problem are likely, after discussion, to hold that belief with considerable confidence; that people tending to believe in the inferiority of a certain racial group will become more entrenched in this belief as a result of discussion.

In my view, there are large lessons here about the formation of outrage and attitudes generally, 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? Why does deliberation drive low punishment ratings down and move high punishment ratings up? There appear to be three answers.[27] 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 a certain way and also to be perceived in a certain way. If you are in a group that is outraged and wants to punish someone severely, you might find it

---

[23]David G. Myers, *Discussion-Induced Attitude Polarization*, 28 HUM. REL. 699, 707–13 (1975).

[24]BROWN, *supra* note 2, at 223–24.

[25]David G. Myers & George D. Bishop, *Discussion Effects on Racial Attitudes*, 169 SCI. (n.s.) 778, 778–79 (1970).

[26]*Id.*

[27]For further discussion of the first two explanations, see BROWN, *supra* note 2, at 230–39.

uncomfortable to be urging relative leniency. To protect your reputation, and perhaps your self-conception, you might move, if you move at all, in the 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. With respect to outrage, the lesson is clear. To preserve their preferred self-image, individuals, finding themselves in an outraged group, will tend to become more outraged still. There is a clue here about the wellsprings of "runaway juries," crowd-control problems, unjustified extremism, and even terrorism (as we shall soon see).[28]

The third explanation begins by noting that people with extreme views tend to have more confidence that they are right and that as people gain confidence, they become more extreme in their beliefs.[29] The intuition here is simple: Those who lack confidence, and who are unsure what they should think, tend to moderate their views. It is for this reason that cautious people, not knowing what to do, are likely to choose the midpoint between relevant extremes.[30] But if other people seem to share one's view, that person is likely to become more confident that that view is right—and hence to move in a more extreme direction. In a wide variety of experimental contexts, people's opinions have been shown to become more extreme simply because their view has been corroborated and because they have been more confident about learning of the shared views of others.[31] If a person is unsure whether he should be outraged, he might become far more outraged when he learns that others are outraged too. The agreement increases confidence and hence extremity.

D. Skewed Debates

In the context of punitive damage awards by juries, a particular finding deserves emphasis. Thus far, the discussion of the relevant study has stressed how deliberation affected punitive intentions, measured on a bounded numerical scale. But jurors were also asked to record their dollar judgment, in advance of deliberation, and then to deliberate to reach dollar verdicts. As I have noted, the highest awards increased by the largest amount—but all awards increased.[32] This

---

[28] *See infra* Part II.F.

[29] *See* Robert S. Baron et al., *Social Corroboration and Opinion Extremity*, 32 J. EXPERIMENTAL SOC. PSYCHOL. 537, 558 (1996).

[30] *See* Mark Kelman et al., *Context-Dependence in Legal Decision Making*, 25 J. LEGAL STUD. 287, 287–88 (1996).

[31] Baron et al., *supra* note 29, at 558.

[32] *See supra* text accompanying notes 17–20.

might seem to be a surprise. An understanding of group polarization might suggest that low awards would drop and high awards would be raised, with the difference pivoting around some neutral point, say, $60,000. But this is not what was observed.

Why did dollar awards systematically increase? A possible explanation, consistent with group polarization, is that any positive median award suggests a predeliberation tendency to punish, and deliberation aggravates that tendency by increasing awards. But even if correct, this explanation seems insufficiently specific. The striking fact is that those arguing for higher awards seem to have an automatic "rhetorical advantage" over those arguing for lower awards. A subsequent study of University of Chicago law students confirmed this finding, suggesting that given prevailing social norms, people find it much easier to defend high awards than the opposite.[33] Similar findings have been made in seemingly distant areas. Suppose that doctors are deciding what steps to take to resuscitate patients; are individuals more or less likely to support heroic efforts? Evidence suggests that as individuals, doctors are less likely to do so than groups, probably because those who favor such efforts have a rhetorical advantage over those who do not.[34]

Or consider the difference between individual behavior and team behavior in the Dictator Game.[35] In this game, subjects are told they can allocate a sum of money, say $10, between themselves and some stranger. Contrary to the standard prediction, most subjects do not keep all or almost all of the money for themselves, choosing instead to keep somewhere between $6 and $8.[36] How is individual behavior affected once people are placed in teams? The answer is that they choose still more equal divisions.[37] This result can be explained by reference to a rhetorical advantage, disfavoring the suggestion of selfishness, even within the group.[38] Within the relevant population, and many imaginable other groups, people do not want to appear to be greedy. Of course the outcome, and the effect of group influence, would change if the dictator team had some reason for hostility to the beneficiaries of their largesse. We can easily imagine a variation of the Dictator Game in which, for example, people of one group were deciding how much to allocate to another group that was thought to be far wealthier. In this variation, the rhetorical advantage might favor greater selfishness.

---

[33]*See* Schkade et al., *supra* note 1, at 1161–62.

[34]*See* Caryn Christenson & Ann Abbott, *Team Medical Decision Making*, *in* DECISION MAKING IN HEALTH CARE: THEORY, PSYCHOLOGY, AND APPLICATIONS 267, 273–76 (Gretchen Chapman & Frank Sonnenberg eds., 2000).

[35]Timothy N. Cason & Vai-Lam Mui, *A Laboratory Study of Group Polarisation in the Team Dictator Game*, 107 ECON. J. 1465, 1465–66 (1997).

[36]*See id.* at 1468–73.

[37]*Id.*

[38]Cason and Mui themselves explain the outcome in terms of social comparison, a closely related account. *See id.*

What produces a rhetorical advantage? The simplest answer points to existing social norms, which of course vary across time and place. Among most Americans, existing norms make it easier to argue, other things equal, for higher penalties against corporations for egregious misconduct; but it is possible to imagine subcommunities (corporate headquarters?) in which the rhetorical advantage runs the other way. In extreme cases, those with a rhetorical advantage are on the correct side of a social taboo, whether mild or strong. In any case it is easy to imagine many other contexts in which one or another side has an automatic rhetorical advantage. Consider, as possible examples, debates over whether there should be higher penalties for those convicted of drug offenses, or whether tax rates should be reduced; in modern political debates, those favoring higher penalties and lower taxes seem to have the upper hand. Of course there are limits on the extent of change that can be made. But when a rhetorical advantage is involved, group deliberation will produce significant shifts in individual judgments. Undoubtedly, legislative behavior is affected by mechanisms of this sort, and it is likely that many movements within judicial panels can be explained in similar terms.

Are rhetorical advantages unhelpful or damaging? In the abstract, this is an impossible question to answer, because shifts have to be evaluated on their merits and in substantive terms. Perhaps the higher punitive awards that follow deliberation are simply better; so too, perhaps, for the movements by doctors, taking more heroic measures, and by groups deciding how equally to spread funds. The major point is that such advantages exist; and it would be most surprising if they were always benign. There is a minor point as well: With respect to dollars, the severity shift produces an increase in variance and unpredictability that is an increase over the already-high levels of unpredictability shown by "statistical juries." This is an independent problem, one to which I will return.

E. Some Extensions

When we consider the ingredients of punishment judgments, group polarization and rhetorical advantage have large implications for behavior inside and outside the courtroom. Apparently, people who begin with a high level of outrage become still more outraged as a result of group discussion. Moreover, the degree of the shift depends on the antecedent level of outrage; the higher the antecedent outrage, the greater the shift as a result of internal deliberations.[39]

---

[39]*See* Schkade et al., *supra* note 1, at 1152 (showing that in top five outrage cases, mean shift was 11%, higher than in any other class of cases). The effect is more dramatic still for dollars; high dollar awards shifted upwards by a significant margin. *See id.* This finding is closely connected to another one: Extremists are most likely to shift, and likely to shift most, as a result of discussions with one another. *See* JOHN TURNER ET AL., REDISCOVERING THE SOCIAL GROUP: A SELF-

There is a point here about the sources not only of severe punishment, but also of crowd behavior, vengeance, rebellion, and even violence, for outrage lies behind these as well.[40] If like-minded people—those predisposed to be outraged—are put together, significant changes are to be expected. Moving well beyond the domain of punitive damages, I offer some speculative remarks on possible extensions here.

It should be easy to see that group polarization is inevitably at work in feuds, ethnic and international strife, and war. One of the characteristic features of feuds is that members of feuding groups tend to talk only to one another, fueling and amplifying their outrage, and solidifying their impression of the relevant events. Group polarization occurs every day within Israel and among the Palestinian Authority. Many social movements become possible through the heightened effects of outrage on group members;[41] consider the movement for rights for deaf people, which was greatly enhanced by the fact that the deaf have a degree of geographical isolation.[42] And group polarization is not a social constant. It can be increased or decreased, and even eliminated, by certain features of group members or their situation. Consider three factors in particular.

1. Antecedent Extremism

Extremists are especially prone to polarization. It is more probable that they will shift, and it is probable that they will shift more. When they start out at an extreme point, and are placed in a group of like-minded people, they are likely to go especially far in the direction with which they started.[43] As I have suggested, there is a lesson here about the sources of terrorism and political violence in general, a lesson to which I shall return.[44] And because there is a link between confidence and extremism, the confidence of particular members also plays an important role; confident people are all the more prone to polarization.[45]

CATEGORIZATION THEORY 154–59 (1987).

[40]*See* Russell Hardin, *The Crippled Epistemology of Extremism*, *in* POLITICAL EXTREMISM AND RATIONALITY 3, 16 (Albert Breton et al. eds., 2002).

[41]*See* Edna Ullmann-Margalit & Cass R. Sunstein, *Inequality and Indignation*, 30 PHIL. & PUB. AFF. 337, 360 (2001).

[42]*See* Sharon Groch, *Free Spaces: Creating Oppositional Consciousness in the Disability Rights Movement*, *in* OPPOSITIONAL CONSCIOUSNESS 65, 67–72 (Jane Mansbridge & Aldon Morris eds., 2001).

[43]This is a finding in the study of punitive damage awards, where groups with extreme medians showed the largest shifts. *See* Schkade et al., *supra* note 1, at 1152. For other evidence, see TURNER ET AL., *supra* note 39, at 158.

[44]*See* Hardin, *supra* note 40, at 16–19; *infra* Part II.F.

[45]*See* Maryla Zaleska, *The Stability of Extreme and Moderate Responses in Different Situations*, *in* GROUP DECISION MAKING 163, 163–64 (Hermann Brandstätter et al. eds., 1982).

2. Solidarity and Affective Ties

If members of the group think that they have a shared identity and a high degree of solidarity, there will be heightened polarization.[46] One reason is that if people feel united by some factor (for example, politics or necessity), dissent will be dampened. Hence if members of the deliberating group are connected by affective ties, polarization will increase.[47] If they tend to perceive one another as friendly, likeable, and similar to them, the size and likelihood of the shift will increase.[48] The existence of such ties reduces the number of divergent arguments and also intensifies social influences on choice. One implication is that mistakes are likely to be increased when group members are united mostly through affective ties and not through concentration on a particular task; it is in the former case that alternative views will less likely find expression.[49] It follows that people are less likely to shift if the direction advocated is being pushed by unfriendly group members; the likelihood of a shift, and its likely size, are increased when people perceive fellow members as friendly, likeable, and similar to them.[50] A sense of "group belongingness" affects the extent of polarization.[51] In the same vein, physical spacing tends to reduce polarization; a sense of common fate and intragroup similarity tend to increase it, as does the introduction of a rival "outgroup."[52]

An interesting experiment attempted to investigate the effects of group identification by manipulating two variables.[53] First, some subjects were given instructions in which group membership was made salient (the "group immersion" condition), whereas others were not (the "individual" condition). For example, subjects in the group immersion condition were told that their group consisted solely of first-year psychology students and that they were being tested as group members rather than as individuals. Second, some subjects were "de-individualized" by having to work on computers in separate rooms, whereas others were asked to work in a single office with desks facing each other (the

---

[46]Abrams et al., *supra* note 22, at 112–16.

[47]*See* Brooke Harrington, The Pervasive Effects of Network Content 2 (2000) (unpublished manuscript, on file with Utah Law Review).

[48]*See* Hermann Brandstätter, *Social Emotions in Discussion Groups*, *in* DYNAMICS OF GROUP DECISIONS 93, 107–08 (Hermann Brandstätter et al. eds., 1978); *see also* TURNER ET AL., *supra* note 39, at 154–59 (attempting to use this evidence as basis for new synthesis, labeled "a self-categorization theory of group polarization").

[49]*See* Harrington, *supra* note 47, at 220–26.

[50]*See* Brandstätter, *supra* note 48, at 107–08; *see also* TURNER ET AL., *supra* note 39, at 158 (implying that group of comparative extremists will show comparatively greater shift toward extremism).

[51]TURNER ET AL., *supra* note 39, at 151.

[52]*Id.*

[53]Russell Spears, *De-Individualization and Group Polarization in Computer-Mediated Communication*, 29 BRIT. J. SOC. PSYCHOL. 121, 123–24 (1990).

"individualized" condition). In the de-individualized condition, visual anonymity was increased. All conditions were held constant in one respect: Every subject was told that people "like them" tended to support one or another view. The relevant issues involved affirmative action, government subsidies for the theatre, privatization of nationalized industries, and phasing out nuclear power plants.

The results were quite striking. Polarization generally occurred. But there was the least such polarization in the individual/individuated condition; group polarization was greatest in the de-individuated/group immersion condition, when group members met relatively anonymously and when group identity was emphasized. Interestingly, there was no significant difference in polarization between the two individuated conditions (with and without emphasis on group immersion).[54] From this experiment, it is reasonable to speculate that polarization is highly likely to occur, and to be most extreme, when group membership is made salient.

3. Exit

Over time, group polarization can be fortified through "exit," as members leave the group because they reject the direction in which things are heading.[55] If exit is pervasive, the tendency to extremism will be greatly aggravated. The point bears on the beliefs and actions of some religious groups, cults, dissident groups, and political outsiders of multiple sorts. As ambivalent members leave, the median group member becomes more extreme (and outraged), and interactions breed more extremity still. At the same time, a number of exits might well breed stronger intergroup norms of solidarity, thus stifling dissent. Political extremism and violence are a predictable result.[56]

F. A Brief Speculative Note on Outrage and Terrorism

I have referred in several places to the relationship among outrage, group polarization, and terrorism. It will be useful to offer some notes on this topic, partly because it is of independent interest and partly because the processes involved in terrorism can find analogues in far more benign processes.[57] When groups become caught up in hatred and violence, it is rarely because of economic

---

[54] *Id.* at 130.

[55] *See* Hardin, *supra* note 40, at 9.

[56] *See id.*

[57] In this section I draw on substantially similar material that appears in Cass R. Sunstein, *Why They Hate Us: The Role of Social Dynamics*, 25 HARV. J.L. & PUB. POL'Y 429, 435–39 (2002).

deprivation[58] or primordial suspicions;[59] it is far more often a product of social mechanisms.[60] Indeed, unjustified extremism frequently results from a "crippled epistemology," in which extremists react to a small subset of relevant information, coming mostly from one another.[61]

A central point is that terrorist leaders act as polarization entrepreneurs. They create enclaves of like-minded people. They stifle dissenting views and do not tolerate internal disagreement. They take steps to ensure a high degree of internal solidarity. They restrict the relevant argument pool and take full advantage of reputational forces, above all by using the incentive of group approval. Terrorist acts themselves are motivated by these forces and incentives. Consider, for example, the following account:

> Terrorists do not even consider that they may be wrong and that others' views may have some merit. . . . They attribute only evil motives to anyone outside their own group. [The] common characteristic of the psychologically motivated terrorist is the pronounced need to belong to a group . . . . Such individuals define their social status by group acceptance.
>
> Terrorist groups with strong internal motivations find it necessary to justify the group's existence continuously. A terrorist group must terrorize. As [sic] a minimum, it must commit violent acts to maintain group self-esteem and legitimacy. Thus, terrorists sometimes carry out attacks that are objectively nonproductive or even counterproductive to their announced goal.[62]

In fact, terrorist organizations impose psychological pressures to accelerate the movement in extreme directions. Here, too, group membership plays a key role. Thus:

> Another result of psychological motivation is the intensity of group dynamics among terrorists. They tend to demand unanimity and be intolerant of dissent. With the enemy clearly identified and unequivocally evil, pressure to escalate the frequency and intensity of operations is ever present. The need to belong to the group discourages resignations, and the fear of compromise disallows their acceptance.

---

[58]*See* Alan B. Krueger & Jitka Maleckova, Education, Poverty, Political Violence and Terrorism: Is There a Causal Connection? 29 (2002) (unpublished manuscript, on file with author).

[59]*See* Timur Kuran, *Ethnic Norms and Their Transformation Through Reputational Cascades*, 27 J. LEGAL STUD. 623, 648 (1998).

[60]*See* Sunstein, *supra* note 57, at 429–35.

[61]*See* Hardin, *supra* note 40, at 18–20.

[62]TERRORISM RESEARCH CENTER, THE BASICS OF TERRORISM Part 2: The Terrorists, *available at* http://www.geocities.com/CapitolHill/2468/basics.html (n.d.).

> Compromise is rejected, and terrorist groups lean toward maximalist positions. . . . In societies in which people identify themselves in terms of group membership (family, clan, tribe), there may be a willingness to self-sacrifice seldom seen elsewhere.[63]

Training routines specifically reinforce the basic message of solidarity amidst outrage and humiliation. Hitler similarly attempted to create group membership, and to fuel movements toward extremes, by stressing the suffering and the humiliation of the German people. This is a characteristic strategy of terrorists of all stripes, for humiliation fuels outrage.[64] "Many al-Qaida trainees saw videos . . . daily as part of their training routine. Showing hundreds of hours of Muslims in dire straits—Palestinians . . . Bosnians . . . Chechens . . . Iraqi children—was all part of al-Qaida's induction strategy."[65] In the particular context of Al Qaeda, there is a pervasive effort to link Muslims all over the globe, above all by emphasizing a shared identity, one that includes some and excludes others. Thus Osama bin Laden "appeals to a pervasive sense of humiliation and powerlessness in Islamic countries. Muslims are victims the world over . . . Bosnia, Somalia, Palestine, Chechnya, and . . . Saudi Arabia. . . . [H]e makes the world simple for people who are otherwise confused, and gives them a sense of mission."[66] Hence there are unmistakable cult-like features to the indoctrination effort: "[T]he military training [in Al Qaeda camps] is accompanied by forceful religious indoctrination, with recruits being fed a stream of anti-western propaganda and being incessantly reminded about their duty to perform jihad."[67]

Intense connections are built into the structure: "The structure of Al Qaida, an all-male enterprise, . . . appears to involve small groups of relatively young men who maintain strong bonds with each other, bonds whose intensity is dramatised and heightened by the secrecy demanded by their missions and the danger of their projects."[68]

This discussion, brief and speculative as it is, should be sufficient to show the central role of outrage and group dynamics in producing terrorists, and indeed in answering the much-disputed question, "Why do they hate us?" Terrorists are made, not born. More particularly, terrorists are made through emphatically social

---

[63] *Id.*

[64] On humiliation, see AVISHAI MARGALIT, THE DECENT SOCIETY 9–27 (Naomi Goldblum trans., 1996).

[65] Giles Foden, *Secrets of a Terror Merchant*, MELBOURNE AGE, Sept. 14, 2001, http://www.theage.com.au/news/world/2001/09/14/FFX1ONZFJRC.html.

[66] Jeffrey Bartholet, *Method to the Madness*, NEWSWEEK, Oct. 22, 2001, at 55.

[67] Stephen Grey & Dipesh Gadher, *Inside Bin Laden's Academies of Terror*, THE SUNDAY TIMES (LONDON), Oct. 7, 2001, at 10.

[68] Vithal C. Nadkarni, *How to Win Over Foes and Influence Their Minds*, THE SUNDAY TIMES OF INDIA, Oct. 7, 2001, 2001 WL 28702843 (quoting Lionel Tiger, Darwin Professor of Anthropology, Rutgers University).

processes. Things could easily be otherwise. Timur Kuran has shown that "ethnification"—close identification with one's ethnic group, in a way that involves hatred of others—is not a matter of history but of current social processes, closely akin to those discussed here.[69] With relatively small changes, a nation that suffers from intense ethnic antagonism could be free from that scourge. So too, I am suggesting, for terrorism. If enclaves of like-minded and susceptible people are an indispensable breeding ground for terrorism, then it is easy to imagine a situation in which nations, not radically different from the way they are today, could be mostly free from terrorist threats.

The simplest and most important lesson for law and policy is that if a nation aims to prevent terrorist activities, a good strategy is to prevent the rise of enclaves of like-minded people. Many of those who become involved in terrorist activities could end up doing something else with their lives. Their interest in terrorism comes, in many cases, from an identifiable set of social mechanisms (generally from associating with extremist leaders and like-minded others). If the relevant associations can be disrupted, terrorism is far less likely to arise.

III. RETRIBUTION AND DETERRENCE

Now let us return to the question of punishment. On the economic account, 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.[70] In the first, we gave people cases of wrongdoing, arguably calling for punitive damages,

---

[69]*See* Kuran, *supra* note 59, at 648–49.

[70]The two experiments are discussed in Sunstein et al., *Optimal Deterrence?*, *supra* note 1, and SUNSTEIN ET AL., PUNITIVE DAMAGES, *supra* note 1, at 132–41.

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. Outrage is what matters, not the 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 be intuitive retributivists. They come to the social role of juror with moral intuitions inconsistent with the economic theory of deterrence. Those intuitions are grounded in outrage. Facts of the case that dampen or weaken outrage will have an effect on punitive intentions. And those intentions will interact in complex ways with dollar awards, precisely because of the difficulty of translating moral outrage into a number that can be used by the legal system.

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? The question seems quite important, because it bears on the question whether the legal system can produce coherent patterns in which distinctions, between plaintiffs and defendants alike, make a degree of sense.

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 not as 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: People were asked to assess a case involving a personal injury on a bounded scale and also on a dollar scale. People were also asked to assess a case involving financial injury on a bounded scale and also on a dollar scale. When the two cases were judged in isolation, the financial injury case received a more severe rating and a higher dollar award. But when the two cases were seen together, there was a significant judgment shift in which people tried to ensure that the financial award was not higher than the personal injury award. People's decisions about the two cases were very different, depending on whether they saw a case alone or in the context of a case from another category.

Exactly the same kind of shift was observed 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 wanted to pay more to protect coral reefs and registered more satisfaction from doing that. But looking at the two cases together, people were quite disturbed at this pattern and generally wanted to pay more to protect elderly people from skin cancer. Here too there was 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 Beetle big or small, you are likely to respond that it is small. But people are well aware that a German shepherd is smaller that a Volkswagen Beetle. People answer as they do because a German shepherd is compared with dogs, whereas a Volkswagon Beetle 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.) What happens, in ordinary communication, is innocuous. It does not breed error or confusion.

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. Indeed there is reason to believe that the pattern, in many domains, is quite senseless. And it may not be too much of a stretch to suggest that the same is true of reactions, some of the time, by both individuals and institutions—that people are quite outraged about behavior that, in a broader or different comparison set, would outrage them little or not at all.

V. CONCLUSION

In this Article I have attempted to cast some light on the relationship among outrage, 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 the difficulty of using a dollar scale; and that with respect to moral judgments, discussion moves people toward a more extreme point in light of their initial predisposition. We have also seen 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. I have also offered some extensions and speculations. Perhaps most importantly, I have suggested that outrage is highly vulnerable to group influences and that a set of like-minded people is highly likely to go to extremes. The point helps to explain behavior within ordinary legislatures as well as behavior by political extremists of all kind.

These are descriptive points. It is far from clear what, if anything, should be done by way of legal reform. But in the context of civil litigation, 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.[71] 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.

[71] *See, e.g.*, PRICE V. FISHBACK & SHAWN EVERETT KANTOR, A PRELUDE TO THE WELFARE STATE: THE ORIGINS OF WORKERS' COMPENSATION (2000) (discussing rise of workers' compensation programs in early twentieth century America).