

ISSN 1045-6333

HARVARD

JOHN M. OLIN CENTER FOR LAW, ECONOMICS, AND BUSINESS

THE DENOMINATOR BLINDNESS EFFECT:
ACCIDENT FREQUENCIES AND THE
MISJUDGMENT OF RECKLESSNESS

W. Kip Viscusi
Richard J. Zeckhauser

Discussion Paper No. 387

10/2002

Harvard Law School
Cambridge, MA 02138

This paper can be downloaded without charge from:

The Harvard John M. Olin Discussion Paper Series:
http://www.law.harvard.edu/programs/olin_center/

The Denominator Blindness Effect: Accident Frequencies and the Misjudgment of Recklessness*

W. Kip Viscusi

Cogan Professor of Law and Economics
Harvard Law School
Cambridge, MA 02138

and

Richard J. Zeckhauser

Ramsey Professor of Political Economy
John F. Kennedy School of Government
Harvard University
Cambridge, MA 02138

Abstract

People seriously misjudge accident risks because they routinely neglect relevant information about exposure. Such risk judgments affect both personal and public policy decisions, e.g., choice of a transport mode, but also play a vital role in legal determinations, such as assessments of recklessness. Experimental evidence for a sample of 422 jury-eligible adults indicates that people incorporate information on the number of accidents, which is the numerator of the risk frequency calculation. However, they appear blind to information on exposure, such as the scale of a firm's operations, which is the risk frequency denominator. Hence, the actual observed accident frequency of accidents/exposure is not influential.

*This research was supported in part by the Sheldon Seevak Research Fund, the Olin Center for Law, Economics and Business, and a grant to the authors from the Exxon Corporation. DeYett Law provided superb programming assistance and expositional suggestions. Two anonymous referees, participants at the NBER law and economics workshop, and Miriam Avins provided helpful comments.

**The Denominator Blindness Effect:
Accident Frequencies and the Misjudgment of Recklessness**

W. Kip Viscusi & Richard J. Zeckhauser

© 2002 W. Kip Viscusi and Richard J. Zeckhauser. All Rights Reserved.

I. Introduction

Juries determining whether punitive damages should be awarded often must confront the issue of whether the defendant has engaged in reckless behavior. When the defendant is an individual, malice is often an additional concern but malice generally is not a factor for actions of corporate entities.¹ Standard jury instructions with respect to punitive damages often call for the jury to determine whether the behavior of the company was reckless. The following example, from an actual court case, is typical:

The purposes of punitive damages are to punish a defendant and to deter a defendant and others from committing similar acts in the future.

Plaintiff has the burden of proving that punitive damages should be awarded by a preponderance of the evidence. You may award punitive damages only if you find that the defendant's conduct

- (1) was malicious; or
- (2) manifested reckless or callous disregard for the rights of others.

Conduct is malicious if it is accompanied by ill will, or spite, or if it is for the purpose of injuring another.

In order for conduct to be in reckless or callous disregard of the rights of others, four factors must be present. First, a defendant must be subjectively conscious of a particular grave danger or risk of harm, and the danger or risk must be a foreseeable and probable effect of the conduct. Second, the particular danger or risk of which the defendant was subjectively conscious must in fact have eventuated. Third, a defendant must have disregarded the risk in deciding how to act. Fourth, a defendant's conduct in ignoring the danger or risk must have involved a

¹ Polinsky and Shavell (1998) provide a detailed overview of the role of malice and other possible factors pertinent to assessing punitive damages.

gross deviation from the level of care which an ordinary person would use, having due regard to all circumstances.

Reckless conduct is not the same as negligence. Negligence is the failure to use such care as a reasonable, prudent, and careful person would use under similar circumstances. Reckless conduct differs from negligence in that it requires a conscious choice of action, either with knowledge of serious danger to others or with knowledge of facts which would disclose the danger to any reasonable person.²

Despite the detail of these instructions, what is meant by recklessness is not well defined. Because a risk-free society is not feasible, presumably recklessness is a failure to strike an appropriate balance between risk and cost in situations where additional expenditures would have reduced the risk. Experimental evidence suggests that people are not able to make reliable judgments about such matters in a wide variety of legal contexts.³ Hindsight bias often intrudes: people view a risk situation after an accident as having been more preventable than it was.⁴ In addition, if corporations undertake explicit efforts to balance risk and cost through a risk analysis, the very act of explicitly making such tradeoffs in contexts where people's health is at risk may be viewed as a form of reckless disregard for individual life or limb. Ideally, however, companies should be encouraged to balance these concerns, thereby ensuring that whatever risks remain were not addressed because the costs of reducing them were too high.

Any determination of whether a company was reckless in its balancing of risk and costs requires some judgment of the resulting risk level. Can people think systematically about risks and accurately assess how hazardous were various activities? Various generic

² *Jardel Co. Inc. et al. v. K. Hughes.*

³ See, among others, Hastie, Schkade, and Payne (1998, 1999a, 1999b); Kahneman, Schkade, and Sunstein (1998); Schkade, Sunstein, and Kahneman (2000); Sunstein, Kahneman, and Schkade (1998); Viscusi (1999, 2000, 2001), and Sunstein, Schkade, and Kahneman (2000). For a synthesis of these and other results, see Sunstein et al. (2002).

⁴ Assessments of the role of hindsight for juror and judge decisions appear in Rachlinski (1998); Hastie and Viscusi (1998); Hastie, Schkade, and Payne (1999b); and Viscusi (1999).

biases in gauging risks are well documented, and in some cases may intrude on a jury's decision making. For example, the observed pattern in which people overestimate small risks may lead to an exaggerated response to a hazard. In this paper, we examine whether people have systematic biases in how they process risk information pertaining to legal cases involving accidents. We present jury-eligible individuals with the kinds of information they are likely to receive in courtroom settings, and see if they process this information in a way that enables them to form sensible judgments pertaining to the risk.

More specifically, their task is to assess the risk level based on the observed accident history, where information on the number of adverse outcomes and a measure of the level of exposure are provided.

Real world jurors typically receive information about a particular accident, past accidents of that type that have resulted from a firm's behavior, and information on the scale of the firm's operations that will generate the risk exposure.

From an analytic standpoint, the frequency of accidents is a useful concept:

$$\text{Frequency of Accidents} = \frac{\text{Number of Adverse Outcomes}}{\text{Level of Exposure}} \quad .^5 \quad (1)$$

As equation (1) indicates, at least in theory, the task of combining the number of adverse outcomes and the level of exposure to calculate the accident frequency is a straightforward arithmetic exercise.

How well do people process information pertaining to the number of accidents and the level of risk exposure? For example, a juror might be told how many accidents occurred in a given number of product deliveries. The number of accidents is the

⁵ As we discuss below, the scale of an activity, not just the frequency of accidents, is required to know how risky it is. Thus, in an everyday activity thought to be safe, one accident in 100 trials probably indicates little, but 10 in 1,000 may be significant.

numerator and the number of deliveries is the denominator when determining the rate of accidents per delivery. Do people combine this information in a reliable manner in making judgments about the recklessness of particular activities? Our hypothesis is that people are much more responsive to information about the numerator, or the total number of accidents, than they are to the denominator, which is the measure of the total level of the particular economic activity. If so, then large-scale operations would be at a disadvantage in terms of the public perception of the riskiness of their activities when compared to smaller-scale organizations. Such biases arise quite apart from a range of “deep pocket” biases, which also tend to disadvantage large firms.

The denominator blindness bias does not appear to have been fully explored previously in the literature. Discussion of public risk perception efforts often focus on the risk numerator, such as the total number of people killed by a certain cause of death, which may account for the observed overestimation of such risks.⁶ Related evidence suggests that for any given probability of winning a prize, people prefer lotteries offering more prizes. For example, people would prefer 10 chances out of 190 to win a prize to 1 chance out of 19. This result is consistent with the hypothesis that people process the numerator more reliably than the denominator.⁷ However, such a bias could also be due

⁶ In particular, Viscusi (1992), p. 7 notes: “This pattern of overestimation may surprise many participants in the smoking debate, but it is quite consistent with other evidence on highly publicized hazards. People frequently overassess widely publicized risks, whether the risks are those of smoking or the chance of being killed by lightning or a tornado. One contributor to this overassessment of the risk is that these public accounts call individuals’ attention to the adverse outcome but do not indicate the probability that the event will occur. Media accounts provide frequent and selective coverage of the numerator of the risk (e.g., the number of tornado deaths) without information on the denominator (e.g., the size of the reference population), making incorporation of public information into risk judgments difficult. The annual reports of the Surgeon General have a similar emphasis on tallies of the adverse health outcome without indicating the number of smokers or the intensity of the product’s use.”

⁷ See Denes-Raj and Epstein (1994) for discussion of experimental evidence on this issue.

to distrust of experimental lotteries and a belief that they are more likely to be legitimate if there are many prizes.

This paper reports upon an experiment in which 422 jury-eligible respondents analyzed a series of cases in legal settings. Section II describes the sample and the general information given to subjects. It also outlines our model of how judgments of the risk probability affect assessments of recklessness. The principal case studies involve a hazardous chemical transportation risk (Section III) and a pizza delivery risk example (Section IV). Our experimental methodology presents scenarios that vary different parameters that determine the accident rate. We then assess the responsiveness of individuals to these parameters in their assessing of the company's recklessness. In each instance, we find that the scale of operations is not influential, but the number of accidents is.

II. Conceptual Framework and Sample Characteristics

Judging Recklessness

We presented people with information about the number of accidents and level of exposure as indicated by the total amount of economic activity producing the accidents. Based on this information, they were asked to assess whether the defendant was reckless. From a conceptual standpoint, for any given type of activity, a firm is reckless if, in the judgment of the juror,

$$p^* > s, \tag{2}$$

where p^* is the assessed probability of an accident associated with the company's activity and s is some critical value for that activity above which the juror will believe that the

company has been reckless. The critical level s will vary depending on the character of the economic activity involved. For example, the construction industry has a fatal accident rate an order of magnitude larger than manufacturing industries, which in turn have fatal accident rates that dwarf those that face college professors. If the professors at a college had the same fatality rate as an average construction firm, then one might well conclude that the school was being operated in a reckless manner. Thus, implicit in any judgment of recklessness is some sense of how expensive it is to achieve and enhance the safety level for that particular activity, and an assessment of what the resulting risk level implies about the balance the defendant has struck between risk and cost.⁸

How jurors will respond to information about an accident history depends on the way in which they form their probability judgments. We will address three alternative approaches -- classical statistics, Bayesian analysis, and reliance solely on accident levels. In some circumstances, it may not be possible to distinguish among these approaches empirically. For purposes of the discussion below, consider an accident situation in which there have been c accidents in n trials, or an accident frequency of $f = c/n$.

Consider first the approach of classical statistics. The accident frequency is given by f as noted above. Any increase in the number of trials for any given number of accidents reduces the value of f . Increasing n also increases the precision of the estimate of the frequency, which has a 95 percent confidence interval given by

$$f \pm 1.96 \sqrt{\frac{f(1-f)}{n}} . \quad (3)$$

⁸ See Posner (1986) for a description of the Learned Hand formula and related economic issues with respect to efficient levels of care.

As n increases, there are two effects. First, f is diminished directly. Second, the confidence interval around f tightens.

Classical statistical tests using confidence intervals often are employed with respect to judgments of the efficacy of medical interventions or the frequency of side effects for pharmaceuticals, but are rarely employed in accident contexts since with small probabilities the numbers of observations tend to be quite limited. Accidents that make it into the courtrooms by their very nature tend to be rare events. Suppose that an acceptably safe driver has an accident causing serious injury every 100,000 miles on average. We decide to employ classical statistics to determine whether someone who had three accidents in 80,000 miles is a safe driver, i.e., does not have an accident rate that is statistically significantly different from that of the average driver. Because of the low probabilities involved we use the Poisson distribution rather than the normal distribution in calculating the likelihoods of various outcomes. A rate of 3 accidents per 80,000 miles falls just short of the level needed to conclude statistically that the outcome was not from an acceptably safe driver; i.e., such a driver would have a more extreme outcome (4 or more accidents) more than 5% of the time. Whatever the findings of a classical statistical test, we might not wish a driver with “merely” three accidents in 80,000 miles to escape liability, particularly if we thought at the outset that there were many reckless drivers.

The second framework we consider enables jurors to incorporate their prior risk beliefs alongside accident information in accordance with principles of rational Bayesian analysis. Bayesian analysis, unlike the classical statistics approach, incorporates prior knowledge but does not undertake any formal test of the accident risk and its associated confidence interval. For concreteness, we assume that individuals have prior beliefs

about accident frequencies that are characterized by a beta distribution. This widely employed distribution is extremely flexible; it can assume a wide variety of shapes, both skewed and symmetric. Suppose that such individuals' prior risk beliefs are tantamount to having observed b accidents out of d trials. Then from their standpoint the risk of an accident is simply b/d as their prior belief pertaining to the accident frequency. Suppose then as part of the legal case the individuals receive information that the firm's activity led to c accidents out of n trials. Based on this information, the individual's posterior beliefs are governed by

$$p^* = \frac{b + c}{d + n}. \quad (4)$$

That is, their initial beliefs get updated to $b+c$ accidents out of $d+n$ trials. Where n is great relative to d , the information on the risk levels conveyed at trial will tend to play a much more influential role than people's prior beliefs.

To see how such a learning process might work, suppose people begin with prior beliefs characterized by a value of b equal to 1 and d equal to 10,000. Thus, their prior beliefs would make an accident 0.0001 likely on a single trial. People will, however, alter their risk beliefs based on experience. Suppose that in the courtroom the individual learns that there have been 2 accidents out of 10,000 situations in which an accident might have occurred. Thus, the actual accident frequency is 0.0002. Combining this information with the individual's prior beliefs leads to a perceived probability of 0.00015.

$$p^* = \frac{1 + 2}{10,000 + 10,000} = 0.00015. \quad (5)$$

Increasing the number of trials n to 50,000, with the same number of accidents, will lead to posterior risk beliefs of

$$p^* = \frac{1 + 2}{10,000 + 50,000} = 0.00005, \quad (6)$$

a value one-third as high despite representing the same number of accidents.

The role of the number of trials n is potentially complex. For any given observed accident frequency (i.e., ratio of c/n), increasing n could either lower or raise the value of p^* depending on one's prior beliefs. For that reason, the comparisons below will be structured in a manner that yields unambiguous predictions. Holding the number of accidents, c , constant, increasing the value of n will always lower p^* . Similarly, for any given value of n , increasing the number of accidents c will always raise p^* . Experimental scenario comparisons in which both c and n vary will not offer such an unambiguous reference point and will consequently not be the focus of attention.

The third approach we consider, has no statistical validity. However, we believe that it does capture important elements of individuals' decision making behavior when judging probabilities in general and recklessness in particular. This approach looks solely to the number of accidents to determine the level of risk. Thus, jurors assess the riskiness of an activity without drawing on any prior knowledge or taking into account the number of times n the activity occurred.

We label this approach "denominator blindness." With it, jurors form a risk assessment $p(c)$ that depends solely on c , the number of accidents. This relationship could be linear, but it could also be nonlinear, as doubling c need not double $p(c)$ for jurors to be ignoring the denominator. The value of $p(c)$ may also depend on the accident context, such as whether it stemmed from a transportation accident or construction

activity. Thus, some notion of the underlying riskiness of the enterprise may enter, but not necessarily in a manner that is consistent with the formal Bayesian learning model. However, for there to be complete denominator blindness, the assessed risk should be insensitive to the value of n for any given number c of accidents.

Characteristics of the Sample and General Instructions

Our experimental design presented subjects with different case scenarios. By comparing their responses to the different cases, we were able to assess the effects of two critical case characteristics: the number of accidents and the scale of the firm's operations.

Our sample consisted of 422 jury-eligible adults. In July 2000, a marketing research firm in Austin, Texas, recruited this sample by phone. Subjects came to a central location to participate in the study. Each individual received \$40 for completing the survey, which required approximately half an hour.

The sample, as summarized in Table 1, included a broad population cross section. One-third of the sample was black or Hispanic. The educational levels were quite diverse. Just under half of the sample had either completed high school or some college education; the remainder was college or post-college graduates. The mean age was 41, and women were somewhat overrepresented in the sample. Subsequent analysis will, in many cases, control for demographic characteristics that might have influenced answers.

Before beginning the survey, each respondent received general instructions indicating that their task would involve the analysis of legal contexts:

You will consider a series of legal case situations. You will be allowed as much time as you need to review the information. Please indicate your

best judgment with respect to each question. In almost all instances there are no right or wrong answers. We are interested in your assessments, and people can feel differently about the cases.

In addition, respondents also received general guidance about punitive damages similar to the information often received as part of a jury's punitive damages instructions:

Below you will consider a series of legal cases. In every instance, the trial jury has already ordered each defendant to pay compensatory damages as full compensation for the harm suffered by the plaintiff. We would like you to imagine that you are a member of the punishment jury. Your job is to decide whether and how much each defendant should be punished, in addition to paying compensatory damages.

As a jury member, you are instructed to award punitive damages if a preponderance of the evidence shows that the defendants acted either maliciously or with reckless disregard for the welfare of others.

Defendants are considered to have acted maliciously if they intended to injure or harm someone or their property. Defendants are considered to have acted with reckless disregard for the welfare of others if they were aware of the probable harm to others or their property but disregarded it, and their actions were a gross deviation from the standard of care that a normal person would use.

Each respondent received one scenario for each type of case. Scenarios were assigned randomly to respondents so there should be no systematic differences across the different samples of respondents who considered the different case scenarios. The different cases considered involved chemical spill accidents and pizza delivery accidents. Oil and chemical spill cases have led to some of the most prominent punitive damages awards in excess of \$100 million. the threat of punitive damages for rapid pizza delivery is also quite real, as is exemplified in the \$79 million punitive damages award in *Kinder v. Hively Corp.* (No. 902-01235, Cir. Ct., St. Louis, verdict Dec. 17, 1993).

Each scenario asked jurors to make various judgments about punitive damages after being told that compensatory damages had been awarded. This approach is

consistent with many past studies.⁹ However, one could hypothesize that the frequency of punitive awards would be different if the experimental scenario had included additional components, such as a stage in which the participants first assessed liability. While such variations might be consequential, our main concern is responses to experimental scenarios in which all these elements have been held constant. The main matter of interest is not the absolute frequencies of judgments of recklessness but whether changes in the accident frequency denominator have effects on such judgments across the different scenarios in accordance with theoretical predictions.

III. Chemical Spill Accidents: The Role of Numerators and Denominators

The first set of scenarios involved a company's delivery of hazardous chemicals by truck. Respondents were given information about the number of chemical spills and the number of deliveries. In this context, the number of deliveries measures the risk exposure. Subsequently, we shall use the terminology numerator and denominator to identify the number of accidents and the level of exposure.

The respondents were also told that chemical spills endangered fish and other wildlife, and were potentially hazardous to people if they contaminated the groundwater. The appendix includes the complete text for one version of this scenario. After reading the scenario, respondents assessed the probability that the company was reckless and should therefore be subject to punitive damages. Each respondent received a series of five possible probabilities ranging from 0 to 1 on a linear risk scale with intervals of 0.25 and verbal characterizations of the probabilities. For example, $\frac{1}{2}$ was characterized as "possibly, 50-50."

⁹ See, for example, the various studies synthesized in Sunstein et al. (2002).

We used four versions of this scenario (see Table 2 for an overview). We present several comparisons that are unambiguous from a theoretical standpoint. In Scenario A, the chemical company had had two chemical spills out of 10,000 deliveries, or an accident frequency rate of 0.0002. In Scenario B there were 5 accidents out of 10,000 deliveries. Because the accident frequency is 2.5 times as great for Scenario B as for Scenario A and the number of deliveries is held constant, one would expect respondents to be more likely to judge the company as being reckless in Scenario B.¹⁰ The greater riskiness of Scenario B should be evident whichever of our three decision approaches one employs, that is whether one acts as a classical statistician, a Bayesian analyst, or falls prey to denominator blindness and bases risk beliefs solely on the number of spills.¹¹ All three formulations of risk belief, even the one that is blind to the denominator, will yield higher recklessness estimates for Scenario B than Scenario A. In the Bayesian approach, this result holds for all possible prior beliefs.

Scenario C increases the denominator from 10,000 to 50,000. Its accident frequency rate of 0.00004 is consequently one-fifth that of Scenario A, which presumably should decrease the assessed likelihood of recklessness whether employing classical statistics¹² or the Bayesian approach. Irrespective of one's prior beliefs, Scenario C implies a smaller risk than Scenario A because the number of spills is unchanged, but the number of deliveries is far greater. The first test of denominator blindness compares

¹⁰ There is no reason why if the risk posterior goes up from .001 to .002 that we should double the number of people who assign recklessness: (a) the prior plays its role, and (b) thresholds for recklessness need not have any particular distribution. Let's say that half the people had a threshold of .0009 and the other half were at .0025. Then this doubling would not affect the percentage assigning recklessness.

¹¹ Note that the 95 percent confidence interval for Scenario A is 0.0002 ± 0.0003 , while that for Scenario B is 0.0005 ± 0.0004 . These confidence intervals overlap, so that the assessed likelihood of recklessness in the two cases is not significantly different. However, in the case of Scenario B it is possible to reject the hypothesis that the accident frequency is zero.

¹² The 95 percent confidence interval for Scenario C is 0.00004 ± 0.00006 .

these two. Given blindness, i.e., if subjects do not take into account the scale of activity, the Scenarios will yield the same recklessness judgments..

The final scenario, Scenario D, involves five chemical spills out of 50,000 deliveries. Compared to Scenario C it represents an increase in the number of spills but the same number of deliveries so that Scenario D unambiguously involves a greater risk whether the approach is classical or Bayesian statistics. Compared to Scenario B, Scenario D has a greater number of deliveries but the same number of spills and is consequently unambiguously less risky. Since the changes are in the numerator not the denominator, judgments of recklessness should be higher for Scenario D than Scenario C irrespective of whether subjects are subject to denominator blindness or are classical or Bayesian statisticians. Comparing Scenario B to Scenario D provides a second test of denominator blindness, as the number of spills is 5 for both, but the number of deliveries increases from 10,000 to 50,000.

For all rational Bayesian learners irrespective of one's prior beliefs the different scenarios should produce:

Perceived Risk Levels.

Scenario B > Scenario A > Scenario C,

and

Scenario B > Scenario D > Scenario C.

Table 3 illustrates these relationships for two different prior belief (b, d) pairs: the uniform prior beliefs of (1, 2) and prior beliefs (1, 1,000,000). The first implies little information. The second indicates substantial information, but with a very low perceived risk of an accident. The one ambiguous relationship is between Scenarios A and D. As

the values in Table 3 indicate, Scenario A poses a greater risk for the uniform prior belief case, whereas Scenario D implies a greater risk when the value of the denominator for the prior beliefs is large. Because of this ambiguity, our tests of denominator blindness will not compare Scenarios A and D.

We used the respondent's assessed probability that the company was reckless to measure the extent to which the information about the scale of the operation and the number of accidents influenced people's judgments. The mean assessed probability that the company was reckless ranged from 0.25 for Scenario A to 0.36 for Scenario B, as shown in Table 2.

Consider first the results for pairs of scenarios in which judgments of recklessness should increase for all three models of how respondents incorporate risk information. In both, the numerator increased with no change in the denominator. Scenario B, in which there are five spills, yielded an assessed probability of recklessness of 0.36 as compared to 0.25 for Scenario A, which had the same risk denominator of 10,000; this difference is statistically significant.¹³ For the two scenarios in which there were 50,000 deliveries, the change in the number of spills from two in Scenario C to five in Scenario D yields a somewhat smaller increase in recklessness risk, from 0.26 to 0.33, which also proves statistically significant.¹⁴ Thus, the relationships that are expected to hold under all three models of risk beliefs do hold, which provides a test of the validity of the experiment.

¹³ In particular, the t-value is 2.96, which is statistically significant at the 95% confidence level, two-tailed test.

¹⁴ The calculated t-value for this comparison was 1.74, which is statistically significant based on a one-sided test at the 95% confidence level, which seems appropriate given that an increased number of spills should boost the assessed probability of recklessness rather than decrease it. It is noteworthy that this smaller increase in the recklessness estimate mirrors the smaller increase in the accident frequency, which is 0.00006 for Scenario D as compared to C and 0.0003 for Scenario B as compared to Scenario D.

We turn now to tests of denominator blindness, to determine what happens if the denominator in the risk frequency expression changes while holding the numerator fixed. As conjectured, significant changes in the denominator failed to produce any significant differences in the likelihood that the company would be considered reckless. In Scenarios A and C, the number of spills is two, but increasing the number of deliveries from 10,000 (A) to 50,000 (C) altered the probability of being assumed reckless from 0.25 to 0.26, which is a difference that is not statistically significant.¹⁵ Likewise, the shift in the number of deliveries from 10,000 to 50,000 in Scenarios B and D, which both involve five spills, altered the recklessness estimate from 0.36 in Scenario B to 0.33 in Scenario D; this difference is also not statistically significant.¹⁶ Across all scenarios, shifts in the number of spills increase the assessed probability of recklessness, but changes in the number of deliveries do not.

The relationship between the frequency of accidents and the respondents' assessments of the probability that the company was reckless is intriguing. The recklessness assessment increased much less than proportionally. For example, Scenario B has an accident frequency that is 2.5 times as great as Scenario A, and respondents are 1.4 times as likely to believe that the company was reckless. This less than proportional response would be expected with Bayesian learning models if people had strong prior beliefs on the likelihood of recklessness.

The personal characteristics of our respondents turned out to affect their propensity to find recklessness, as is shown in Table 4 and in a subsequent experiment. For example, Hispanics are 0.1 more likely to find recklessness, hence award punitive

¹⁵ In particular, the calculated t-statistic is 0.41.

¹⁶ The calculated t-statistic is 0.77.

damages, than are whites, the omitted group. This 0.1 represents a 28-40% increase over the base rate. Respondents who have some college or are college graduates are less likely to award punitive damages. Cigarette smokers, who have revealed through their decision to smoke a greater willingness to incur risks, are less likely to award punitive damages, by a probability of 0.09, a significant difference. Accepting more risky behavior by the company is consistent with smokers' own risk-taking patterns. Seatbelt use, however, does not predict any statistically significant difference.

That personal characteristics affect risk judgments, however, in no way diminishes our central results about the number of accidents being influential and the level of exposure being overlooked when evaluating recklessness. Consider the coefficients for the three scenarios in Table 4, which show the impact of the scenario relative to Scenario A (the base or omitted case). In Scenario B, in which there are five spills out of 10,000 deliveries, respondents had a 0.12 higher probability of awarding punitive damages than in Scenario A, with 2 spills out of 10,000 deliveries. Similarly, respondents had a 0.08 higher probability of awarding punitive damages in Scenario D with five accidents out of 50,000 deliveries, than in Scenario A. Both B and D multiply the numerator of the risk calculation by five, which significantly increases the likelihood that punitive damages will be awarded.

When it is the denominator that has shifted, however, results are quite different. One cannot reject the hypothesis that the Scenario B and Scenario D coefficients are identical, i.e., increasing the number of trials from 10,000 to 50,000 doesn't matter. Similarly, there is no statistically significant effect for Scenario C, in which the

denominator is changed by a factor of five relative to Scenario A. Thus, controlling for personal characteristics, the denominator blindness effects continue to hold.

Table 5 presents the key coefficients from two different regression results. Estimates of the personal characteristic variables are not reported because they closely parallel the findings in Table 3. The first specification reports the simple regression of the probability of awarding punitive damages on the number of deliveries and the number of spills. The number of deliveries has no statistically significant effect and has a negligible influence on the assessed probability of recklessness. In contrast, the increased number of spills boosts the assessed probability of recklessness by 0.03 per spill. The second specification in Table 5 regresses the natural logarithm of the probability that the company was found reckless against the log of the number of deliveries and the log value of the number of spills.¹⁷ For this formulation, which is commonly used in empirical analysis, the logarithm of the assessed risk should be positively related to the logarithm of the number of accidents and negatively related to the logarithm of the number of deliveries. In our calculations, however, we find that the number of deliveries does not play a statistically significant role, but the number of spills is statistically significant.

The consistent pattern that emerges is that the observed accident frequency is not influential, but rather the absolute number of accidents. This result holds controlling for personal characteristics and is true for both specifications in Table 5. The level of

¹⁷ This formulation would be appropriate if people based their risk assessments solely on the information provided in the survey, using the formula

$$p^* = \frac{\text{Number of Accidents}}{\text{Number of Deliveries}}.$$

Then

$$\ln p^* = \ln (\text{Number of Accidents}) - \ln (\text{Number of Deliveries}).$$

economic activity that generates a series of accidents plays an insignificant role in respondents' assessments of recklessness.

IV. Pizza Delivery Accidents: Does the Scale of Operations Matter?

We conducted a second set of experiments involving a quite different setting, namely low-consequence automobile accidents arising out of pizza deliveries. This enables us to determine whether our earlier results generalize to commonplace settings, or whether they reflect the sensitive issue of chemical spills.

The experimental design held the number of accidents constant at three accidents per scenario, but varied the number of pizza locations to assess whether respondents would be sensitive to this manipulation of the scale of operations.

The appendix includes a copy of a representative pizza delivery operation scenario. The risk was that of automobile accidents that arose while a driver for the pizza chain was delivering pizzas. In each case there was property damage to vehicles but no personal injury. The scenarios asked respondents to assess the probability that the company called Best Pizza was reckless. A separate question asked respondents to rate the importance of different kinds of information, which helps us determine whether the scale of operations influenced their thinking.

Table 6 summarizes the experimental design. In each instance there were three accidents. In Scenario A the firm was a local firm with an unspecified number of locations. Scenario B indicates that the firm is local but has 15 locations, whereas in

Scenario C the local firm has two locations. Presumably, the decrease in the number of locations should make liability judgments more likely, as the accident rate is 7.5 times as great for Scenario C as for Scenario B. In Scenario D there are 15 locations, as in Scenario B, but the company is a national chain, which may be a less sympathetic defendant. Respondents may also view the national chain, e.g., as having a different activity level as being a large-scale enterprise no matter how many locations it has in the area.¹⁸

The assessed probabilities of recklessness in this example ranged from 0.41 in Scenario B to 0.48 in Scenario A. These assessed values of the probability of reckless behavior are higher than for the hazardous chemical delivery scenario.

We consider first the results for the scenarios in which the number of locations is specified. The dramatic increase in the number of accidents per location from Scenario B to Scenario C increases the mean assessed probability of recklessness modestly, from 0.41 to 0.46, a difference that is statistically significant based on a one-tailed test but not a two-tailed test.¹⁹ The assessed probability of recklessness in Scenario C is almost identical to that in Scenario D even though the risk levels differ by a factor of 7.5.²⁰ That comparison involved not only a change in the number of locations but also a shift in the identity of the firm from a local to a national firm. We isolate the role of a national firm by comparing Scenarios B and D, for which the number of accidents and number of

¹⁸ Scenarios B, C, and D specified the number of locations, but none of the scenarios specified the level of activity per location.

¹⁹ In particular, the calculated t-statistic is 1.45, which falls short of statistical significance based on a one-tailed t-test at the 95% confidence level.

²⁰ The calculated t-statistic for this comparison is 0.37.

locations is identical. The shift to a national firm increases the assessed probability of recklessness from 0.41 to 0.47, a statistically significant difference.²¹

If the number of locations is unspecified, as in Scenario A, then the mean assessed probability of recklessness reaches its highest value of 0.48. This estimate is statistically different only from that in Scenario B, which has the lowest accident frequency rate, 0.2 per location, for a local firm.²² In short, and parallel to our earlier results about chemical spills, the number of locations did not influence assessments of recklessness, despite its immediate link to level of exposure, the denominator of frequency of accidents.

We wished to determine whether personal characteristics affected recklessness assessments in the pizza case as they did with chemical spills. Table 7 reports a regression analysis that parallels Table 4. The results in Table 7 examine how various personal characteristics affected respondents' assessments of the probability of recklessness. Female respondents assess a greater degree of recklessness, as do Hispanic respondents and respondents who are in the other nonwhite group. The omitted education group variable consists of those with no more than a high school education, and this group assesses a greater degree of recklessness than do the three included education group variables for different levels of college education.

The omitted scenario indicator variable is that for Scenario A. Only Scenario B has a statistically significant influence, which implies a negative effect on the assessed probability of recklessness of 0.08. Being a local firm with a large number of locations proves to have some influence, though not perhaps as stark as one might expect based on

²¹ The calculated t-statistic is 1.79, which is statistically significant at the 95% confidence level, based on a one-tailed test only. This result is plausible if one acts with the working hypothesis the jurors will be more likely to assess recklessness if the firm is not local.

the change in the number of accidents per location. Moreover, the comparable risk performance of the national firm in Scenario D does not play a significant role.

V. Conclusion

Judging the magnitude of a risk -- how often an accident occurs per unit of exposure -- is essential to determining whether the party responsible for an accident was reckless. Until one knows whether a risk is consequential or trivial, it is impossible to assess whether efforts to address the risk were adequate. This kind of concern arises not only with respect to liability judgments but also with respect to regulatory policy. For example, the U.S. Supreme Court has ruled that the Occupational Safety and Health Administration can only regulate risks that are judged to be “significant”; any judgment of significance necessarily must entail some consideration of the frequency with which the risk occurs.

To properly assess a risk, one must investigate the probability of various adverse consequences. The risk of an accident consists of two components, the number of adverse accidental outcomes divided by some measure of the economic activity that generates the accident. Thus, a primary task is to construct a measure of the accident frequency, such as the risk of automobile accidents per 100,000 miles driven or the probability that any given launching of a space shuttle will lead to a fatality.

The experimental evidence presented here indicates that people often do quite badly in making such judgments even when presented with all the information they need to assess accident frequency. The number of accidents influences assessments of

²² The pertinent t-test for Scenario A in comparison to the other scenarios are 1.94 for Scenario B, 0.46 for Scenario C, and 0.08 for Scenario D.

recklessness, but people tend to ignore or give slight attention to information pertaining to the scale of the economic activity, the denominator of risk frequency. That we detected these biases does not mean that jurors cannot be educated to think more analytically about risk frequency issues. However, our results suggest that eliminating such biases in risk belief is an important task that should be addressed in order to promote sounder judgments of liability and of risk levels themselves.

Table 1
Sample Characteristics
(422 Observations)

	Mean (Standard Deviation)
Age	41.31 (12.34)
Female	0.59 (0.49)
White	0.63 (0.48)
Black	0.12 (0.33)
Hispanic	0.20 (0.40)
Other nonwhite races	0.05 (0.21)
High school	0.14 (0.34)
Some college	0.32 (0.47)
College grad	0.36 (0.48)
Professional degree	0.17 (0.38)
Smoker	0.15 (0.36)
Seatbelt user	0.89 (0.32)

Table 2
Assessed Probability of Recklessness for Chemical Spills, by Scenario^a

Case Scenario	Number of Spills	Number of Deliveries	Accident Frequency	Recklessness Estimate	
				Mean	Std. Error of Mean
A	2	10,000	0.0002	0.25	0.02
B	5	10,000	0.0005	0.36	0.03
C	2	50,000	0.00004	0.26	0.02
D	5	50,000	0.0001	0.33	0.03

^aThe question asked of respondents was: “How likely do you think it is that Apex [Chemical Company] was reckless in its delivery operations and hence should be subjected to punitive damages?”

Table 3
Summary of Risk Beliefs for Illustrative Priors

Prior Belief Parameters (b, d)	Risk Beliefs for Scenarios			
	A 2 Accidents 10,000 Deliveries	B 5 Accidents 10,000 Deliveries	C 2 Accidents 50,000 Deliveries	D 5 Accidents 50,000 Deliveries
(1, 2)	0.0003	0.0006	0.00006	0.00012
(1, 1,000,000)	3.0×10^{-6}	6.0×10^{-6}	2.9×10^{-6}	5.7×10^{-6}

Table 4
Regression of the Assessed Probability of Recklessness for Chemical Spills on Personal Characteristics

Variable	Coefficient (Standard Error)
Constant	0.323* (0.073)
Age	3.22E ⁻⁴ (0.001)
Female	0.040 (0.027)
Black	0.024 (0.042)
Hispanic	0.102* (0.035)
Other nonwhite races	-0.005 (0.065)
Some college	-0.101* (0.042)
College graduate	-0.089* (0.042)
Professional degree	-0.027 (0.049)
Smoker	-0.091* (0.038)
Seatbelt user	-0.034 (0.043)
Scenario B (5 spills; 10,000 deliveries)	0.119* (0.037)
Scenario C (2 spills; 50,000 deliveries)	0.027 (0.038)
Scenario D (5 spills; 50,000 deliveries)	0.076* (0.037)

*Coefficient is significant at the 95% confidence level, two-tailed test.

Table 5
 Probit Regression of the Assessed Probability of Recklessness as a Function of the
 Number of Spills and Deliveries

Variable	Coefficient (Standard Error)	
	Probability	Ln (Probability)
Spills	0.028** (0.009)	
Deliveries	-2.14E ⁻⁷ (6.61E ⁻⁷)	
Ln (Spills)		0.066** (0.021)
Ln (Deliveries)		-0.004 (0.012)

**Coefficients are significant at the 99% confidence level, two-tailed test.

Note: Each equation also includes the demographic variables listed in Table 4 and a constant term.

Table 6
Assessed Probability of Recklessness in Pizza Delivery Operations, by Scenario^a

Case Scenario	Number of Accidents	Number of Locations	Accident Frequency	Firm	Recklessness Estimate	
					Mean	Std. Error of Mean
A	3	Unspecified	Unspecified	Local	0.48	0.03
B	3	15	0.2	Local	0.41	0.02
C	3	2	1.5	Local	0.46	0.03
D	3	15	0.2	National	0.47	0.03

^aRespondents were asked to assess whether Best Pizza was reckless in its delivery operations and did not exercise appropriate care.

Table 7
Regression of Assessed Probability of Recklessness in Pizza Delivery Operations on
Personal Characteristics and Scenarios

Variable	Coefficient (Standard Error)
Constant	0.492* (0.070)
Age	0.001 (0.001)
Female	0.053* (0.026)
Black	-0.043 (0.040)
Hispanic	0.086* (0.034)
Other nonwhite races	0.134* (0.062)
Some college	-0.081* (0.040)
College graduate	-0.074** (0.040)
Professional degree	-0.132* (0.047)
Smoker	-0.042 (0.037)
Seatbelt user	-0.016 (0.041)
Scenario B (3, 15, Local)	-0.077* (0.036)
Scenario C (3, 2, Local)	-0.017 (0.036)
Scenario D (3, 15, National)	-0.006 (0.036)

*Coefficient is significant at the 95% confidence level, two-tailed test.

**Coefficient is significant at the 95% confidence level, one-tailed test.

Appendix

Chemical Spill Accident Scenario

The *Apex Chemical Company* transports hazardous chemicals for important industrial uses. These chemicals are toxic to fish and wildlife. Moreover, if the chemicals get into the water supply or the groundwater, they can create significant health hazards for people as well. Because these chemicals are transported by truck, there is some risk of a traffic accident, which in turn can cause a chemical spill. Last year, *Apex* had 2 chemical spills out of 10,000 deliveries.

How likely do you think it is that *Apex* was reckless in its delivery operations and hence should be subjected to punitive damages? Your best estimate will do.

0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Not at All Likely	Somewhat Likely	Possibly, 50-50	Very Likely	Definitely

Pizza Delivery Accident Scenario

In calendar 1998, *Best Pizza*, a local pizza chain with 15 locations, had 3 of its employees involved in separate automobile accidents while delivering pizzas in the Austin, Texas area. Each of these accidents caused property damage to other vehicles, but no personal injury. You have been asked to assess whether the court should award punitive damages against *Best Pizza* because they believe its delivery operations were reckless. Improperly

maintained vehicles, poor worker training, or emphasis on rapid delivery schedules that compromise safety all could be classified as reckless if they led to accidents.

How likely do you think it is that *Best Pizza* was reckless in at least one of these different safety dimensions? Use the scale below to indicate the probability that *Best Pizza* was reckless and did not exercise appropriate care, based on your best guess given the information you have been given above.

Probability That *Best Pizza* Was Reckless

0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Not at All Likely	Somewhat Likely	Possibly, 50-50	Very Likely	Definitely

Rank the following different types of additional information that you would like to assist in your determination of whether punitive damages are warranted. Rate these factors from 1 to 5 with 1 being most important.

- _____ Car maintenance practices
- _____ Driver training and experience
- _____ Incentives given to driver for fast delivery
- _____ Number of deliveries
- _____ Average length of delivery trip

References

- Denes-Raj, Veronika, and Seymour Epstein. 1994. "Conflict Between Intuitive and Rational Processing: When People Behave Against Their Better Judgment," 66 *Journal of Personality and Social Psychology* 819-829.
- Fischhoff, Baruch, Sarah Lichtenstein, Paul Slovic, Stephen L. Derby, and Ralph L. Keeney. 1981. *Acceptable Risk*. Cambridge: Cambridge University Press.
- Hastie, Reid, David Schkade, and John Payne. 1998. "A Study of Juror and Jury Judgments in Civil Cases: Deciding Liability for Punitive Damages," 22 *Law and Human Behavior* 287-314.
- Hastie, Reid, David Schkade, and John Payne. 1999a. "Juror Judgments in Civil Cases: Effects of Plaintiff's Requests and Plaintiff's Identity on Punitive Damage Awards," 23 *Law and Human Behavior* 445-470.
- Hastie, Reid, David Schkade, and John Payne. 1999b. "Juror Judgments in Civil Cases: Hindsight Effects on Judgments of Liability for Punitive Damages," 23 *Law and Human Behavior* 597-614.
- Hastie, Reid, and W. Kip Viscusi. 1998. "What Juries Can't Do Well: The Jury's Performance as a Risk Manager," 40 *Arizona Law Review* 901-921.
- Kahneman, Daniel, David Schkade, and Cass Sunstein. 1998. "Shared Outrage and Erratic Awards: The Psychology of Punitive Damages," 16 *Journal of Risk and Uncertainty* 49-86.
- Kunreuther, Howard, et al. 1978. *Disaster Insurance Protection: Public Policy Lessons*. New York: John Wiley.

- Lichtenstein, Sarah, et al. 1978. "Judged Frequency of Lethal Events," 4 *Journal of Experimental Psychology* 551-578.
- Morgan, M. Granger. 1983. "On Judging the Frequency of Lethal Events: A Replication," 3 *Risk Analysis* 11-16.
- Posner, Richard. 1986. *Economic Analysis of Law, 3rd Edition*. Boston: Little Brown.
- Rachlinski, Jeffrey. 1998. "A Positive Psychological Theory of Judging in Hindsight," 65 *University of Chicago Law Review* 571-625.
- Schkade, David A., Cass R. Sunstein, and Daniel Kahneman. 2000. "Deliberating about Dollars: The Severity Shift," 1000 *Columbia Law Review* 1139-1175.
- Sunstein, Cass, Daniel Kahneman, and David Schkade. 1998. "Assessing Punitive Damages (with Notes on Cognition and Valuation in Law)," 107 *Yale Law Journal* 2071-2153.
- Sunstein, Cass, David Schkade, and Daniel Kahneman. 2000. "Do People Want Optimal Deterrence?" 29 *Journal of Legal Studies* 237-253.
- Sunstein, Cass, Reid Hastie, John Payne, David Schkade, and W. Kip Viscusi. 2002. *Punitive Damages: How Juries Decide*. Chicago: University of Chicago Press.
- Viscusi, W. Kip. 1992. *Smoking: Making the Risky Decision*. New York: Oxford University Press.
- Viscusi, W. Kip. 1998. *Rational Risk Policy*. Oxford: Oxford University Press.
- Viscusi, W. Kip. 1999 "How Do Judges Think about Risk?" 1&2 *American Law and Economics Review* 26-62.
- Viscusi, W. Kip. 2000. "Corporate Risk Analysis: A Reckless Act?" 52 *Stanford Law Review* 547-597.

Viscusi, W. Kip. 2001. "Jurors, Judges, and the Mistreatment of Risk by the Courts,"
XXX Journal of Legal Studies 107-142.

Footnotes:

1. Polinsky and Shavell (1998) provide a detailed overview of the role of malice and other possible factors pertinent to assessing punitive damages.
2. *Jardel Co. Inc. et al. v. K. Hughes*.
3. See, among others, Hastie, Schkade, and Payne (1998, 1999a, 1999b); Kahneman, Schkade, and Sunstein (1998); Schkade, Sunstein, and Kahneman (2000); Sunstein, Kahneman, and Schkade (1998); Viscusi (1999, 2000, 2001), and Sunstein, Schkade, and Kahneman (2000). For a synthesis of these and other results, see Sunstein et al. (2002).
4. Assessments of the role of hindsight for juror and judge decisions appear in Rachlinski (1998); Hastie and Viscusi (1998); Hastie, Schkade, and Payne (1999b); and Viscusi (1999).
5. As we discuss below, the scale of an activity, not just the frequency of accidents, is required to know how risky it is. Thus, in an everyday activity thought to be safe, one accident in 100 trials probably indicates little, but 10 in 1,000 may be significant.
6. In particular, Viscusi (1992), p. 7 notes: “This pattern of overestimation may surprise many participants in the smoking debate, but it is quite consistent with other evidence on highly publicized hazards. People frequently overassess widely publicized risks, whether the risks are those of smoking or the chance of being killed by lightning or a tornado. One contributor to this overassessment of the risk is that these public accounts call individuals’ attention to the adverse outcome but do not indicate the probability that the event will occur. Media accounts provide frequent and selective coverage of the numerator of the risk (e.g., the number of tornado deaths) without information on the denominator (e.g., the size of the reference population), making incorporation of public information into risk judgments difficult. The annual reports of the Surgeon General have a similar emphasis on tallies of the adverse health outcome without indicating the number of smokers or the intensity of the product’s use.”
7. See Denes-Raj and Epstein (1994) for discussion of experimental evidence on this issue.
8. See Posner (1986) for a description of the Learned Hand formula and related economic issues with respect to efficient levels of care.
9. See, for example, the various studies synthesized in Sunstein et al. (2002).

10. There is no reason why if the risk posterior goes up from .001 to .002 that we should double the number of people who assign recklessness: (a) the prior plays its role, and (b) thresholds for recklessness need not have any particular distribution. Let's say that half the people had a threshold of .0009 and the other half were at .0025. Then this doubling would not affect the percentage assigning recklessness.
11. Note that the 95 percent confidence interval for Scenario A is 0.0002 ± 0.0003 , while that for Scenario B is 0.0005 ± 0.0004 . These confidence intervals overlap, so that the assessed likelihood of recklessness in the two cases is not significantly different. However, in the case of Scenario B it is possible to reject the hypothesis that the accident frequency is zero.
12. The 95 percent confidence interval for Scenario C is 0.00004 ± 0.00006 .
13. In particular, the t-value is 2.96, which is statistically significant at the 95% confidence level, two-tailed test.
14. The calculated t-value for this comparison was 1.74, which is statistically significant based on a one-sided test at the 95% confidence level, which seems appropriate given that an increased number of spills should boost the assessed probability of recklessness rather than decrease it. It is noteworthy that this smaller increase in the recklessness estimate mirrors the smaller increase in the accident frequency, which is 0.00006 for Scenario D as compared to C and 0.0003 for Scenario B as compared to Scenario D.
15. In particular, the calculated t-statistic is 0.41.
16. The calculated t-statistic is 0.77.
17. This formulation would be appropriate if people based their risk assessments solely on the information provided in the survey, using the formula

$$p^* = \frac{\text{Number of Accidents}}{\text{Number of Deliveries}}.$$

Then $\ln p^* = \ln(\text{Number of Accidents}) - \ln(\text{Number of Deliveries})$.

18. Scenarios B, C, and D specified the number of locations, but none of the scenarios specified the level of activity per location.

19. In particular, the calculated t-statistic is 1.45, which falls short of statistical significance based on a one-tailed t-test at the 95% confidence level.
20. The calculated t-statistic for this comparison is 0.37.
21. The calculated t-statistic is 1.79, which is statistically significant at the 95% confidence level, based on a one-tailed test only. This result is plausible if one acts with the working hypothesis the jurors will be more likely to assess recklessness if the firm is not local.
22. The pertinent t-test for Scenario A in comparison to the other scenarios are 1.94 for Scenario B, 0.46 for Scenario C, and 0.08 for Scenario D.