Tailoring Negligence Standards to Accident Records

Alice Guerra and Tobias M. Hlobil

ABSTRACT
The standard 1-period, unilateral care accident model assumes identical injurers and perfect information of an individual's risk type. However, these assumptions are unlikely to hold in real-world accident scenarios. This paper considers a 2-period, unilateral care accident model in which injurers differ by probabilities of accident (their risk types) and have incomplete information about their risk types. We find that courts should optimally examine an individual's accident history to accurately infer the risk type and adjust the due level of care accordingly. We show that tailoring due levels of care in the second period affects the definition of the due level of care in the first period. When judges have access to accident records but the risk type is hidden, they should relax the due level of care for first-time offenders to generate more information about an individual's risk type, which helps to establish more efficient differentiated standards in the subsequent period.

1. INTRODUCTION
The standard unilateral care accident model is generally a 1-period model in which potential injurers are identical and have complete information about their probabilities of accident (their risk types). In this traditional framework, the optimal due level of care—which is set to minimize the

Alice Guerra is an Assistant Professor and Marie Curie Fellow at Copenhagen Business School. Tobias Hlobil is a PhD Candidate at the European Doctorate in Law and Economics, Hamburg. The authors would like to thank Emanuela Carbonara, Goran Dominioni, Luigi Alberto Franzoni, Nuno Garoupa, Steffen Juranek, Elena Kantorowicz-Reznichenko, Henrik Lando, Eric Langlais, Shay Lavie, Alain Marciano, Francesco Parisi, Daniel Pi, Giovanni Battista Ramello, Massimiliano Vatiero, Louis Visscher, and Abraham Wickelgren for helpful suggestions and seminar audiences at the 2016 annual meeting of the Italian Society of Law and Economics (Campus Luigi Einaudi of the University of Turin) and the 2015 European Association of Law and Economics annual conference (University of Vienna). The authors are extremely grateful to Edward Morrison, William Hubbard, and an anonymous referee for their constructive and insightful comments.

[Journal of Legal Studies, vol. 47 (June 2018)]
© 2018 by The University of Chicago. All rights reserved. 0047-2530/2018/4702-0021$10.00
social cost of accidents—is invariant over time and uniform across all potential injurers (see, for example, Shavell 1980, 1987). Accident records are neither available nor needed. However, these assumptions are very unlikely to hold for most real-world accidents for a number of reasons. First, accidents can repetitively occur over time, and accident records are usually available to courts and public authorities. Second, potential injurers normally differ in their probabilities of accident (for example, individuals are not equally good at driving). Third, individuals and social planners generally have incomplete information about their own risk types (see, for example, Bajtelsmit and Thistle 2008, 2009). Indeed, not all variations in accident rates can be explained by observed factors, and information about each individual’s probability of accident inevitably remains hidden.

This paper analyzes the optimal due levels of care in a more realistic 2-period, unilateral care accident scenario. We introduce three extensions to the standard tort model (see, for example, Shavell 1980, 1987). First, we move from the traditional 1-period framework to a 2-period model in which accident records become available after the first period. This allows us to consider first-time and repeat offenders, with judges evaluating the costs and benefits for different actors of consistently meeting the due level of care over time.

Second, the probability of accident is not the same for all individuals. We consider two types of potential injurers, namely, high-risk and low-risk types. For any given level of precaution, the two risk types differ in the probability of an accident: as intuition suggests, high-risk types are more likely to incur accidents than low-risk types. Low-risk types can be interpreted as more talented or more experienced and high-risk types as less talented or less experienced. For example, for a given level of care, experienced or more talented surgeons are less likely to cause harm with a diagnosis or surgery than inexperienced or less talented surgeons. Similarly, experienced drivers such as taxi drivers are less likely to be involved in accidents compared with inexperienced private drivers.

The third extension represents the key analytical feature of our analysis: information about an individual’s risk type is hidden not only to the social planner but also to each potential injurer. Only the proportions of each risk type in the population are known.¹ This setup can be interpreted under the lens of behavioral biases about risk perception. As several ex-

¹. Bajtelsmit and Thistle (2008, 2009) use a similar framework to analyze the demand for liability insurance.
perimental and psychological research studies reveal, individuals generally ignore and routinely misjudge their own accident risks, for example, because of optimistic or pessimistic dispositions, lack of knowledge, or an inability to accurately process available information (for example, Viscusi and Zeckhauser 2004; Viscusi and Evans 2006).

By relying on this theoretical framework, we bring to light two novel considerations of the optimal definition of due levels of care. First, we find that, in the presence of heterogeneous risk and imperfect information, negligence standards should be optimally tailored to accident records. Courts should examine the defendant’s history of accidents to more accurately infer the type of risk that he or she poses and adjust his or her due level of care accordingly, as insurance companies do to set individualized premiums. The rationale for tailoring due levels of care is indeed quite similar to the practice in the insurance industry of setting individualized premiums or assigning demerit points on the basis of past reported behavior (see, for example, Bourgeon and Picard 2007; Frihe 2008; Dionne, Michaud, and Pinquet 2013).^{2}

More interestingly, we show that the possibility of tailoring due levels of care in the second period affects the definition of the due level of care in the first period. This is the less obvious and more novel insight, which we label the first-period effect: the efficiency of setting one-size-fits-all due levels of care at a level lower than the literature has suggested. The current literature commonly argues that when information about heterogeneous risk is hidden, the optimal due level of care should be at an averaged, static level. However, from an intertemporal perspective, courts might deviate from this static optimization. Our analysis reveals that by relaxing the due level of care for first-time offenders, more accidents occur in the first period, but more information is subsequently gathered about the differential propensities to create accident risks. This added information later helps to establish more efficient differentiated standards. Counter to intuition, inducing more accidents in the first period may be socially valuable since it can help to minimize the social costs associated with accidents across time. The logic is intuitive: accidents today can reduce the total social cost tomorrow. Essentially, the cost of an accident is the price of screening which potential injurers are more likely to be the higher- and lower-risk types.

By considering heterogeneous injurers, this paper is linked to stud-

---

2. Closely related to our paper are Crocker and Doherty (2000) and Bajtelsmit and Thistle (2008, 2009), which focus on the incentives to purchase liability insurance.
ies analyzing the implications of individual heterogeneities in negligence law. This literature argues that when the social planner cannot observe individual heterogeneities, the optimal solution is to set an averaged, static due level of care (for example, Ganuza and Gomez 2005; Miceli 2006; Bajtelsmit and Thistle 2009; Endres and Friese 2014; Korsmo 2016). The relevance and timely nature of the present paper is underscored by the fact that the discussion on personalizing due levels of care in the legal literature has been revived by Ben-Shahar and Porat (2016), in which past information about similar behavior is discussed as a proxy for potential injurers’ abilities and tendencies with respect to risk creation and precaution taking. This suggests that a driver’s traffic violation or an instance of medical malpractice by a physician can be used by the court in personalizing the standard of care.

The idea of tailoring due levels of care to accident records bears some formal similarities to the practice in criminal law of punishing repeat offenders more severely than first-time offenders. Instead of applying escalating penalties, tort law generally imposes punitive damages or civil monetary penalties on repeat tortfeasors (see, for example, Polinsky and Shavell 1998a; Rhee 2012). A major factor in imposing punitive damages is whether harmful conduct involved repeated actions or was an isolated incident, as the Supreme Court found in State Farm Mutual Automobile

3. The heterogeneities between individuals considered in the literature relate to the costs of taking care and the wealth of potential injurers (Arlen 1992; Miceli and Segerson 1995), the ability to take and cost of taking care (Ganuza and Gomez 2005), and the difference in relative gain from potentially harmful activities (Emons 1990a, 1990b; Emons and Sobel 1991). See also the principle of accident proneness (James and Dickinson 1950).

4. The previous literature has recognized the inefficiency of applying the reasonable-person standard in the presence of heterogeneous parties, although it has defended its application when parties’ heterogeneities are costly to assess (Diamond 1974; Landes and Posner 1987; Shavell 1987).

5. Ben-Shahar and Porat (2016) distinguish between similar past behaviors such as several violations in similar circumstances and different past behaviors. Intuitively, it is easier to process information about the defendant’s risk-creation and precaution-taking tendencies from similar past behaviors.

6. This is a common practice in many countries, not only for crimes such as theft and murder but also for criminal violations of environmental and labor regulations (see, for example, Polinsky and Shavell 1998a; Chu, Hu, and Huang 2000; Dana 2001). Notwithstanding this common practice, answers to the question of how repeat offenders should be punished are mixed. Under some circumstances, enhancing the efficiency of deterrence may require more severe punishment for repeat offenders, the use of prior convictions to set individualized sanctions, or both (see, for example, Dana 2001; Shepherd 2002; Emons 2003, 2007).
Tailoring punitive damages to the defendant’s recurring behavior is explicitly mentioned in several state laws, such as repeat violations of consumer protection laws (D.C. Code Ann., sec. 28-3813 [1981]), unlawful trade practices (Idaho Code, sec. 48-608), abusive recruitment practices (Iowa Code, sec. 91E.4), and profiteering in drug prescriptions (Me. Rev. Stat. Ann., title 22, sec. 2697). However, unlike repeat offenders in criminal law, repeat offenders in civil law have received little attention in the standard tort model. Exceptions include Levmore (1990) and Polinsky and Shavell (1998a), which analyze whether sanctions should depend on prior offenses. We contribute to this literature by showing that the current system of imposing punitive damages or greater sanctions on repeat offenders is less efficient at minimizing social costs than our proposal of tailoring negligence standards to accident records.

The remainder of this paper proceeds as follows. Section 2 characterizes the socially optimal due levels of care and analyzes whether tailoring due levels of care to accident records outperforms awarding punitive damages to repeat offenders. Section 3 discusses the practical relevance of the results, including a numerical example that provides the basic intuitions for the general findings. Section 4 concludes and suggests possible directions for future research. Appendix A presents the theoretical model, and Appendix B sketches the proofs of the main results.

2. TAILORING NEGLIGENCE STANDARDS

In this section, we analyze whether and how due levels of care should be tailored to past accidents.\(^7\) We consider a 2-period, unilateral accident model in which potential injurers differ in their probabilities of accident (their risk types). For comparative reasons, we discuss the socially optimal due levels of care in three cases: when information about an individual’s risk type is not hidden and accident records are available; when information about an individual’s risk type is hidden to potential injurers and the social planner and accident records are available, which is the main case of interest in this paper; and when information about an individual’s risk type is hidden to potential injurers and the social planner

---

7. See also Dobbs (1993, sec. 3.11[2]), which stresses that “[r]epeated misconduct or a policy of misconduct . . . is often an element in punitive damages cases.”

8. As is common in the literature, we assume due levels of care to be set at the socially optimal levels. For an analytical description of the model, see Appendix A.
but accident records are not available. The first and third cases represent the two benchmark cases of perfect and imperfect information already studied in the literature in 1-period models. Our interest is in comparing them with the novel second case.

When information about an individual’s risk type is not hidden and accident records are available, the first-best solution is implemented by tailoring due levels of care to each potential injurer’s risk type—and only to the risk type—in each period. In this benchmark case of perfect information, no information about past accidents is required to reach the global minimum of the total social cost function. At the other extreme, when information about an individual’s risk type is hidden and accident records are not available, conventional results in the literature of tort law and economics propose an averaged due level of care as the second-best solution (for example, Ganuza and Gomez 2005).

Instead, our analysis shows that when information about an individual’s risk type is hidden but accident records are available, the second-best solution is implemented by tailoring due levels of care to incidents of past accidents, whereas the uniform, averaged due level of care characterizes the third-best solution. The intuition is clear: records of past accidents convey crucial information about an individual’s risk type and help to define more efficient due levels of care. In particular, when the first-best solution is not achievable, the socially efficient standard of due care in the second period should be higher (lower) for potential injurers who were (were not) involved in an accident in the first period than when information about past accidents is hidden. This result is summarized in proposition 1.

Proposition 1: Tailoring Negligence Standards to Accident Records. In the second period, when information about an individual’s risk type is hidden but accident records are available, the second-best solution is implemented by tailoring due levels of care to incidents of past accidents. The second-best due levels of care are higher for repeat offenders than for first-time offenders.

As intuition would suggest, the rationale for proposition 1 is that the group of potential injurers who had an accident in the first period have a higher marginal probability of being involved in another accident in the second period. For this reason, it is socially efficient that the due level of care in the second period is set higher than the standard applied in the first period. The same reasoning holds true for those who did not have an
accident in the first period. On average, this group has a lower marginal probability of being involved in accidents in the second period. Therefore, the due level of care for this group in the second period should be set lower than the standard applied in the first period.

More interestingly, the possibility of personalizing due levels of care in the second period influences the definition of the due level of care in the first period. This follows by comparing the optimal due level of care for the first period in two cases: when due levels of care in the second period are tailored to past accident experiences (second-best solution) and when a uniform due level of care is set over time (third-best solution). The result is summarized in proposition 2.

**Proposition 2: Information-Forcing Negligence Standard.** In the first period, when information about an individual’s risk type is hidden but accident records are available, the second-best due level of care should be optimally set lower than the averaged, static due level of care. This allows for gathering more information about individuals’ risk types and establishing more efficient differentiated standards in the second period.

Proposition 2 highlights the simple and important first-period effect: by deviating downward from the static, one-size-fits-all due level of care suggested in the previous literature, more accidents occur and more information is subsequently learned about the differential propensities to create accident risks. This added information later helps to establish more efficient differentiated standards, thus reducing social costs in the second period.9

Since the tort system already treats first-time and repeat offenders differently through punitive damages or escalated administrative sanctions (for example, suspension or revocation of a driver’s license), one might ask whether and when—that is, under which conditions—tailoring negligence standards outperforms awarding punitive damages or greater civil monetary sanctions for repeat offenders.10 Consider the current legal system in which the due level of care is averaged and time invariant and civil

---

9. This result has some similarities with prior models showing that deviations from static optimality are justified because of the information-forcing effect that these deviations have. For example, information-forcing defaults in contract law compel individuals to state terms in their contracts expressively. Among others, see Porat and Strahilevitz (2013), which explores the role of Big Data in the personalization of default rules. We thank the reviewer for this insight.

10. Our analytical argument relies solely on incentives for precaution without referring to administrative cost savings.
penalties or punitive damages depend on whether the defendant has a record of prior offenses. By comparing the respective social costs, we find that tailoring negligence standards is more efficient than awarding punitive damages: the information-forcing role of tailoring negligence standards is crucial in minimizing total social costs over time and across risk types. Lowering the first-period due level of care increases the costs of accidents in the first period, which yields informational benefits in subsequent periods. In this case, simply treating repeat offenders disadvantageously by awarding punitive damages in the second period—while keeping the negligence standard fixed at the average level—would neither yield any informational benefit nor enhance deterrence. We refer to this result as the desirability of tailoring negligence standards, and it is formally derived in Appendix B.

3. PRACTICAL MEANING

We now discuss the practical meaning of our findings. Beyond the novel, theoretical contribution of this paper, a crucial question that emerges is when—that is, under which circumstances—courts could in practice deviate from optimal standards downward for the purpose of generating better information for the future. Are there contexts in which the information-forcing aspect of the first-period effect is sufficiently relevant that judges would follow this suggestion in practice? The answer is a qualified yes: the first-period effect is particularly relevant in cases in which the population of interest is heterogeneous in risk type, whereby each actor’s risk type is hidden information.

Moreover, the fine-grained tweaking that we propose would be perceptible in situations in which the negligence standard is not vaguely defined in practice but rather precisely determined by the law that was violated, for example, by safety statutes, regulations, or municipal ordinances. Traffic violations are the most obvious examples. In these cases, the judge still has to evaluate whether the conduct violated the statute and caused the accident, although the due level of care is precisely determined: it was either violated or it was not, with no intermediate possibilities. For instance, the standard of due care in a speeding case assumes that a reasonable person would not drive a vehicle over the speed limit. Our findings suggest that the decision concerning whether to squarely adhere to a legislative standard should lie within the discretion of the court.
which should be free to deviate from statutory or regulatory limits wherever they appear unreasonable given the circumstances of the accident.\textsuperscript{11}

As an example, let us consider speeding. Speeding—which is a persistent, costly social problem—is particularly well suited to elucidate the practical meaning of our findings for the following reasons.\textsuperscript{12} Some drivers have a much higher risk of crash or injury when driving above the speed limit than others. Moreover, drivers do not generally know their own risk of accident: some individuals can drive safely above the speed limit, while others are at higher risk of accident even when driving below the speed limit. Clearly, in this case one size does not fit all. However, information about an individual's risk type is normally hidden not only to the social planner but also to each driver. Thus, national or municipal legislation generally imposes uniform and static speed limits on all drivers.

Our paper proposes a negligence system in which low-risk and high-risk drivers are better identified and efficiently targeted. The fact that a driver has a record of prior speeding-related accidents may provide information about his or here type of risk, such as a higher-than-average propensity to cause accidents when speeding. This information-based reason justifies tailoring negligence standards to accident records. Moreover, it is possible to go one step further. Gathering information about risk types comes with a price, namely, a speed limit for first-time offenders that is higher than the static, averaged speed limit. This will increase accident costs in the first period but reduce the total social costs over time.

To provide clearer explanations of our results, we present the following numerical example. We use a unilateral care scenario in which the population of potential injurers comprises high-risk and low-risk drivers. Each driver lives for 2 periods, and in each period each driver may cause an accident that results in harm \( d = 100 \). Suppose that there are 33 low-risk drivers for every seven high-risk drivers in this population. The proportion of low-risk drivers, \( \lambda_L \), is equal to .825, and the proportion of high-risk drivers, \( \lambda_H \), is equal to .175.

\textsuperscript{11} In this respect, our paper contributes to the extensive debate on judicial discretion in claims involving negligence per se. See, among others, the discussion in Yowell (1997).

\textsuperscript{12} Speeding is one of the most important factors in the causation and severity of many accidents (see, for example, Aarts and van Schagen 2006; Fitzpatrick, Rakasi, and Knodler 2017). For instance, in 2016 speed-related fatal accidents accounted for 27 percent of all traffic fatalities in the United States. For extensive records on speeding-related fatalities, see the reports published by the National Highway Traffic Safety Administration (for example, National Center for Statistics and Analysis 2018).
Let us describe the traditional, static scenario. In each period, each potential injurer can take different levels of care \( x \) at a unit cost of 1 to reduce the probability of an accident. In our practical example, the level of care is related to the rate of speed. A speed of 65 miles per hour (mph) corresponds to a level of care \( x \) equal to 0, and a speed of 15 mph corresponds to a level of care \( x \) equal to 10. Let \( p_L(x) \) and \( p_H(x) \) denote the probability of accident of a low- and a high-risk type, respectively. The probability of an accident decreases as levels of care increase, that is, as speed decreases.\(^{13}\) The relationship among the level of care, the speed, the probability of accident, and social costs is presented in Table 1.

Ideally, under perfect information about an individual’s risk type, the first-best solution would be to tailor the speed limit to each risk type, that is, 40 mph for a low-risk driver \((x^{*}_{L} = 5)\) and 25 mph for a high-risk driver \((x^{*}_{H} = 8)\). In each period, the social cost of accidents would then be equal to 5.50 for low-risk drivers and 9.63 for high-risk drivers (see Table 1). The total social cost would then be equal to 6.22 \((= .825 \times 5.50 + .175 \times 9.63)\) per period and thus 12.44 over 2 periods.

However, information about an individual’s risk type is normally hidden. It is a standard result in the economic analysis of accidents and liability that a due level of care invariant over time and averaged across individuals induces the second-best level of precaution. This leads to an averaged speed limit equal to 35 mph \((\bar{x}^{*} = 6)\). The social cost of accidents would then be 7.05 per period and thus 14.107 over 2 periods, which is obviously higher than the social costs under perfect tailoring.

For the sake of clarity, consider two drivers: driver A is a low-risk type, and driver B is a high-risk type. In case of perfect information, the social planner would optimally set personalized speed limits at 40 mph for driver A and 25 mph for driver B. However, information about their risk types is hidden, and the social planner sets a speed limit averaged at 35 mph. Suppose that the two drivers are involved in separate, independent, but identical accidents: they each hit a legally parked vehicle, and the owner of the vehicle sues. Driver A was driving at 40 mph, and driver B was driving at 30 mph. Given the speed limit of 35 mph, driver A is deemed negligent, whereas driver B is considered nonnegligent. It is clear that the uniform speed limit targets the low-risk driver—who was driving safely at his or her optimal speed level—rather than the high-risk driver—who was instead driving unsafely above his or her optimal speed.

\(^{13}\) The functional form of the probability function used in this example is \( p_L(x) = 2/(1 + e^{1.2x}) \) for low-risk types and \( p_H(x) = 2/(1 + e^{.6x}) \) for high-risk types. Other functional forms and parameters can be used for the probability functions (for example, logistic functions), subject to the assumptions described in Appendix A.
level. The result is that low-risk drivers will be overdeterred and more frequently sanctioned, while high-risk drivers will be underdeterred and more likely to repeat the same offenses over time. This overall implies higher social costs.

We now expand this setting so that judges may use accident records to tailor due levels of care in the second period. Suppose that the social planner keeps the averaged speed limit in the first period, that is, 35 mph. In the second period, there are two groups of potential injurers: those who had an accident in the first period and those who did not. It is possible to update the probabilities of high or low risks on the basis of prior accidents. The probability of being a high-risk type given that an accident occurred in the first period is equal to .883 (\(=\frac{.175 \times .0532}{.175 \times .0532 + .825 \times .0015}\)), whereas the probability of being a low-risk type given that an accident occurred in the first period is equal to .117 (\(=\frac{.825 \times .0015}{.175 \times .0532 + .825 \times .0015}\)). Similarly, the probabilities of being a high-risk or a low-risk type given that an accident did not occur in the first period are equal to .167 and .833, respectively.

It is possible to show that in the second period the optimal speed limit should be 25 mph for repeat offenders (due level of care equal to 8) and 40 mph for first-time offenders (due level of care equal to 5).\(^{14}\) Therefore,

\[14. \text{Formally, let } x_{2|A}^{**} \text{ and } x_{2|NA}^{**} \text{ denote the optimal due levels of care for the group that incurred an accident in the first period and the group that did not, respectively. The optimal due levels of care are derived from the following optimization problems: } x_{2|A}^{**} \in \arg \min \{1.17 p_H(x_{2|A}) + 0.883 p_H(x_{2|A})\} 100 + x_{2|A}, \text{ and similarly, } x_{2|NA}^{**} \in \arg \min \{0.833 p_H(x_{2|NA}) + 1.167 p_H(x_{2|NA})\} 100 + x_{2|NA}. \]

### Table 1. Accidents and Social Costs

<table>
<thead>
<tr>
<th>Speed (mph)</th>
<th>Care (x)</th>
<th>Probability of Accident (p_H(x))</th>
<th>Social Costs (x_H^*)</th>
<th>(x_L^*)</th>
<th>(\bar{x}^{**})</th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
<td>0</td>
<td>1</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>60</td>
<td>1</td>
<td>.708</td>
<td>71.868</td>
<td>47.295</td>
<td>51.595</td>
</tr>
<tr>
<td>55</td>
<td>2</td>
<td>.462</td>
<td>48.295</td>
<td>18.634</td>
<td>23.825</td>
</tr>
<tr>
<td>50</td>
<td>3</td>
<td>.283</td>
<td>31.370</td>
<td>8.319</td>
<td>12.353</td>
</tr>
<tr>
<td>45</td>
<td>4</td>
<td>.166</td>
<td>20.634</td>
<td>5.632</td>
<td>8.257</td>
</tr>
<tr>
<td>40</td>
<td>5</td>
<td>.094</td>
<td>14.485</td>
<td>5.494</td>
<td>7.067</td>
</tr>
<tr>
<td>35</td>
<td>6</td>
<td>.053</td>
<td>11.319</td>
<td>6.149</td>
<td>7.053</td>
</tr>
<tr>
<td>30</td>
<td>7</td>
<td>.029</td>
<td>9.954</td>
<td>7.044</td>
<td>7.554</td>
</tr>
<tr>
<td>15</td>
<td>10</td>
<td>.004</td>
<td>10.494</td>
<td>10.001</td>
<td>10.087</td>
</tr>
</tbody>
</table>

**Note.** The values for social costs are \(x_H^* \in \arg \min p_H(x)d + x, \) \(x_L^* \in \arg \min p_L(x)d + x, \) and \(\bar{x}^{**} \in \arg \min [\lambda_H p_H(x) + \lambda_L p_L(x)]d + x.\)
the social costs for the drivers who had an accident in the first period are equal to 9.44 (=.883 × 9.63 + .117 × 8.01). Similarly, the social costs for the drivers who did not have an accident in the first period are equal to 7.00 (=.167 × 14.49 + .833 × 5.50). By pooling the two groups, in the second period the total social costs are thus equal to 7.026.15 The social costs in the first period are equal to 7.053, and thus the overall social costs over the 2 periods are 14.079.

We now show that total social costs can be further reduced by raising the first-period speed limit from 35 mph to 40 mph (or, equivalently, by reducing the first-period due level of care from 6 to 5). The accident costs in the first period increase, which leads social costs to increase from 7.053 to 7.067. However, the accident costs in the second period decrease, especially those related to high-risk types, which leads social costs to decrease from 7.026 to 6.997. Overall, total social costs decrease from 14.079 to 14.064.

The results from the practical example are summarized in Table 2. In Appendix B, we demonstrate that these findings hold general validity.

Lowering the due level of care in the first period amplifies the effect of tailoring negligence standards to accident records in the second period. Roughly speaking, we see that a slight reduction in the first-period due level of care yields an increase in the proportion of high-risk drivers in the group of injurers who had an accident in the first period and a decrease in the proportion of high-risk drivers in the group of injurers who did not have an accident in the first period. Essentially, inducing accidents in the first period better signals individuals’ risk types and amplifies the benefits of tailoring negligence standards in the next period.

For the sake of clarity, let us consider again the example of drivers A and B, in which driver A was driving at 40 mph and driver B was driv-

Table 2. Optimal Speed Limits

<table>
<thead>
<tr>
<th>Period 1</th>
<th>First-Time Offender</th>
<th>Repeat Offender</th>
<th>Social Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Policy 1</td>
<td>35</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>Policy 2</td>
<td>35</td>
<td>40</td>
<td>25</td>
</tr>
<tr>
<td>Policy 3</td>
<td>40</td>
<td>40</td>
<td>25</td>
</tr>
</tbody>
</table>

15. That is, (.0532 × .175 + .0015 × .825) × 9.44 + [(1 − .0532) × .175 + (1 − .0015) × .825] × 7.00 = 7.026.
ing at 30 mph when their accidents occurred. Suppose that the judge is more tolerant of first-time offenses and considers an injurer to be negligent if his or her speed did not exceed 40 mph \((x = 5)\) when the accident occurred. In this case, neither driver A nor driver B will be found negligent in the first period. This slight upward deviation of the speed limit approaches the optimal speed limit of driver A (that is, 40 mph), while it departs from the optimal speed limit of driver B (that is, 25 mph). Consequently, if an accident occurs in the next period, the judge can infer that the individual is more likely a high-risk type than a low-risk type and can efficiently require a higher due level of care, for example, a speed limit equal to 30 mph or less. Overall, compared with other policies, our proposal reduces the total social costs over time and approximates the first-best solution.

4. Conclusions

This paper adds important aspects of reality to the conventional unilateral care accident model, namely, that accidents may repetitively occur over time, potential injurers differ in their probabilities of accident, and neither the individual nor the social planner can perfectly observe an individual’s risk types. In this kind of context, we find that the accident history of an injurer should be used by courts to more accurately infer his or her risk type and adjust his or her due level of care accordingly over time. The less obvious and more novel insight is what we label the first-period effect: the efficiency of setting one-size-fits-all due levels of care at a level lower than prior literature has suggested.

Under current thinking, when courts have no information about heterogeneous risk, the optimal due level of care should be at some intermediate level. However, looking ahead might drive courts to deviate from this static optimization: by reducing the uniform, first-period due level of care, more accidents occur, and more information is subsequently gathered about individuals’ risk types. This added information helps to establish more efficient differentiated negligence standards.

Despite appearing counterintuitive at first glance, a slight reduction in the first-period due level of care is likely to substantially reduce social costs over time since it allows policy makers to better focus their safety efforts on the highest-risk proportion of the population. Our proposed policy helps judges to efficiently screen potential injurers over time and adjust the due level of care accordingly, thus better approaching the first-
best solution. Although we do not formally analyze a model with 3 or more periods, we believe that the optimal definition of due levels of care would have two properties: in each period, optimal due levels of care are nondecreasing in the number of prior violations, and the due level of care for individuals with the highest possible number of prior violations is maximal. We further believe that by considering an extended time frame, the first-period effect would be more salient and relevant.

Our theoretical contribution finds its practical relevance within the realm of judicial discretion in adhering to legislative standards of care, for example, speeding and standards of behavior for professional drivers. Cases in which statutory standards are so unreasonable that the court should optimally deviate from them might be relatively frequent, especially whenever the population of interest is heterogeneous. For example, driving 40 mph where the legal limit is 35 mph should not be considered inherently wrong or unreasonable. Moreover, within the unlimited realm of judicial discretion, our findings suggest the reason why and the circumstances under which judges should optimally deviate from strict adoption of legislative standards. It is important to note that accident records are informative and useful when the population of potential injurers is substantially heterogeneous in risk type. In this case, tailoring negligence standards to accident records is practically relevant and socially valuable.

The ability of judges and policy makers to efficiently tailor negligence standards to accident records might encounter practical difficulties given the possibly high administrative costs and frequent legal errors. The application of a vaguer uniform standard of care for heterogeneous injurers is generally defended when producing information on individuals’ subjective characteristics such as costs of private care is costly and error prone (see, for example, Ganuza and Gomez 2005; Emons 2007). Nonetheless, these problems with error are indeed a point of strength in tailoring negligence standards to accident records, which is expected to facilitate—rather than complicate—judicial discretion in assessing negligence. As intuition suggests, it is relatively costless to access information about individuals’ accident histories since most accident records are generally available to public authorities (Ben-Shahar and Porat 2016). Moreover, the occurrence of similar accidents involving the same individual is a reliable, uncontroversial, and objective piece of information. For these reasons, tailoring standards to individuals’ objective accident records is not expected to be excessively costly and error prone, at least if compared with individuals’ other subjective characteristics that are more difficult
for judges to correctly observe and assess, such as mental disability and costs of private care. In general, tailoring negligence standards to accident records is justified whenever the benefits of reducing social costs over time outweigh the costs of gathering and processing information about risk types.

Whether our theoretical proposal to tailor negligence standards to accident records would reduce the social costs of accidents in real-life cases is ultimately an empirical question. Accordingly, future research can be developed in this direction. For example, field experiments may be designed to test the efficiency of our proposal in reducing speeding violations over time. If a slight increase in the first-period speed limit and speed limits tailored to prior speeding-related accidents yield lower social costs over time than applying a uniform and static speed limit, then the first-period effect revealed in our analysis is not only theoretically but also practically relevant.

Another interesting direction for future studies would be to analyze situations in which courts do not observe individuals’ risk types while potential injurers do. In such cases, opportunistic behavior might arise, and the admissibility of past records at trial might not be warranted. Another possible extension is to model a high-risk type as implying a higher cost of care, as in Ganuza and Gomez (2005), instead of a higher probability of accident. In these cases, results can be different, especially in settings with asymmetric information. We would predict that, in a 2-period setting such as in this paper, where due care standards may depend on accident history, the classical ratchet effect would arise (Freixas, Guesnerie, and Tirole 1985; Laffont and Tirole 1990): low-cost injurers would mimic high-cost ones by being involved in accidents, which would allow them to be seen as high-cost injurers deserving a more lenient standard.

An important implication of our results is related to activity levels. By considering that activity levels are generally assumed to be inversely related to due levels of care (Shavell 1980, 1987), our findings suggest that it would be reasonable to discourage high-risk individuals from engaging in the relevant risky activity, at least for a certain period of time. This can explain the common practice of suspending licenses of drivers who are persistent violators. Similarly, truck drivers who have been repeatedly involved in accidents are generally moved to departments with low potentials for accidents, and young, less-experienced surgeons are not allowed to perform high-risk surgeries (at least without proper supervision).
It is worth noting that due levels of care tailored to accident records might produce not only positive but also potentially negative incentive effects. Given that injuries in the first period will raise due levels of care in the second period, individuals may refrain from engaging in risky yet beneficial activities, being afraid that their past will be used against them. Our framework can be fruitfully adapted to investigate such potential negative effects.

APPENDIX A: THE MODEL

A1. Setting the Stage

We consider a 2-period, unilateral care accident model with heterogeneous risk types and imperfect information. Each potential injurer in the society lives 2 periods, and the discount rate is 0 (no time preferences). We consider the case of risk-neutral agents and symmetric information; that is, the information available to each potential injurer and social planner is the same. An accident may occur in both or either of the periods of the model. In case of an accident, each victim suffers harm $d > 0$. We assume perfect compensation in cases in which the injurer is found liable. Potential injurers and victims are strangers to each other or at least do not have the possibility of bargaining.

A2. Notation and Assumptions

Let us now define the terms and articulate the assumptions of the model. Let $r \in \{L, H\}$ denote an individual’s risk type, either low risk (L) or high risk (H). Potential injurers and the social planner do not observe individuals’ risk types. However, they know the proportion of low-risk and high-risk types in the population. Let $\lambda_L$ and $\lambda_H$ be the proportions of low-risk and high-risk types, respectively, where $\lambda_L, \lambda_H > 0$, and $\lambda_L + \lambda_H = 1$ (for a similar framework, see, among others, Rothschild and Stiglitz 1992; Hoy 1982; Crocker and Doherty 2000; Bajtelsmit and Thistle 2008, 2009).

Let $A_1$ ($NA_1$) denote the event that an accident has occurred (has not occurred) in the first period. Let $t \in \{1, 2 | A_1, 2 | NA_1\}$ denote the following three states, respectively: first period, second period given that an accident occurred in the first period, and second period given that an accident did not occur in the first period.

16. The analysis is confined to a 2-period model to stylize the intertemporal allocation problem in the simplest way. Similar 2-period analyses have been used in economic models of repeat offenders (see, for example, Polinsky and Rubinfeld 1991; Chu, Hu, and Huang 2000; Endres and Bertram 2006).

17. We discuss the implications of this assumption in Section 4.
In each period, each potential injurer can invest in precautions $x_t \in \mathbb{R}^+_0$ at a unit cost of 1 to reduce the probability of an accident.

Let $p_L(x_t)$ ($p_H(x_t)$) be the probability of an accident when a low-risk type (high-risk type) spends $x_t$ on care. The probability of an accident is a strictly decreasing, strictly convex function of expenditure on care. Low-risk types have a lower accident probability for any expenditure on care. Formally, $0 < p_L(x_t) < p_H(x_t) < 1$, $\forall x_t.18$ Following Bajtelsmit and Thistle (2008, 2009), we assume that $0 > p_L'(x_t) > p_H'(x_t)$, $\forall x_t$; that is, an additional investment in care by a low-risk type is less productive in reducing the probability of accident than an equivalent investment in care by a high-risk type (for an extensive discussion of this assumption, see Bajtelsmit and Thistle 2008).

A3. Period 1

The probabilities that a potential injurer is involved in an accident or not are respectively given as follows:19

$$P(A_t) = \lambda_L p_L(x_t) + \lambda_H p_H(x_t)$$

(A1)

and

$$P(NA_t) = \lambda_L [1 - p_L(x_t)] + \lambda_H [1 - p_H(x_t)].$$

(A2)

A4. Period 2

In the second period, each potential injurer may have already been involved in an accident in the first period. Each potential injurer’s and the social planner’s prior beliefs about an individual’s risk type are updated on the basis of whether an accident happened in the first period. Let $\lambda_{IA_t}$ ($\lambda_{INA_t}$) denote the probabilities of belonging to a certain risk type $r \in \{L, H\}$ given that an accident happened (did not happen) in the first period. Using Bayes’s rule, $\lambda_{IA_t}$ and $\lambda_{INA_t}$ are respectively given as follows:

$$\lambda_{IA_t} = \lambda_r p_r(x_t)/P(A_t)$$

(A3)

and

$$\lambda_{INA_t} = \lambda_r [1 - p_r(x_t)]/P(NA_t)$$

(A4)

for each $r \in \{L, H\}$.


19. In particular, $P(A_t)$ in equation (A1) refers to the probability with which a randomly selected individual may be one who has had an accident, given that everyone in society chooses care level $x_t$. Therefore, it also can be interpreted as the measure of individuals who have accidents in period 1. This interpretation also holds for the other similar specification in the model.
In the second period, each potential injurer can belong to one of two categories: those who incurred an accident in the first period and those who did not. In each category, each potential injurer can be either a high-risk or a low-risk type.

Let $x_{2A_1}$ ($x_{2NA_1}$) denote the potential injurer’s expenditure on care in the second period given that an accident occurred (did not occur) in the first period. The probabilities that a potential injurer is involved in an accident in the second period given that an accident either occurred or did not occur in the first period are respectively defined as follows:

$$P(A_2 \mid A_1) = \lambda_{A_1} p_L(x_{2A_1}) + \lambda_{HA_1} p_H(x_{2A_1}) \quad (A5)$$

and

$$P(A_2 \mid NA_1) = \lambda_{NA_1} p_L(x_{2NA_1}) + \lambda_{BNA_1} p_H(x_{2NA_1}). \quad (A6)$$

In the following sections we use this setup to characterize the social and private optimization problems.

**A5. Social Optimization Problem**

The social optimization objective is to minimize the overall expected social cost of accidents $S$, which is the sum of the expected social cost of accidents in both periods. The overall expected social cost of accidents includes the cost of taking care and the expected damages from accidents. The social maximization problem can be defined as follows:  

$$\min_{x_1, x_{2A_1}, x_{2NA_1}} S(x_1, x_{2A_1}, x_{2NA_1}) = S_1 + P(A_1) S_{2A_1} + P(NA_1) S_{2NA_1}, \quad (A7)$$

where $S_1$ denotes the social costs of accidents in the first period, and $S_{2A_1}$ ($S_{2NA_1}$) denotes the social costs of accidents in the second period given that an accident occurred (did not occur) in the first period. These period-specific social costs of accidents are respectively defined as follows:

$$S_1 = P(A_1) d + x_1, \quad (A8)$$

$$S_{2A_1} = P(A_2 \mid A_1) d + x_{2A_1}, \quad (A9)$$

and

$$S_{2NA_1} = P(A_2 \mid NA_1) d + x_{2NA_1}. \quad (A10)$$

20. We limit our attention to primary effects. Given risk neutrality, secondary effects (namely, issues related to risk spreading) are absent in this paper. Tertiary effects (namely, issues related to transaction costs) are ignored.

21. For simplicity and without loss of generality, expressions in the first period are not discounted. Discounting is irrelevant for our results as long as individual discount rates do not deviate from the social discount rate. For similar formulations, see Endres and Bertram (2006) and Friehe and Langlais (2017).
The first-order conditions for the solution of the social minimization problem (A7) define the due levels of care in each period, which are denoted $x_{1i}^{**}$, $x_{2iA_i}^{**}$, and $x_{2iNA_i}^{**}$. Analytically, $x_{1i}^{**}$, $x_{2iA_i}^{**}$, and $x_{2iNA_i}^{**}$ are respectively defined as follows:

\[-\lambda_L p_L(x_i) [x_{2iA_i} - x_{2iNA_i}] + d [1 + p_L(x_{2iA_i}) - p_L(x_{2iNA_i})] \]
\[-\lambda_H p_H(x_i) [x_{2iA_i} - x_{2iNA_i}] + d [1 + p_H(x_{2iA_i}) - p_H(x_{2iNA_i})] = 1, \tag{A11}\]

\[\lambda_L p_L(x_i) [1 + dp_L'(x_{2iA_i})] + \lambda_H p_H(x_i) [1 + dp_H'(x_{2iA_i})] = 0, \tag{A12}\]

and

\[\lambda_L [1 - p_L(x_i)] [1 + dp_L'(x_{2iNA_i})] + \lambda_H [1 - p_H(x_i)] [1 + dp_H'(x_{2iNA_i})] = 0. \tag{A13}\]

### A6. Private Optimization Problem

We consider the private optimization problem under a simple negligence regime. The court is assumed to be able to perfectly observe the level of care adopted by the injurer. The private optimization problem is defined as follows:

\[
\min_{x_{1i}^{**}, x_{2iA_i}^{**}, x_{2iNA_i}^{**}} T(x_{1i}, x_{2iA_i}, x_{2iNA_i}) = T_1 + P(A_i)T_{2iA_i} + p(NA_i)T_{2iNA_i}, \tag{A14}
\]

where $T_1$ denotes the private costs of accidents in the first period, and $T_{2iA_i}$ ($T_{2iNA_i}$) denotes the private costs of accidents in the second period given that an accident occurred (did not occur) in the first period. These period-specific private costs of accidents are defined as follows:

\[
T_1 = \Phi_t P(A_i) d + x_{1i}, \tag{A15}
\]

\[
T_{2iA_i} = \Phi_{2iA_i} P(A_{2i} | A_i) d + x_{2iA_i}, \tag{A16}
\]

and

\[
T_{2iNA_i} = \Phi_{2iNA_i} P(A_{2i} | NA_i) d + x_{2iNA_i}. \tag{A17}
\]

The first-order conditions for the solution of the private minimization problem in equation (A14) define the privately optimal care levels in each period, which are denoted $x_{1i}^{**}$, $x_{2iA_i}^{**}$, and $x_{2iNA_i}^{**}$.

Under a simple negligence rule, parameters $\Phi_t$ are defined as follows:

\[
\Phi_t = \begin{cases} 
0 & \text{given that } x_t \geq x_t^{**} \\
1 & \text{given that } x_t < x_t^{**}.
\end{cases} \tag{A18}
\]

For a similar formulation of the private minimization problem given a rule of negligence, see Endres and Bertram (2006). For a study of the effect of legal error on the incentives to exercise care, see Polinsky and Shavell (1989).
As in the traditional literature on the economics of negligence, the switch parameters $\Phi_t$ guarantee that the compensation that the potential injurer has to pay is equal to damages if he or she is found negligent. Comparing the expected private costs of the potential injurer when his or her level of care equals the due levels of care with a situation in which the injurer is negligent leads to the well-known result that individual compliance with the due level of care is the private optimum.  

**APPENDIX B: PROOFS**

**B1. Proof of Proposition 1**

By rearranging equations (A12) and (A13), $x_{2A_1}^*$ and $x_{2NA_1}^*$ can be expressed as follows:

$$-\frac{\lambda_{1H}}{\lambda_{H1}} = \frac{p_{H}(x_{1}^* | x_{2}^*)[1 + dp_{H}^l(x_{2A_1})]}{p_{L}(x_{1}^* | x_{2}^*)[1 + dp_{L}^l(x_{2A_1})]}$$  (B1)

and

$$\frac{\lambda_{L}}{\lambda_{H1}} = \frac{[1 - p_{H}(x_{1}^* | x_{2}^*)][1 + dp_{H}^l(x_{2NA_1})]}{[1 - p_{L}(x_{1}^* | x_{2}^*)][1 + dp_{L}^l(x_{2NA_1})]}$$  (B2)

respectively. Since the left-hand side in equation (B1) is equal to the left-hand side in equation (B2), for the first-order conditions (B1) and (B2) to hold, the right-hand side of equation (B1) should be equal to the right-hand side of equation (B2). This requires that (rearranging)

$$\frac{p_{L}(x_{1}^* | x_{2}^*)[1 - p_{H}(x_{1}^* | x_{2}^*)]}{p_{H}(x_{1}^* | x_{2}^*)[1 - p_{L}(x_{1}^* | x_{2}^*)]} = \frac{[1 + dp_{H}^l(x_{2A_1})][1 + dp_{L}^l(x_{2NA_1})]}{[1 + dp_{H}(x_{2A_1})][1 + dp_{L}(x_{2NA_1})]}.$$  (B3)

Since by construction $0 < p_{H}(x_{1} | x_{2}) < p_{L}(x_{1} | x_{2}) < 1, \forall x_{1},$ where $t \in \{1, 2 | A_1, 2 | NA_1\},$ the left-hand side of equation (B3) is less than 1. Condition (B3) holds if its right-hand side is less than 1 as well. This requires that (rearranging)

$$\frac{1 + dp_{H}^l(x_{2A_1})}{1 + dp_{L}^l(x_{2NA_1})} < \frac{1 + dp_{L}^l(x_{2A_1})}{1 + dp_{H}^l(x_{2NA_1})}.$$  (B4)

If $x_{2A_1}^{*} = x_{2NA_1}^{*},$ the left-hand side of expression (B4) becomes equal to the right-hand side of expression (B4); this yields a result that violates expression (B4). Since by construction $p'(x_{1}) < 0,$ $p''(x_{1}) < 0, \forall x_{1},$ where $t \in \{1, 2 | A_1, 2 | NA_1\},$ if $x_{2A_1}^{*} < x_{2NA_1}^{*},$ then $p_{H}^l(x_{2A_1}) > p_{H}^l(x_{2NA_1}), \forall r \in \{L, H\}.$ By construction, $0 > p_{L}^l(x_{1}) > p_{H}^l(x_{1}), \forall x_{1}.$ This implies that if $x_{2A_1}^{*} < x_{2NA_1}^{*},$ the left-hand side of expression (B4) increases faster than the right-hand side of expression (B4). This yields a result that violates expression (B4). By contrast, if

23. Proofs are available from the authors on request.
$x_{21A} > x_{21NA}$, then $p'(x_{21A}) < p'(x_{21NA})$, $\forall r \in \{L, H\}$. Since by construction $0 > p'_l(x_i) > p'_h(x_i)$, $\forall x_i$, this implies that if $x_{21A} > x_{21NA}$, the left-hand side of expression (B4) decreases faster than the right-hand side of expression (B4). This yields a result that satisfies expression (B4). It follows that expression (B3) holds if $x_{21A} > x_{21NA}$.

Q.E.D.

B2. Proof of Proposition 2

Let us first analytically derive the third-best due level of care. When information about an individual’s risk type is hidden and information about accident records is not available (or not accessible), the social optimization problem over 2 periods simplifies as follows:

$$\min_{\bar{x}} S_0(\bar{x}) = 2[P(A)d + \bar{x}], \quad (B5)$$

where $P(A) = \lambda_l p_l(\bar{x}) + \lambda_h p_h(\bar{x})$. The solution gives the third-best, uniform, averaged due level of care, which is denoted $\bar{x}^{**}$ and defined as follows:

$$-(\lambda_l p'_l(\bar{x}) + \lambda_h p'_h(\bar{x}))d = 1. \quad (B6)$$

Let us compare the second-best due level of care $x^*_1$—as defined in equation (A11)—and the third-best due level of care—as defined in equation (B6). It can be easily proved that $x^*_1 = \bar{x}^{**}$ if $x_{21A} = \bar{x}_{21A}$, which is never satisfied as proved in proposition 1. It follows that $x^*_1 \neq \bar{x}^{**}$.

We now prove that $x^*_1 < \bar{x}^{**}$. The left-hand side in equation (A11) (equation [B6]) represents the marginal benefit of $x^*_1 (\bar{x}^{**})$. Thus, $x^*_1 < \bar{x}^{**}$ if the marginal benefit of $x^*_1$ is lower than the marginal benefit of $\bar{x}^{**}$. This happens when $x_{21A} + p_l(x_{21A})d > \bar{x}_{21A} + p_l(\bar{x}_{21A})d$, $\forall r \in \{L, H\}$, which is always satisfied given that $x_{21A} > \bar{x}_{21A}$, as shown in the proof of proposition 1. Q.E.D.

B3. Proof of the Desirability of Tailoring Negligence Standards

We first recall that tailoring due levels of care to accident records represents the second-best solution to the social minimization problem described in equation (A7), whereas averaging due levels of care represents the third-best solution. Thus, the social costs of accidents under tailored due levels of care are lower than under averaged due levels of care. Formally, $S(x^*_1, x^*_{21A}, x^*_{21NA}) < S_0(\bar{x}^{**})$.

Let us characterize the social welfare function when punitive damages are awarded to repeat offenders. Let $D > 0$ denote punitive damages. Let the due levels of care be fixed to the averaged, third-best level $\bar{x}^{**}$ as defined in equation (B6). The social welfare function $S_D$ is defined as follows:

$$S_D(\bar{x}^{**}) = 2[P(A)d + \bar{x}^{**}] + P(A)^2D, \quad (B7)$$

where $P(A) = \lambda_l p_l(\bar{x}^{**}) + \lambda_h p_h(\bar{x}^{**})$ denotes the probability of an accident in each period. Since $D > 0$ by construction, it follows straightforwardly that
Thus, \( S(x_1^*, x_2^*, x_{2|A_i}^*) < S_D(\vec{x}^*), \) which proves that tailoring negligence standards to accident records outperforms awarding punitive damages to repeat offenders in minimizing the social costs of accidents. Q.E.D.

REFERENCES


Dionne, Georges, Pierre-Carl Michaud, and Jean Pinquet. 2013. A Review of Recent Theoretical and Empirical Analyses of Asymmetric Information in Road

24. This is in line with the standard result in the torts literature. Under the negligence rule, the use of punitive damages is generally either socially detrimental (since it may lead to excessive levels of care with respect to the standard of due care) or socially irrelevant (since potential injurers will optimally invest in the standard of due care, thus avoiding both compensatory and punitive damages). See, among others, Polinsky and Rubinfeld (1991), Polinsky and Shavell (1998b), and Shavell (2007, p. 166). There are two fairly general circumstances in which damages exceeding harm are socially desirable: when injurers are hard to identify as the sources of harm (for example, if victims do not bring suit) and when injurers gain illicit utility—that is, utility not included in the social welfare function (Polinsky and Rubinfeld 1991)—from harmful acts (for example, destroying someone’s property). These two exceptions lie beyond the scope of this paper.


