But-for Causation and the Implementability of Compensatory Damages Rules

Urs Schweizer*, University of Bonn

July 27, 2018

Abstract

This paper considers two parties whose choice of actions imposes external effects on each other. All uncertainty is captured by a random state of nature. The interaction among the two parties is governed by damages rules. Evidence is generated by a signal, that is, a state-contingent function of actions. Parties are assumed to choose Nash equilibria of the game induced by the damages rule in place. Incentives are derived from compensation requirements. The performance of such a rule is evaluated in terms of welfare levels of these Nash equilibria. The paper provides sufficient criteria for signals to allow implementing compensatory and, hence, welfare enhancing damages rules.

JEL classification: K13

Key words: legal damages, tort law, causation, limited observability, inefficient standards, bidirectional externality, compensation principle, Nash implementation

*address: Urs Schweizer, Department of Economics, University of Bonn, Kaiserstr. 1, 53113 Bonn, Germany, schweizer@uni-bonn.de
1 Introduction

This paper considers two parties whose choice of actions imposes external effects on each other. You may think of a bidirectional accident model where both parties are in the role of injurer and victim. Actions may consist of choosing a one-dimensional level of care. Yet, the focus will be on more general action spaces, not endowed with a linear ordering. You may think of two-dimensional actions combining the choice of precaution expenditures and activity levels, of multi-dimensional precaution choice, of mixed strategies over a set of pure strategies or of complete contingent plans in a setting with asymmetric information.

The interaction among the two parties is governed by damages rules. Parties are assumed to choose a Nash equilibrium of the game induced by the damages rule in place. The performance of such a rule is evaluated in terms of welfare levels achieved in Nash equilibrium. A profile of actions and the welfare level under this profile serve as reference points. This reference profile may but need not maximize welfare.

A damages rule is called welfare enhancing if all Nash equilibria (multiple equilibria are not ruled out) come with a welfare level not below the one from the reference profile. In particular, if the reference profile is efficient (welfare maximizing) and if the damages rule is welfare enhancing then all Nash equilibria will be efficient.

Comparing welfare of Nash equilibrium under inefficient due care standards with welfare from keeping these standards has been pioneered by Dari-Mattiacci and Hendriks (2013). The present paper examines the notion of welfare enhancing damages rules within the general externality setting. Extending Schweizer (2016) beyond efficient reference profiles, incentives are derived from compensation requirements. A damages rule is called compensatory relative to the reference profile if no party is worse off under unilateral deviations from the reference profile by the other party. It will be shown that compensatory damages rules are welfare enhancing, a result referred to as compensation principle.

Compensation requirements are analytically easy to handle and lead to welfare results of great generality. Action spaces without linear ordering remain tractable without extra effort. Moreover, compensation requirements
are separable in each direction of the externality, which provides an easy link from unidirectional to bidirectional models.

The compensation principle as used by the present paper also extends findings by Jain and Singh (2002) who have considered an unidirectional accident model. Among all conceivable damages rules, they classify some of them as liability rules. Within that class, they characterize those which generate efficient incentives. Their criterion requires that a victim who has not spent less than due care can fully recover whereas an injurer who has not spent less than due care is never held liable. Within this class of liability rules, their criterion is equivalent to the compensation requirement of the present paper so that their propositions 1 and 2 directly follow from the more general compensation principle.

In any case, from an incentive perspective, it seems desirable to implement compensatory damages rules. Yet, as evidence available to courts may only be correlated with the action profile without fully revealing it, this need not always be possible.

Courts may err in assessing a party’s action (see Craswell and Calfee (1986) for a study where legal standards are uncertain). Or a party may suffer from lapses of attention. The party intends to take an action (input) but effective becomes another action (output) which differs from the input by an error term. Given lapses of attention, courts are assumed to observe the output but not the input (see Cooter and Porat (2014) for a study on lapses of attention).

The existing literature typically models such uncertainty by postulating distribution functions that depend on actions. For the general externality model, however, all uncertainty is assumed being driven by a random state of nature. In particular, evidence is generated by a signal, that is a state-contingent function of the chosen actions. The main contribution of the present paper consists of providing criteria on signals which, if satisfied, allow implementing compensatory and, hence, welfare enhancing damages rules.

These criteria turn out being related to but-for causation. Introducing a state space has the advantage of generating counterfactuals needed for the but-for test endogenously. The following implementability result (among others) will be established. Suppose the evidence reveals of each party whether or not she has caused any harm to the other party according to the but-
for test. Then the signal allows implementing a compensatory damages rule even if the size of harm (if any) and the action profile itself remain hidden to courts.

The implementability criteria of the present paper are then confronted with various tort models from the existing literature. As a first application, accident models are identified as special cases of the general externality model. This includes unidirectional accident models as first studied by Brown (1973) as well as bidirectional accident models where both parties may suffer from accidents (for bidirectional models, see Leong (1989) and Arlen (1990) among others).

This literature has investigated incentives under negligence rules. For a party to be held liable, her behavior as an injurer must be deemed negligent. If it is then the other party as victim can recover the full harm from accidents.

Negligence rules do not explicitly make use of a but-for test. In fact, just knowing accident probabilities as functions of the action profile would not even be sufficient to deal with but-for causation properly.

Moreover, to implement a negligence regime literally, the available evidence should reveal whether or not a party was negligent, an accident has occurred and, in case of an accident, what size total harm is of. If the signal reveals less, implementing sort of a negligence rule may still be feasible. To this end, a negligence criterion and a proxy for total harm will be specified leading to a compensatory damages rule whenever the implementability criteria are met. Negligence rules, in fact, compensate even excessively. Yet, overcompensation does not distort incentives as long as the appropriate negligence criterion is in use.

Bidirectional externalities also occur in models where each party’s precaution directly affects the other party’s cost of taking precaution. Dharmapala and Hoffmann (2005) have shown that none of the standard (unilateral) tort liability rules induce socially optimal behavior by both parties. Damages rules compensatory relative to an efficient reference profile, in contrast, provide efficient incentives and can be implemented whenever the implementability criteria of the present paper are satisfied.

As a second application, the same criteria will be confronted with activity-precaution models as pioneered by Shavell (1980, 1987). For legal reasons (or because activity levels remain hidden to courts), the literature mainly deals
with damages rules that condition exclusively on precaution expenditures. Quite generally, such rules fail being compensatory and they are well-known to distort incentives.

In contrast, the implementability criteria of the present paper provide insights as to what additional evidence would be needed to fully internalize such external effects. In particular, if courts observe whether or not the activity has been operated at an excessive level (relative to a given reference level) then they could implement a damages rule, which is welfare enhancing relative to this level.

As a third application, I deal with the acquisition of information before parties choose actions. Kaplow and Shavell (1996) have studied an unilateral accident model where the injurer has the option to learn, at costs, the true level of harm in case of an accident before deciding on precaution expenditures. The present setting captures this by a particular signal and, to apply the compensation principle, by a set of complete contingent plans as the injurer’s action space.

Outside the lab, complete contingent plans are hardly ever observable. Nonetheless, a compensatory damages rule can be implemented as long as the victim’s actual level of harm remains observable by courts. A compensatory damages rule generates incentives for the injurer to acquire information even if the acquisition activity itself remains hidden to courts.

In addition, Kaplow and Shavell also investigate acquisition of information ex post under the assumption that both parties learn the true level of harm and have the option of revealing it to courts. So far, however, I have not been able to make use of the compensation principle if revelation ex post by the involved parties is at stake.

As a fourth application, the model by Kaplow (2012) on the burden of proof is rephrased in the setting of the present paper. Kaplow examines the trade-offs from raising the strength of burden of proof on welfare. In an unilateral accident model, he postulates a signal attaining values from the real line. Courts specify the level of the signal beyond which the injurer is held liable. Within this class of liability, Kaplow propagates the level of strength which maximizes welfare. Quite generally, this liability criterion outperforms preponderance of the evidence but may remain second best nonetheless.

The present paper aims, instead, at implementing welfare enhancing dam-
ages rules. Not every signal allows doing this. But, whenever the signal reveals the victim’s level of harm (if the injurer engages in a harmful activity) or level of benefit (if the injurer engages in a benign activity), then a compensatory and, hence, welfare enhancing rule can be implemented. If the reference profile is efficient then such a rule would even be first best.

Finally, an extension of the externality model is considered where the legal system has the option to affect the signal generating evidence. To this end, the multi-dimensional precaution model by Dari-Mattiacci (2005) on the scope of negligence is embedded into the externality setting. The scope of negligence concerns the dimensions that enter the negligence criterion. Different scopes differ with respect to costs to society and are subject of social choice.

Dari-Mattiacci considers only damages rules which induce zero precaution in dimensions outside the scope of negligence. It is this assumption which allows making use of the compensation principle. For any given scope, a damages rule is implemented, which would be compensatory for precaution choice restricted to the dimensions within the scope. The scope of negligence is predicted to emerge which leads to the highest welfare net of social costs.

The paper is organized as follows. Section 2 provides an introductory example to show why a state space is useful for addressing but-for causation. Section 3 introduces the general externality model and establishes the compensation principle. Section 4 deals with but-for causation and its relation to the implementability criterion. Section 5 confronts the findings with various tort models from the existing literature. Section 6 concludes.

2 Introductory example

In this section, an unilateral accident model is considered where party $A$ faces a binary precaution choice $x_A$ from \{0, 1\} at cost $c \cdot x_A$. For illustration, think of $x_A = 0$ as meaning that party $A$ has illegally disposed his toxic waste in a lake. All fish are dead. Party $B$ owns the fish, their market value being $h$.

If it were known that fish of market value $h^o$ (by assumption, $0 < h^o < h$) would have been dead even if $A$ had not disposed illegally, that is, if $A$ had decided $x_A = 1$, then party $A$’s illegal act has caused harm of size $\Delta = h - h^o > 0$ so that $B$ would be entitled to recover the difference $\Delta$ from
A (according to the but-for test and as proposed by Grady (1988) and Kahan (1989)). Yet, in general, the counterfactual value $h^o$ need not be known for sure and the impact $\Delta$ may remain uncertain.

The traditional accident model captures uncertainty by specifying the distribution of harm as a function of precaution choice $x_A$. Assume that victim $B$ suffers from total harm $h_n$ with probability $p_n(x_A)$ where $h_0 = 0 < h_1 < \ldots < h_n < \ldots < h_N$. A higher index means a higher level of total harm. At due care $x_A = 1$, low levels of harm are more likely (in the sense of stochastic dominance) than at precaution $x_A = 0$ below due care, that is

$$\sum_{m=0}^{n} p_m(0) < \sum_{m=0}^{n} p_m(1)$$

holds for all $n < N$ and $\sum_{m=0}^{N} p_m(0) = \sum_{m=0}^{N} p_m(1) = 1$ for $n = N$.

Suppose $A$ has actually spent (ex ante) below due care, that is $x_A = 0$, and $B$ suffers (ex post) from harm $h = h_n$. The but-for test compares the actual level $h_n$ of harm with the counterfactual level $h^o = h_m$, which would have resulted if $A$ had met her obligation $x^o_A = 1$. The probabilistic setting underlying traditional accident models does not generate counterfactual values explicitly.

As a remedy, a state space $\Omega$ is postulated instead as a primitive of the model. This state space is endowed with an exogenous probability measure $\pi$ driving all uncertainty. You may think of states $\omega = (h, h^o)$ simply as consisting of the actual and counterfactual level of harm $h$ and $h^o$ where state $\omega_{nm} = (h_n, h_m)$ occurs with probability $\pi_{nm}$. These probabilities are primitives of the model whereas the distribution functions can now be calculated endogenously from them as

$$p_n(0) = \sum_{m=0}^{N} \pi_{nm} \text{ and } p_m(1) = \sum_{n=0}^{N} \pi_{nm}.$$ 

Conversely, however, the probabilities $\pi_{nm}$ of states do not follow unambiguously from the probabilities $p_n(0)$ and $p_m(1)$ unless the distributions of harm under $x_A = 0$ and $x_A = 1$ are independent, in which case $\pi_{nm} = p_n(0) \cdot p_m(1)$ would hold.

Independence, however, seems a severe restriction as it rules out all cases where observing the actual harm level allows updating beliefs with respect to the counterfactual level of harm. Think, for instance, of cases where the
level of harm under due precaution can never be higher than under precaution below due care (no benefits from negligence). In such cases, \( \pi_{nm} = 0 \) would hold for all \( m > n \) which contradicts independent distributions. In the presence of correlation, knowing the probabilities \( p_n(0) \) and \( p_m(1) \) is not sufficient to calculate the probabilities \( \pi_{nm} \) of states unambiguously.

In any case, if a state space serves as a primitive of the model, then utility (profit) of party \( i = A, B \) will be expressed as function \( U_i(x_A, \omega) \) of \( A \)'s precaution expenditures \( x_A \) and the state \( \omega \) randomly chosen by nature. In the above example, these functions would be \( U_A(x_A, \omega_{nm}) = -c \cdot x_A \) and

\[
U_B(0, \omega_{nm}) = -h_n \quad \text{and} \quad U_B(1, \omega_{nm}) = -h_m,
\]

whereas the impact from \( A \)'s deviation, according to the but-for test, would be \( \Delta(x_A, \omega) = U_B(1, \omega) - U_B(0, \omega) \). If positive, then \( \Delta(x_A, \omega) \) corresponds to harm caused by \( A \)'s negligence and \( B \) may be entitled to recover \( \Delta(x_A, \omega) \) from \( A \). If negative, then the impact \( \Delta(x_A, \omega) \) would consist of a benefit of size \( -\Delta(x_A, \omega) \) caused by \( A \)'s negligence. Under tort law, such benefits are typically kept for free, in which case damages in line with the but-for test may be summarized as \( \max\{\Delta(x_A, \omega), 0\} \).

Courts have to specify damages based on the available evidence \( s \). Typically, such evidence will be correlated with precaution \( x_A \) and the state \( \omega \) of nature without fully revealing neither of them. In the model, evidence \( s \) is generated by a signal \( s = \sigma(x_A, \omega) \), that is a state-contingent function of precaution choice \( x_A \). The signal is understood as a primitive of the model. To be implementable, damages \( D_{AB}(s) \) owed by \( A \) to \( B \) are based on the available evidence, that is they are functions exclusively of \( s \).

Tort law aims at compensating victims for negligent behavior of the injurer. From an incentive perspective, imposing compensation requirements proves desirable as well. In fact, requiring injurers to compensate victims for deviations from due care generates efficient incentives (provided that due care is specified at the efficient level). Since precaution is chosen ex ante, compensation in expected terms would be sufficient.

The but-for test, in contrast, is applied ex post. Yet, not even ex post, evidence need fully reveal the counterfactual value underlying the test. To stay as close to the but-for test as the available information permits, it is proposed to award expected damages based on the but-for test, for lack of
more accurate information though, conditional on the available evidence.

In the above example, this means the following. Suppose it is just known that \( A \) has spent \( x_A = 0 \) below due care and that the victim suffers from harm \( h_n \). Then damages based on the but-for test and conditional on the available evidence amount to

\[
D_{AB}(x_A = 0, h = h_n) = \frac{1}{p_n(0)} \cdot \sum_{m=0}^{n} \pi_{nm} \cdot (h_n - h_m).
\]

With these damages, the victim would (in expected terms) be compensated or, in fact, even overcompensated by the amount

\[
\sum_{n=0}^{N-1} \sum_{m=n+1}^{N} \pi_{nm} \cdot (h_m - h_n) > 0
\]

(unless \( \pi_{nm} = 0 \) holds for all \( n < m \)). Such overcompensation would be due to the legal practice that, under tort law, benefits from deviations must not be returned.

The above example is simple enough and the signal sufficiently informative for implementing the above damages regime which is compensatory in expected terms. The externality model to be introduced in the next section generalizes the above example in many directions. First, both parties take decisions (bilateral case). Second, the externalities may be bidirectional and of general type, not necessarily involving accidents. Third, spaces from which to choose actions may be multi-dimensional and may concern precaution expenditures as well as activity levels (or, where appropriate, complete contingent plans).

For this general externality model, signals will then be characterized, which allow implementing compensatory damages regimes.

3 The general externality model

In this section, a general model is introduced with two parties \( A \) and \( B \), whose actions may generate external effects in both directions. Parties \( A \) and \( B \) choose actions \( x_A \) and \( x_B \) from given sets \( X_A \) and \( X_B \) of alternatives. Both parties’ decisions may be one- or multi-dimensional and may consist of discrete or continuous choice or even a mixture of both (for applications, see
section 5 below). No linear ordering is imposed on the action spaces $X_A$ and $X_B$, that is, they need not be subsets of the real line.

After parties have chosen the action profile $x = (x_A, x_B)$ from $X = X_A \times X_B$, nature randomly selects the state $\omega$ from state space $\Omega$ according to probability measure $\pi$. Party $i$’s profit (utility) is a state-contingent function $U_i(x, \omega)$ of the action profile. Expected welfare amounts to $w(x) = E[U_A(x, \omega) + U_B(x, \omega)]$ where $E$ denotes expectation with respect to the exogenously given probability measure $\pi$ of state space $\Omega$.

Evidence, being generated by a signal $\sigma : X \times \Omega \rightarrow S$, is also state-contingent so that it is correlated with the state and the action profile. To remain implementable, damages are based on the available evidence $s$ from the set $S$ of conceivable alternatives.

Whenever externalities are of bidirectional nature, a damages rule $D(s) = [D_{AB}(s), D_{BA}(s)]$ specifies damages in both directions. Here, $D_{AB}(s)$ denotes damages owed by $A$ to $B$ and $D_{BA}(s)$ damages in the opposite direction. In sum, $A$ owes net damages $D_{AB}(s) - D_{BA}(s)$ to $B$ and $B$ owes $D_{BA}(s) - D_{AB}(s)$ to $A$.

Damages rule $D(s)$ leads to a game in normal form with $X_A$ and $X_B$ as strategy spaces and with payoff functions
\[
\phi_i(x) = E[U_i(x, \omega) + D_{ji}(\sigma(x, \omega)) - D_{ij}(\sigma(x, \omega))]
\]
for $i = A, B$. Rational parties choose an action profile $x^N \in X$ consisting of mutually best responses, that is a Nash equilibrium of the game induced by the damages rule in place.

In the following, a profile $x^o \in X$ serves as reference profile leading to $w(x^o)$ as a reference level for welfare comparisons. A damages rule $D(s)$ is called welfare enhancing if $w(x^N) \geq w(x^o)$ holds for any Nash equilibrium $x^N$ of the game induced by $D(s)$. The damages rule is called compensatory relative to the reference profile if no party is worse off under unilateral deviations by the other party, that is if the compensation requirements
\[
\phi_i(x^o_i, x_j) \geq \phi_i(x^o)
\] (1)
hold for all $x_j \in X_j$ and $i = A, B$. The following result is referred to as compensation principle.
Proposition 1 If, for a given signal \( \sigma : X \times \Omega \rightarrow S \), the damages rule \( D(s) \) is compensatory relative to the reference profile, then the regime is welfare enhancing relative to the same profile. Moreover, if the reference profile maximizes welfare then it is a Nash equilibrium and all Nash equilibria (if more than one exists) are payoff equivalent.

Proof. This is a generalized version of the compensation principle as established in Schweizer (2016). The proof is straightforward. Suppose \( x^N \) is a Nash equilibrium. Since \( x^N_A \) and \( x^N_B \) are mutually best responses, in particular, the inequalities

\[
\phi_A(x^N) \geq \phi_A(x^o_A, x^N_B) \quad \text{and} \quad \phi_B(x^N) \geq \phi_B(x^N_A, x^o_B)
\]

must hold. It then follows from (1) that

\[
\phi_A(x^o_A, x^N_B) \geq \phi_A(x^o) \quad \text{and} \quad \phi_B(x^N_A, x^o_B) \geq \phi_B(x^o)
\]

and, hence,

\[
\phi_A(x^N) \geq \phi_A(x^o) \quad \text{and} \quad \phi_B(x^N) \geq \phi_B(x^o)
\]

(2) must hold. From (2), it follows that \( w(x^N) \geq w(x^o) \) and, hence, the first claim is established.

To establish the remaining claims, suppose that the reference profile maximizes welfare so that the inequality \( w(x^N, y^N) \geq w(x^o, y^o) \) must be binding and, hence, both inequalities in (2) must be binding as well. This establishes payoff equivalence. For a proof that the reference profile is a Nash equilibrium (provided that the reference profile is efficient), the reader is referred to Schweizer (2016).

Before introducing implementability criteria in the next section, the following signals may serve as illustration. Consider, first, the signal \( \sigma(x, \omega) = x \) which reveals the true action profile but nothing else. Then the damages rule

\[
D_{ij}(x) = \max \{ E[U_j(x^o_i, x_j, \omega) - U_j(x_i, x_j, \omega)], 0 \}
\]

is compensatory as is easily seen (see Kaplow and Shavell (1996) for a closely related rule).

Consider, second the signal \( \sigma(x, \omega) = [x_B, U_B(x, \omega)] \), which reveals the action \( x_B \) and the realized profit \( u_B = U_B(x, \omega) \) of one party, say \( B \), but
nothing else. This signal allows implementing the following damages rule which turns out being compensatory:

\[ D_{AB}(x_B, u_B) = E[U_B(x_A^o, x_B, \omega)] - u_B \]

and

\[ D_{BA}(x_B, u_B) = D_{BA}(x_B) = \max \{ E[U_A(x_A^o, x_B^o, \omega) - U_A(x_A^o, x_B, \omega)] , 0 \} \]

Under this rule, again no party is held liable who has kept to the reference profile, that is

\[ E[D_{ij}(x_i^o, U_B(x_i^o, x_j, \omega))] = 0 \]

holds for \( i = A, B \). Moreover,

\[ E[D_{AB}(x_B^o, U_B(x_A^o, x_B^o, \omega))] = E[U_B(x_A^o, x_B^o, \omega)] - E[U_B(x_A^o, x_B, \omega)] \]

holds for all unilateral deviations \((x_A, x_B^o)\) by \( A \) and

\[ E[D_{BA}(x_B)] = D_{BA}(x_B) \geq E[U_A(x_A^o, x_B, \omega)] - E[U_A(x_A^o, \omega)] \]

for all unilateral deviations \((x_A^o, x_B)\) by \( B \). Therefore, all compensation requirements are fulfilled so that this damages rule must be compensatory indeed.

As a drawback, damages \( D_{AB}(x_B, u_B) = E[U_B(x_A^o, x_B, \omega)] - u_B \) under this rule may be of either sign and, if negative, party \( B \) enjoys a benefit from \( A \)'s deviation. For the above damages rule to remain compensatory, such benefits would have to be returned (which would be unusual under tort law).

In the next section, a general implementability criterion based on but-for causation will be introduced. Any signal satisfying this criterion allows implementing compensatory damages rules under which benefits need not be returned.

## 4 But-for causation and implementability

Causation is meant relative to the same reference profile \( x^o \) as introduced for welfare comparisons in the previous section. Suppose parties have actually
chosen action profile \( x = (x_A, x_B) \in X \) and the state of nature is \( \omega \in \Omega \). Then
\[
\Delta_{ij}(x, \omega) = U_j(x_i^*, x_j, \omega) - U_j(x_i, x_j, \omega)
\]
is referred to as \textit{impact} caused (in the sense of the but-for test) by party \( i \)'s deviation from the reference profile. This difference can be of either sign. If \( \Delta_{ij}(x, \omega) > 0 \) then \( i \)'s deviation has caused a negative impact on party \( j \). Let \( \gamma_{ij}(x, \omega) \) denote the corresponding indicator function, that is \( \gamma_{ij}(x, \omega) = 1 \) if \( \Delta_{ij}(x, \omega) > 0 \) and \( \gamma_{ij}(x, \omega) = 0 \) else.

The signal \( \sigma : X \times \Omega \to S \) generating evidence may not be informative enough to detect deviations for sure, let alone the exact size of the impact from them. Suppose courts have obtained evidence \( s \), their conjecture being that \( x \) is the chosen action profile. They then can update their beliefs that the true state \( \omega \) of nature must be from the event
\[
\Omega^s(x) = \{ \omega \in \Omega : \sigma(x, \omega) = s \}.
\]
Let \( \pi^s(x) = \pi \{ \Omega^s(x) \} \) denote the probability of this event and \( E[\Delta_{ij}(x, \omega) \mid \Omega^s(x)] \) the conditionally expected impact from \( i \)'s deviation under the premise that \( x \) is the true action profile.

From the perspective of the compensation principle, only unilateral deviations \( (x_i, x_j^o) \) matter. Therefore, courts may restrict attention to unilateral deviations consistent with the observed evidence \( s \). Let \( X_i^s = \{ x_i \in X_i : \pi^s(x_i, x_j^o) > 0 \} \) denote the set of all unilateral deviations by \( i \) which could have generated evidence \( s \) with positive probability. If \( x_i \in X_i^s \) then there exists a state \( \omega_i \) such that \( s = \sigma(x_i, x_j^o, \omega_i) \).

Moreover, let \( S_j^o \) denote the subset of \( S \), for which \( X_i^s \) is not empty, that is, if \( s \in S_j^o \), then it cannot be ruled out that party \( j \) has kept to the reference profile.

Courts are aware that their conjectures on the chosen actions need not be correct. They know, however, that the least upper bound (maximum, if it exists) about all consistent conjectures amounts to
\[
\delta_{ij}(s) = \sup_{x_i \in X_i^s} E[\Delta_{ij}(x_i, x_j^o, \omega) \mid \Omega^s(x_i, x_j^o)].
\]
Notice, if \( \delta_{ij}(s) \leq 0 \), then party \( i \) cannot have caused any negative impact on \( j \), at least not if \( j \) has actually kept to the reference profile. With this notation at hand, the following result on implementability can be established.
Proposition 2  (i) Suppose the criterion \( \delta_{AB}(s) + \delta_{BA}(s) \leq 0 \) is met for all \( s \in S_A^o \cap S_B^o \). Then the signal allows implementing a compensatory damages rule.

(ii) If \( \delta_{AB}(s) \leq 0 \) and \( \delta_{BA}(s) \leq 0 \) hold even separately for all \( s \in S_A^o \cap S_B^o \), then any damages rule satisfying

\[
D_{ij}(s) = 0 \quad \text{if} \quad s \in S^o_i \quad \text{and} \quad D_{ij}(s) \geq \delta_{ij}(s) \quad \text{if} \quad s \in S^o_j \setminus S^o_i
\]

is compensatory, no matter how damages are specified for \( s \notin S_A^o \cup S_B^o \).

(iii) If the externality is of unidirectional nature from, say, \( A \) to \( B \). Then \( \delta_{BA}(s) = 0 \) for all \( s \in S_A^o \) and the implementability criterion of (ii) follows from the one of (i).

**Proof.** Notice, if \( \pi^s(x^o) > 0 \) then \( x^o_i \in X^o_i \) and \( s \in S^o_i \) for \( i = A, B \). Moreover, as \( \Delta_{ij}(x^o, \omega) = 0 \), it follows that \( \delta_{ij}(s) \geq 0 \) must hold for such \( s \). If, in addition, \( \delta_{AB}(s) + \delta_{BA}(s) \leq 0 \) then \( \delta_{ij}(s) = 0 \).

Compensatory damages rule are then constructed as follows. For \( s \in S_A^o \cap S_B^o \), if \( \delta_{AB}(s) \leq 0 \) and \( \delta_{BA}(s) \leq 0 \) then choose \( D_{AB}(s) = D_{BA}(s) = 0 \). For any other \( s \in S_A^o \cap S_B^o \) (in case (ii), such \( s \) do not exist), specify damages so that

\[
\delta_{AB}(s) \leq D_{AB}(s) - D_{BA}(s) \leq -\delta_{BA}(s) \quad (3)
\]

holds and, for \( s \in S^o_j \setminus S^o_i \), define \( D_{ji}(s) = 0 \) and \( D_{ij}(s) \geq \delta_{ij}(s) \). Notice, with this specification, (3) is satisfied for all \( s \in S_A^o \cup S_B^o \). Any such damages rule must be compensatory for the following reasons.

First, \( E[D_{AB}(\sigma(x^o, \omega))] = E[D_{BA}(\sigma(x^o, \omega))] = 0 \) holds if both parties keep to the reference profile because \( \delta_{ij}(s) = 0 \) has been shown to hold for such \( s \). Second,

\[
E[ U_B(x^o, \omega) - U_B(x_A, x^o_B, \omega) ] = E[ \Delta_{AB}(x_A, x^o_B, \omega) ] =
\]

\[
= \sum_{s \in S^o_B} \pi^s(x_A, x^o_B) \cdot E[ \Delta_{AB}(x_A, x^o_B, \omega) \mid \Omega^s(x_A, x^o_B) ] \leq
\]

\[
\leq \sum_{s \in S^o_B} \pi^s(x_A, x^o_B) \cdot \delta_{AB}(s) \leq \sum_{s \in S^o_B} \pi^s(x_A, x^o_B) \cdot [D_{AB}(s) - D_{BA}(s)] =
\]

\[
= \sum_{s \in S} \pi^s(x_A, x^o_B) \cdot [D_{AB}(s) - D_{BA}(s)] =
\]

\[
= E[D_{AB}(\sigma(x_A, x^o_B)) - D_{BA}(\sigma(x_A, x^o_B))] \]

holds for all unilateral deviations by party \( A \) so that \( B \) is compensated for
such deviations. For symmetric reasons, $A$ is also compensated for unilateral deviations by $B$ and, hence, the above damages rule must be compensatory, no matter how it is specified for $s \not\in S^o_A \cup S^o_B$.

(ii) Under the stronger condition of claim (ii), any damages rule as specified in the claim satisfies the condition of the first claim and, hence, must be compensatory as was to be shown.

(iii) Unidirectionality means that $A$’s utility $U_A(x_A, \omega)$ remains independent for $B$’s choice $x_B$ of action and, hence, $B$’s choice cannot cause any impact on $A$. Therefore $\delta_{BA}(s) = 0$ holds indeed. But if it does then $\delta_{AB}(s) + \delta_{BA}(s) = \delta_{AB}(s)$ for all $s \in S^o_A \cap S^o_B$, from which claim (iii) follows immediately.

The criterion of claim (i) is sufficient for implementability. In the absence of uncertainty, it is also necessary. In fact, suppose that profits $U_i(x, \omega) = U_i(x)$ and signal $\sigma(x, \omega) = \sigma(x)$ remain independent of any state and that damages rule $D(s)$ is compensatory. Without loss of generality, we may assume this damages rule to be normalized by $D(\sigma(x^o)) = [0, 0]$. Then

$$D_{ij}(\sigma(x^o_i, x_j)) - D_{ij}(\sigma(x^o_i, x_j)) \geq U_i(x^o) - U_i(x^o_i, x_j)$$

must hold for all unilateral deviations $x_j$ (for $i = A, B$). Since, for any $s \in S^o_A \cap S^o_B$, there exist $x_A$ and $x_B$ so that $\sigma(x^o_A, x_B) = \sigma(x_A, x^o_B) = s$, it follows that

$$U_A(x^o) - U_A(x^o_A, x_B) + U_B(x^o) - U_B(x_A, x^o_B) \leq [D_{BA}(s) - D_{AB}(s)] + [D_{AB}(s) - D_{BA}(s)] = 0$$

and, hence, that $\delta_{AB}(s) + \delta_{BA}(s) \leq 0$ necessarily must hold for all $s \in S^o_A \cap S^o_B$.

Under uncertainty, however, the criterion need not be necessary because compensation is required only from the ex ante perspective, but not necessarily ex post, that is after the evidence $s$ has been generated by signal $\sigma$.

The implementability criterion offered by proposition 2 (ii) is of particular interest. If it is met then it allows to check compensation requirements in each direction separately. This is in contrast to the Nash property which does not allow for such separation.

The signal $\sigma(x, \omega) = x$, introduced at the end of the previous section, satisfies criterion (ii). In fact, if $s \in S^o_j$ then there exists $x^*_i$ and $\omega_i$ such
that \( s = \sigma(x'_i, x'_j, \omega_i) = (x'_i, x'_j) \). Moreover, if \( s = \sigma(x, \omega) \in S^o_A \cap S^o_B \) then \( \sigma(x, \omega) = x^o \) and, hence, \( \delta_{ij}(s) = 0 \).

The signal
\[
\sigma^{bft}(x, \omega) = [\gamma_{AB}(x, \omega), \gamma_{BA}(x, \omega)]
\]
also satisfies criterion as will be shown by the next proposition. This signals attain values in \( S = \{0, 1\}^2 \) and reveals, whether or not a party’s deviation has caused (in the sense of the but-for test) a negative impact on the other party. More precisely, if \( \gamma_{ij}(x, \omega) = 1 \) then
\[
\Delta_{ij}(x, \omega) = U_j(x^o_i, x_j, \omega) - U_j(x, \omega) > 0
\]
so that \( i \) must have deviated and this deviation must have hurt party \( j \) according to the but-for test.

**Proposition 3** The signal \( \sigma^{bft}(x, \omega) \) satisfies the implementability criterion of proposition 2 (ii) and, hence, allows implementing a compensatory damages rule.

**Proof.** Suppose \( s \in S^o_A \cap S^o_B \). Then there exist \( x'_i \) and \( \omega_i \) such that
\[
s = \sigma^{bft}(x'_i, x^o_j, \omega_i) = \sigma^{bft}(x^o_A, x'_B, \omega_B)
\]
and, hence, \( s = [0, 0] \) because \( \gamma_{AB}(x^0_A, x'_B, \omega_A) = 0 \) and \( \gamma_{BA}(x'_A, x^0_B, \omega_A) = 0 \). Therefore, if \( s(x, \omega) = s = [0, 0] \) then \( \Delta_{ij}(x, \omega) \leq 0 \) and, hence, \( \delta_{ij}(s) \leq 0 \) must hold for the least upper bound as well. \( \blacksquare \)

The above proposition shows that courts can implement a compensatory damages rule provided that they know whether a party’s deviation has caused any negative impact on the other party at all. They need not know, though, its size.

Along similar lines, it can be shown that the signal
\[
\sigma^B(x, \omega) = [x_B, \gamma_{AB}(x, \omega)]
\]
revealing party \( B \)’s action and whether or not party \( A \)’s deviation has caused a negative impact on \( B \), would satisfy the criterion of proposition 2 (ii) as well.
5 Applications

5.1 Accident models and negligence rules

A general externality model \( U_A(x, \omega) \) and \( U_B(x, \omega) \) with action profiles \( x \) to be chosen by parties \( A \) and \( B \) from \( X = X_A \times X_B \) and state \( \omega \) chosen by nature from state space \( \Omega \) is referred to as an accident model if profit functions are of shape

\[
U_i(x, \omega) = -c_i(x_i) - H_i(x, \omega)
\]

for \( i = A, B \). Here, \( x_i \) denotes party \( i \)'s choice of precaution measures from the set \( X_i \) of alternatives, \( c_i(x_i) \) expected costs of these measures (precaution costs, care level), and \( H_i(x, \omega) \geq 0 \) total harm party \( i \) suffers from. If \( H_i(x, \omega) > 0 \) then an accident has occurred.

In Brown's (1973) pioneering version of the accident model, no distinction is made between precaution measures and precaution expenditures, that is \( c_i(x_i) = x_i \) so that the action consists of choosing a precaution level \( x_i \) from the (linearly ordered) real half-line \( \mathbb{R}_+ \). Expected harm \( E[H_i(x, \omega)] \) is assumed to be a monotonically decreasing function of both parties' precaution expenditures. Due levels \( x^o = (x^o_A, x^o_B) \) of care serve as reference profile so that party \( i \) is negligent if \( i \) has spent less on precaution than \( x^o_i \). Due to the assumed monotonicity, negligence \( x_i < x^o_i \) in terms of care levels is equivalent to \( E[H_j(x_i, x_j, \omega)] > E[H_j(x^o_i, x_j, \omega)] \) in terms of expected effectiveness of a precaution measure.

The negligence criterion in terms of effectiveness seems more appealing and, in particular, it remains applicable even if the action space is not linearly ordered. But notice, with more general action spaces, the criterion \( E[H_j(x_i, x_j, \omega)] > E[H_j(x^o_i, x_j, \omega)] \) for party \( i \)'s choice \( x_i \) may become dependent on the action \( x_j \) chosen by the other party \( j \).

From the incentive perspective (compensation principle), only unilateral deviations matter. Under negligence rules, a party keeping to the reference profile is not held liable whereas an unilaterally deviating party must replace total harm the other party suffers from. To implement such negligence rules literally, the signal should reveal the action profile and, in case of an accident, the level of harm as captured by signal

\[
\sigma^*(x, \omega) = [x_A, x_B, H_A(x, \omega), H_B(x, \omega)].
\]
At evidence \((x_i, x_j^*, h_i, h_j)\), party \(i\) owes damages

\[
D_{ij}(x_i, x_j^*, h_j) = h_j \text{ if } E[H_j(x_i, x_j^*, \omega)] > E[H_j(x_i^0, x_j^*, \omega)],
\]

and \(D_{ij}(x, h_j) = 0\) else, to party \(j\).

If a signal \(\sigma\) is less informative than \(\sigma^*\) but still sufficient for implementing a compensatory damages rule then the notion of a negligence rule may be adapted as follows. Suppose that \(\delta_{ij}(s) \leq 0\) holds at given evidence \(s\). Then party \(i\) can be ruled out as having caused any negative impact on \(j\) (provided that \(j\) herself has actually kept to the reference profile). Under the extended negligence criterion, party \(i\)’s choice of action is deemed non-negligent and \(i\) is not held liable whenever \(\delta_{ij}(s) \leq 0\).

Otherwise, that is, if \(\delta_{ij}(s) > 0\), then party \(i\)’s choice of action is deemed negligent and \(i\) is held liable for total harm. The available evidence need not reveal the exact size of total harm, in which case the least upper bound consistent with the evidence may serve as a proxy. Formally, it is defined as the supremum of the conditionally expected total harm

\[
\eta_{ij}(s) = \sup_{x_i \in X_i^s} E\left[ H_j(x_i, x_j^*, \omega) \mid \Omega^s(x_i, x_j^*) \right]
\]

over all unilateral deviations \(x_i\) from \(X_i^s\) that could possibly have generated evidence \(s\). By definition, total harm is never negative and, hence, \(\eta_{ij}(s) \geq \delta_{ij}(s)\) holds for all \(s\).

Under the extended version of the negligence rule, party \(j\) would be entitled to recover \(\eta_{ij}(s)\) from \(i\) if \(i\)’s behavior was negligent (that is, if \(\delta_{ij}(s) > 0\)) whereas \(j\) cannot recover anything from \(i\) otherwise (that is, if \(\delta_{ij}(s) \leq 0\)).

This extension of the negligence rule would be compensatory whenever the implementability criterion of proposition 2 (ii) is met. Notice, under this negligence rule, victims would recover beyond what is needed for satisfying the compensation requirement of the compensation principle. Yet, as follows from proposition 2 (ii), overcompensation does not distort incentives under the above negligence criterion.

### 5.2 Activity-precaution models

As a second application, consider the following version of an activity-precaution model where party \(i\) faces two-dimensional choice \(x_i = (a_i, p_i)\) from \(X_i =\)
The decision $a_i$ denotes $i$'s activity level, the decision $p_i$ her precaution expenditures per unit of activity ($i = A, B$). Let $x = (x_A, x_B)$ denote the entire action profile, $a = (a_A, a_B)$ the profile of activity levels and $p = (p_A, p_B)$ the profile of precaution expenditures.

The utility functions are of shape

$$U_i(x, \omega) = v_i(x_i) - a_i \cdot a_j \cdot H_i(p, \omega)$$

well in line with the precaution-activity model (see Shavell 1980, 1987). The function $v_i(x_i)$ denotes $i$'s expected profit (before harm) from operating her activity at level $a_i$, net of own total precaution expenditures $a_i \cdot p_i$.

Shavell and the subsequent literature deal with damages rules, which condition exclusively on the precaution profile $p$ and total harm $h_i = H_i(p, \omega)$ party $i$ actually suffers from, to show that the efficient profile fails being a Nash equilibrium.

Suppose it is for lack of information that damages do not condition on activity levels. Then the signal

$$\sigma(a, p, \omega) = [p_A, p_B, H_A(p, \omega), H_B(p, \omega)]$$

revealing precaution expenditures $p = (p_A, p_B)$ per unit of activity as well as the actual levels of harm, the two parties suffer from in case of an accident, would be appropriate. As will be shown below, signal (5) violates the implementability criterion of proposition 2.

For implementing a compensatory damages rule, the signal would have to reveal, on top of precaution expenditures, whether or not activities have been operated at excessive activity levels. This is captured by signal

$$\sigma(a, p, \omega) = [p_A, p_B, \lambda_A(a_A), \lambda_B(a_B)]$$

where $\lambda_i(a_i) = 1$ if $a_i > a^*_i$ and $\lambda_i(a_i) = 0$ else.

Proposition 4 For the above activity-precaution model, the signal (6) satisfies the implementability criterion of proposition 2 (ii) whereas the signal (5) does not.

Proof. Consider any evidence $s$ from $S^o_A \cap S^o_B$. This means that

$$s = \sigma(x'_A, x'_B, \omega_B) = \sigma(x'_A, x'_B, \omega_A)$$

(7)
holds for some actions $x_i'$ and states $\omega_i$. Given signal (6), it then follows that $p' = p^o$ and $\lambda_i(a_i') = \lambda_i(a_i^o) = 0$ must hold for $i = A, B$. Therefore, if $x$ and $\omega$ are such that $\sigma(x, \omega) = s = [p^o, 0, 0]$ then $a_i \leq a_i^o$ and

$$U_j(x^o, \omega) - U_j(x_i, x_j^o, \omega) = a_j^o \cdot (a_i - a_i^o) \cdot H_j(p^o, \omega) \leq 0.$$  

It then follows that $\delta_{ij}(s) \leq 0$ must also hold for such $s$ and, hence, the implementability criterion of proposition 2 (ii) is satisfied for $i = A, B$. This settles the first claim.

As for the second claim, if (7) holds for signal (5) then $p' = p^o$ as well as $h_i = H_i(p^o, \omega_j) = H_i(p^o, \omega_i)$ must hold for $i = A, B$. Suppose this evidence is actually generated by profile $x$ and state $\omega$, that is $s = \sigma(x, \omega)$. It follows that $p = p^o$ and $H_i(p^o, \omega) = h_i$ so that the impact from any unilateral deviation $a_i$ by party $i$ on party $j$ amounts to

$$U_j(x^o, \omega) - U_j(x_i, x_j^o, \omega) = a_j^o \cdot (a_i - a_i^o) \cdot h_j$$

and is strictly positive if $i$ operates his activity at an excessive level $a_i > a_i^o$. It follows that the least upper bound $\delta_{ij}(s) > 0$ must be strictly positive all the more so that the implementability criterion of proposition 2 would be violated indeed. ■

Even if courts can observe, whether or not activities have been operated at excessive levels, on legal grounds, they may be reluctant taking activity levels into account while quantifying damages. From an incentive perspective, they should not be reluctant as the above proposition suggests.

5.3 Acquiring information ex ante

Kaplow and Shavell (1996) have examined an unilateral accident model where the injurer may acquire information before deciding on precaution. The setting of the present paper allows generalizing their findings at least as far as acquiring information ex ante is concerned.

While not ruling out bidirectional externalities, I follow Kaplow and Shavell, though, by assuming that only one party, say $A$, has the option to acquire information ex ante, that is before choosing action $y$ from the set $Y$ of alternatives. Party $A$’s information acquisition decision is denoted by $a \in \{0, 1\}$. If $A$ acquires information ($a = 1$), then $A$ privately learns, which
event $\Omega_t$ from the partition $\Omega = \Omega_1 \cup \ldots \cup \Omega_T$ of state space the true state $\omega$ belongs to. For lack of time or for other reasons, party $A$ cannot reveal the obtained information $t$ to party $B$ before $B$ must choose her action $x_B$ from $X_B$. If $a = 0$, then $A$ remains equally (un-) informed as $B$.

To describe this situation as a special case of the general model, the action space of injurer $A$ is identified with the set of all complete contingent plans $X_A = \{0, 1\} \times Y^{T+1}$ to be interpreted as follows. Action $x_A = (a, y_0, ..., y_T) \in X_A$ means that $A$ chooses $y = y_0 \in Y$ if $a = 0$ and $y = y_t$ if $a = 1$ and $A$ has learned the true state $\omega$ to be from the event $\Omega_t$. Party $B$ as the victim, remaining uninformed, still chooses her action $x_B$ from the set $X_B$. Profit and utility are state contingent functions $U_i(x, \omega)$ of the action profile $x = (x_A, x_B)$ so that not even bidirectional externalities need be ruled out.

By their very nature, complete contingent plans can hardly ever be observed. Therefore, in the setting with information acquisition, we focus on the signal (4) which reveals the victim’s action and whether or not $A$’s deviation from the reference profile has caused any negative impact on $B$. This signal allows implementing a compensatory damages regime quite generally as shown at the end of section 4.

Suppose the reference profile $x^o = (a^o, y^o_0, ..., y^o_T, x^o_B)$ requires party $A$ to seek information, that is $a^o = 1$. Then a compensatory damages rule can be implemented and, as follows from the compensation principle, such rules will be welfare enhancing. In particular, if the ex ante acquisition of information is socially valuable, then party $A$ has the incentive to acquire information under any damages rule compensatory relative to $x^o$.

For illustration, consider the following special case of the above model. Profit functions are of shape

$$U_A(x, \omega) = V_A(y, x_B, t) - c \cdot a$$ and $$U_B(x, \omega) = V_B(y, x_B, t)$$

for $\omega \in \Omega_t$ with the following interpretation. The parameter $c$ captures $A$’s private costs of becoming informed. Profits $V_i(y, x_B, t)$ (before information costs) are type-contingent and depend exclusively on the actions $y$ and $x_B$ actually chosen by $A$ and $B$ (not on the entire contingent plan).

In this setting, the signal $\sigma(x, \omega) = [y, x_B, t]$ would be at least as informative as signal $\sigma^B(x, \omega)$ and, hence, would allow implementing a compensatory damages rule all the more. This signal expresses that courts can observe the
actions actually chosen by the two parties as well as the true type $t$. Therefore, while it costs $c$ to learn the type ex ante (before the decisions are due), ex post, courts learn this type for free.

5.4 Burden of proof

In this subsection, the model of Kaplow (2012) is confronted with the general approach of the present paper. In Kaplow’s model, only one party, say $A$, reaches a binary decision $a$ from $\{0, 1\}$ where $a = 1$ means that $A$ engages in the activity with private benefit $U_A(\omega)$ and $a = 0$ means that $A$ does not. Party $A$’s activity imposes impact $U_B(\omega)$ on party $B$ (all members of society but $A$) who otherwise remains passive (unilateral decision by $A$). Evidence $s = \sigma(\omega)$ is generated by signal $\sigma : \Omega \rightarrow S$ if $A$ engages in the activity.

Activities may either be harmful (type $t = h$) or benign (type $t = b$). In the present setting, this is captured by a partition $\Omega = \Omega_h \cup \Omega_b$ of state space $\Omega$. State $\omega$ belongs to $\Omega_h$ if $U_A(\omega) + U_B(\omega) < 0$ and $\omega$ belongs to $\Omega_b$ else. Party $A$ knows the type $t$ of his activity before deciding on whether to engage in it or not. Party $A$ of type $t$ expects private benefits

$$u_{At} = E[U_A(\omega) \mid \Omega_t]$$

from engaging in the activity.

Kaplow’s design problem consists of specifying a subset $S_L$ of $S$ so that party $A$ is held liable if and only if $a = 1$ and the evidence $s$ belongs to this subset. Moreover, if $A$ is held liable, he must pay an amount $f(s)$ as a sanction. Think of a fine or of damages. Kaplow treats this amount $f(s)$ as exogenously given.

Under preponderance of the evidence, for instance, the legal system would hold $A$ liable at evidence $s$ if and only if the probability of receiving evidence $s$ is higher if $A$’s act was harmful than if it was benign, that is, if the corresponding conditional probabilities satisfy the inequality

$$\pi\{\sigma(\omega) = s \mid \Omega_b\} < \pi\{\sigma(\omega) = s \mid \Omega_h\}.$$ 

Given any subset $S_L$ of $S$ (not necessarily in line with preponderance), party $A$ of type $t$ engages in the activity if and only if the expected private benefit
is not lower than the expected sanction, that is, if and only if

\[ u_{AI} \geq \sum_{s \in S_L} p_t(s) \cdot f(s) \]

where \( p_t(s) = \pi \{ \sigma(\omega) = s \mid \Omega_t \} \) denotes the ex ante probability of evidence \( s \) if \( A \)'s activity is of type \( t \).

Let \( a_t(S_L) \in \{0, 1\} \) denote \( A \)'s decision as a function of the subset \( S_L \) which, in turn, is specified by the legal system. Then the expected welfare amounts to

\[
W(S_L) = \pi \{ \Omega_b \} \cdot E [U_A(\omega) + U_B(\omega) \mid \Omega_b] \cdot a_b(S_L) + \\
+ \pi \{ \Omega_b \} \cdot E [U_A(\omega) + U_B(\omega) \mid \Omega_b] \cdot a_b(S_L)
\]

as a function of \( S_L \). Instead of preponderance of the evidence, Kaplow suggests to use the subset \( S_L \) for which the expected welfare \( W(S_L) \) is highest.

I propose instead to look, first, for a damages rule \( D_{AB}(s) \) compensatory relative to the efficient decision, \( a^e(\omega) = 1 \) if \( \omega \in \Omega_b \) and \( a^e(\omega) = 0 \) else. A compensatory rule has to satisfy the compensation requirements

\[
E [U_B(\omega) + D_{AB}(\sigma(\omega)) \mid \Omega_b] \geq 0 \geq E [U_B(\omega) + D_{AB}(\sigma(\omega)) \mid \Omega_b].
\]

Of course, not all signals are informative enough to implement such a rule, in which case the compensation principle is of no use. The signal \( \sigma^B(\omega) = U_B(\omega) \) revealing the actual impact \( u_B = U_B(\omega) \) on party \( B \), however, allows implementing the damages rule \( D(u_B) = -u_B \), which is compensatory relative to the efficient decision and, as follows from the compensation principle, provides efficient (first best) incentives for \( A \).

With an exogenously given fine \( f(s) \), instead, compensation requirements need not be fulfilled and the rule based on this fine need not be welfare enhancing, no matter how the levels \( S_L \), at which \( A \) is held liable, are specified.

In other words, Kaplow’s rule may end up being second best even in informational settings that would allow first best (such as signal \( \sigma^B(\omega) = U_B(\omega) \) above). To implement first best, imposing sanctions for allegedly harmful acts may not be enough. In addition, rewards for engaging in benign activities may also be needed. Tort law, in general, is not prepared to grant rewards for benign acts. On this account, Kaplow’s second best solution has merit even if an economic explanation why tort law concentrates on harmful acts while neglecting benign ones seems missing.
5.5 Scope of negligence

So far, evidence has been generated by a given signal. But courts may have the option to affect the quality of this signal. Dari-Mattiacci (2005) has proposed a model that endogenizes the scope of negligence in the following sense. Precaution choice is multi-dimensional where party $i$ chooses $N_i$-dimensional precaution from the non-negative orthant $\mathbb{R}_{+}^{N_i}$ for $i = A, B$. Courts decide on the dimensions $n_i \subset \{1, \ldots, N_i\}$ (referred to as scope of negligence) entering the negligence criterion. Only damages rules are considered which induce parties to choose zero precaution in all dimensions which do not enter the criterion. Social costs to install a signal $\sigma^n$ revealing precaution in those dimensions within the scope $n = (n_A, n_B)$ of negligence amount to $C^n$. Let $X^n_i$ denote the subset of $\mathbb{R}_{+}^{N_i}$ with zero precaution in all dimensions outside of scope $n_i$ and $X^n = X^n_A \times X^n_B$ the set of corresponding profiles.

For any scope $n = (n_A, n_B)$, consider a reference profile $x^n$ which maximizes expected welfare

$$W^n = \max_{x \in X^n} E[U_A(x, \omega) + U_B(x, \omega)]$$

over all action profiles from the set $X^n$. The signal $\sigma^n$ allows implementing a damages rule which is compensatory relative to $x^n$ if compensation requirements are restricted to profiles from $X^n$. Any such rule ensures expected welfare $W^n$ in Nash equilibrium of the game induced by the rule, as follows from the compensation principle.

The scope $n^* = (n^*_A, n^*_B)$ of negligence is predicted which maximizes $W^n - C^n$. Along these lines, Dari-Mattiacci’s findings on the scope of negligence can easily be extended to the general externality setting of the present paper.

6 Concluding remarks

This paper has made extensive use of the compensation principle according to which compensation requirements relative to a reference profile are sufficient for a damages rule to be welfare enhancing relative to this profile. In a literal sense, these requirements are not necessary, that is welfare enhancing damages rules may exist, which need not be compensatory. Yet, all efficient or welfare enhancing damages regimes from the literature, which I have checked so far, have turned out being compensatory. Therefore, not
much (if anything) seems lost while much is gained in terms of analytical simplicity and generality of action spaces by concentrating on compensatory damages rule.

The limits of the compensation principle are reached, however, whenever courts must decide cases based on evidence not informative enough for implementing any compensatory damages rule at all. In such cases, exploring incentives under damages rules leads to tedious second best considerations where general insights tend to be rare. Except for inefficient reference profiles, the present paper does not address questions of second best. It rather characterizes the type of information needed to implement first best.

7 References


