Safety Regulation and Monitor Liability*

Eberhard Feess
Aachen Technical University†

Ulrich Hege
HEC School of Management
and CEPR‡

This version: July 2001

* We are grateful to Doug Bernheim, Lewis Kornhauser, Kai-Uwe Kühn, an associate editor and an anonymous referee for helpful comments and suggestions. Financial support from the European Commission (TMR Grant ERBFMBICT 972827) and from Thyssen Foundation is gratefully acknowledged.

† Address: Department of Economics, Aachen Technical University, D-52062 Aachen, Templergraben 64, Germany. E-mail feess@rwth-aachen.de.

‡ Address: HEC School of Management, Department of Finance and Economics, F-78351 Jouy-en-Josas Cedex, France. E-mail hege@hec.fr.

Abstract

We propose a simple liability rule when several agents are jointly responsible for monitoring a risky economic activity or certifying its security. Examples are safety controls for drugs or technical systems, environmental liability, or air safety accidents. The agents have private knowledge of their monitoring or avoidance costs. We adopt a mechanism design approach to ensure optimal monitoring incentives. Our innovation is to focus on information that is available or can be proxied when harm has occurred and when typically regulators and/or courts deliberate over fines and damages. By contrast, earlier proposals require more estimations of hypothetical accident scenarios and their ex ante probabilities. We argue that our rule promises substantial savings in information costs for courts and regulators and excludes likely sources of errors.

Key words: Safety regulation, multiple monitors, Bayesian mechanism.

JEL classification: K13, H41

1. Introduction

We analyze efficient liability rules when several agents are jointly responsible for carrying out a risky economic activity or monitoring its security. As an example, consider the classical DES case. Between 1941 and 1971, about 200 drug companies produced and marketed Diethylstilbestrol (DES) that was prescribed to prevent miscarriage among expectant mothers. It was only in the late 1960s that DES was discovered to cause a vaginal cancer in women whose mothers had taken DES. Any of the manufacturers bore responsibility, since all had failed to carry out thorough enough testing that would have revealed the cancer risk, as was acknowledged by the courts who considered apportioning liability according to market shares (see Cooter and Ulen 1988, 339). All manufacturers were responsible for controlling the safety of the drug, and one company's discovery of the risk would have been sufficient to prevent the spread of vaginal cancer.

Environmental regulation is another area where the safety of procedures and installations is independently monitored by multiple agents. Typically, companies constantly monitor the safety of hazardous plants. In parallel, the regulating agency performs regular or random controls. Finally, insurers also impose standards of safety controls, and they may monitor these standards themselves.

Independent security controls as in these examples are only special cases of a wider class of closely related problems. In other cases, monitors may not be in charge of the *same* tests, but their responsibilities overlap in the sense that one agent has reasons to believe that the control activity of others would be sufficient to guarantee safety. For instance, in case of air safety accidents, there are various layers of security controls, like manufacturers, the airline, the ground staff, the service company, the air traffic control, the safety regulator, the pilot, and it frequently takes the simultaneous failure of two or more of them to trigger an accident.

Designing an efficient liability rule would be trivial if information were perfect. Any kind of negligence rule, where each agent has to pay if she violates the due care level, implements the efficient care choices.[1] However, efficient negligence rules require that the court is perfectly informed about the various agents' avoidance costs. In the DES case, it was difficult to find out what the reasonable standard of care had been. Among the many possible risks of the new drug, which ones deserved special attention at the time? What prior knowledge about the risks did the manufacturers really have? How competent were their test technologies, and hence, what would have been the costs to develop tests that would have been really effective? On all these questions, the drug manufacturers

[1] See e.g. Shavell (1987).

are likely to hold private information. The same is true for our environmental regulation example. Typically, all parties will keep records of their monitoring activity, as a precaution in the event of damage. But if the controls fail and an environmental disaster occurs, the question will be how competent the test procedures and the monitoring staff really were, or translated in economic terms, what the true costs of effective monitoring were for any of the agents in charge. For these reasons, we analyze liability rules for multiple agents when monitoring costs are private information.

We restrict our attention to liability rules without punitive damages, that is to liability rules where total liability payments must not exceed total harm.[2] We refer to this condition as *balancedness*, employing the usual term in the mechanism design literature. This literature has shown that solutions exist which give agents the right incentives and satisfy balancedness. D'Aspremont and Gerard-Varet (1979) (AGV henceforth) have suggested an optimal mechanism which does precisely that. The idea is that agents pay for the entire damage that their own actual choice of monitoring effort causes, relative to an expected choice of effort. This gives agents the desired incentives at the margin. Any surpluses collected in this way are transferred to other agents in a way that ensures that the recipients cannot manipulate the transfers.

The problem with the AGV-mechanism and related contributions (e.g. Demski and Sappington (1984) and Emons and Sobel (1991)) is that it is rather demanding, if it should be used in practice. It requires to assess what agents' efficient behavior would have been, how it would have translated into expected damage, and how this expected damage is affected if agents deviated from their expected behavior. Our main contribution is to show that for the problem at hand, a considerably less complex version of the AGV-mechanism can be used. Similar to an AGV mechanism, we construct side payments such each agent bears total accident costs in expectation, and hence acts efficiently, but these side payments use only the minimal amount of information needed.

Specifically, our liability rule has two attractive features. First, the rule is relatively simple because all side payments are linear in the ratios between the agents' expected and actual failure risks. Second, it is sufficient to make the liability payments dependent on the *actual* accident. This means that the court can safely ignore the impact of hypothetical accident scenarios. We develop our rule for independent security controls where farm occurs only if all of many layers in a safety net fail jointly. Afterwards, we demonstrate that the rule can also be applied to more general patterns of damage functions where the agents' impacts

[2] Punitive damages are not only legally controversial (see Polinsky and Shavell (1998)), but also problematic from an economic point of view: Since injurers will pay more than the total harm caused to society, it will discourage potential injurers from investing into socially beneficial economic activities.

on the damage function are not separable. Unfortunately, the liability rule then becomes more complicated as nothing like an individual failure risk exists for general multicausal damage functions. We conclude that our suggestion is most useful for independent security controls.

Besides the literature on AGV-mechanisms, our paper is related to the law and economics literature on optimal liability rules for accidents with multiple tortfeasors. Shavell (1987) and Landes and Posner (1987) demonstrate that negligence rules solve the problem under perfect information. Kornhauser and Revesz (1989, 1994) and Arlen (1992) extend the analysis to different sharing rules and settlements. Finsinger and Pauly (1990), and Polinsky and Che (1991) propose liability rules with punitive damages. Feess and Hege (1999) consider the moral hazard case and show that the team production problem can be solved through insurance contracts.

The paper is organized as follows. A simple example with only two security agents is presented in Section 2. We introduce the general model and discuss the main result in Section 3. Section 4 extends the analysis to more general patterns. Section 5 concludes.

2. An Example: Two Agents

2.1. The Model

We begin with the simple case of two risk-neutral agents, in order to develop the intuition for our rule. Two agents $i = 1, 2$, are in charge of monitoring an economic activity that produces harm x if the agents fail to detect errors. Agent i's verifiable effort a_i determines the probability $p(a_i)$ that an error escapes her attention. Both agents monitor the activity independently, and harm x will only occur if both agents fail to stop the causes or security gaps. Hence, the probability for the harm occurring is $p(a_1, a_2) = p(a_1) \cdot p(a_2)$.

Agent i's effort costs $c^t(a_i)$ depend on her type t that is private information. There are two possible cost types, high, $t = h$, or low, $t = l$. We assume that $c^h(a_i) > c^l(a_i)$ for all levels of a_i. Outsiders and regulators believe that $t = h$ with probability f_i^h and $t = l$ with probability $f_i^l = 1 - f_i^h$. The types of the two agents are independently distributed. To ensure an interior optimum, we assume

$$\frac{dp(a_i)}{da_i} < 0,\ \frac{d^2p(a_i)}{da_i^2} > 0,\ \frac{dc^t(a_i)}{da_i} > 0,\ \frac{d^2p(a_i)}{da_i^2} > 0\ \forall i,\ \forall t.$$

Social costs are defined as

$$C = c^t(a_1) + c^t(a_2) + p(a_1)\, p(a_2)\, x\,,$$

and depend on the types and on the effort choices, t_1, t_2, a_1 and a_2. Our assumptions ensure that C is strictly concave in a_1 and in a_2.

Let a_i^t denote the efficient choice of action for i, given her cost function is of type $t = (h, l)$. Interim efficiency requires that agent i minimizes expected social costs, given that each of the types of agent j takes her efficient action. Let $\bar{p}_j = f_j^l\, p(a_j^l) + f_j^h\, p(a_j^h)$ denote the expected failure probability of agent j in an interim efficient equilibrium. Thus, a_i^t is agent i's choice of action minimizing

$$p(a_i)\ \bar{p}_j\ x + c^t(a_i)\,. \tag{2.1}$$

This leads to the first order condition determining a_i^t:

$$\bar{p}_j\, x\, \frac{dp(a_i)}{da_i} = -\frac{dc^t(a_i)}{da_i} \tag{2.2}$$

Eqn.(2.2) expresses the familiar condition that agent i's optimal effort requires that her marginal costs of care be equal to the marginal decrease in expected social harm. The marginal decrease in expected social harm is based on the belief that the agent j will also exercise her optimal effort, a_j^t, and therefore fail to detect errors with an expected probability of $\bar{p}_j = f_j^l\, p(a_j^l) + f_j^h\, p(a_j^h)$.

We are looking for a *liability rule* that minimizes total costs C. Let $l_i(a_1, a_2)$ denote the liability attributed to agent i if the harm x occurs. The liability rule $l_i(a_1, a_2)$ is called *efficient* if it leads to agent i choosing her interim efficient effort level a_i^t. We restrict our attention to liability rules where victims are exactly compensated, i.e. where $l_1(a_1, a_2) + l_2(a_1, a_2) = x\,,\ \forall a_1, a_2$ *(balancedness)*. The liability rule $l(a_1, a_2)$ must implement the best attainable effort profile(a_i, a_j) as a Bayesian Nash equilibrium after the rule is announced.

2.2. An Optimal Liability Rule

In this example with two agents, our efficient liability rule will take the following form:

$$l_i(a_1, a_2) = \frac{x}{2}\left[1 + \frac{\bar{p}_j}{p(a_j)} - \frac{\bar{p}_i}{p(a_i)}\right], \quad i = 1, 2 \tag{2.3}$$

The rule consists of three components. First, agent i pays half of the damage, $x/2$. Second, half of damages is multiplied by the ratio of agent j's *expected* failure probability $\bar{p}_j$, divided by j's *actual* failure probability $p(a_j)$. This ratio, $\bar{p}_j/\,p(a_j)$, can be smaller or larger than one, and it magnifies or reduces the part that agent i has to pay in total. In the third term, the same ratio is applied to agent i herself, but since this term has a negative sign, it can be viewed as a transfer that agent i receives from agent j. Throughout, we refer to the second and the third part as side payments.

The *failure ratios* $\bar{p}_j / p(a_j)$ and $\bar{p}_i / p(a_i)$ are the essential element of our mechanism. Agent i's *failure ratio* can be interpreted as the *percentage* increment by which i's expected error tolerance, $\bar{p}_i$, fell short or exceeded i's actual error tolerance $p(a_i)$. This ratio is larger than one if agent i was more careful than expected, and vice versa.

Efficiency requires that each type minimizes expected social costs, given that the other player choose her type-specific efficient action. In the remainder of this Section, we will demonstrate why $l_i(a_1, a_2)$ satisfies this requirement. When choosing her effort, agent i minimizes[3]

$$\tilde{C}_i^t = p(a_i)\, \bar{p}_j\, l_i(a_1, a_2) + c^t(a_i)\,,$$

Substituting $l_i(a_1, a_2)$ from Eqn. (2.3) leads to

$$\tilde{C}_i^t = p(a_i)\, \bar{p}_j\, \frac{x}{2} \left[1 + \frac{\bar{p}_j}{p(a_j)} - \frac{\bar{p}_i}{p(a_i)}\right] + c^t(a_i) \tag{2.4}$$

or

$$\tilde{C}_i^t = \bar{p}_j\, \frac{x}{2} \left[p(a_i) + p(a_i)\frac{\bar{p}_j}{p(a_j)} - \bar{p}_i\right] + c^t(a_i) \tag{2.5}$$

When calculating her efficient care level, agent i is uncertain about agent j's type or effort $p(a_j)$, and she replaces them by expectations $p(a_j) = f_j^l p(a_j^l) + f_j^h p(a_j^h) = \bar{p}_j$. Thus, she assesses the value of the first two terms in the objective square bracket of (2.5) as $p(a_i) + p(a_i)E\left[\frac{\bar{p}_j}{p(a_j)}\right] = p(a_i) + p(a_i)\frac{\bar{p}_j}{\bar{p}_j} = p(a_i)2$. Agent i will operate with the following simple objective function

$$\tilde{C}_i^t = p(a_i)\ \bar{p}_j\ x\ - \bar{p}_i\ \bar{p}_j\ \frac{x}{2} + +c^t(a_i)\,, \tag{2.6}$$

where $\bar{p}_i\ \bar{p}_j \frac{x}{2}$, that is the part that agent i expects to get from agent j, is independent of i's actual behavior since $p(a_i)$ cancels out. It is then immediate that the first order condition of agent i's objective function (2.6),

$$\frac{d\tilde{C}_i^t}{da_i} = 0 \quad\Leftrightarrow\quad \frac{dp(a_i)}{da_i}\bar{p}_j\, x = -\frac{dc^t(a_i)}{da_i}\ ,$$

is identical to the first-order condition (2.2) of social cost minimization.

We add three remarks to sharpen the understanding of liability rule (2.3). First, since the first term and the second term of liability rule (2.3) add up to x in expectations, one

[3]The tilde on $\tilde{C}_i^t$ indicates that this is i's private objective function, as opposed to the social cost function (2.1).

may wonder why these terms are not aggregated immediately. The reason is that simply defining $l_i(a_1, a_2) = \frac{x}{2} + \frac{x}{2}\frac{\bar{p}_i}{p_i(a_i)}$ instead would imply punitive damages of x. And while the second term $\frac{x}{2}\frac{\bar{p}_j}{p(a_j)}$ does not influence agent i's care level, it is clearly important to make sure that agent j behaves efficiently.

Second, i's side payment $\frac{\bar{p}_j}{p_j(a_j)}x$ is decreasing in the agent j's actual failure risk $p_j(a_j)$ compared to agent j's expected failure risk $\bar{p}_j$. This is a desirable property: The higher agent j's expected failure risk, the more important that agent i detects the fault, the higher her efficient liability risk. But the lower agent j's actual failure risk, the lower the probability that harm occurs, i.e. the higher i's optimal payment if a damage happens despite agent j's low failure risk. Note that it cannot be sufficient to make agent i's payment simply depend on agent j's actual failure probability alone, because agent i takes expectations in equilibrium, and variable expected liability payments would no longer add up to total harm.

Third, we have only considered a symmetric liability rule, where each agent is expected to pay half of the harm. However, the liability rule can easily be generalized to any rule of the form $l_i(a_1, a_2) = x\left[\alpha_i + (1-\alpha_i)\frac{\bar{p}_j}{p(a_j)} - \alpha_i\frac{\bar{p}_i}{p(a_i)}\right]$, where $0 < \alpha_i < 1$, as we will see. Expected contributions are asymmetric then, with agent i paying an expected fraction of α_i of the damages.[4]

3. The General Case of Independent Security Controls

In this Section, we extend the analysis to $n \geq 2$ agents and to k different accident scenarios, where scenario k leads to harm $x^k \in (x^i, ..., x^K)$. Each of the n agents $i = 1, ..., n$ chooses an effort $a_i \in A_i$, determining the probability $p^k(a_i)$ that agent i's care fails with respect to a risk leading to harm x^k. We assume that harm x^k depends on the care of $m \leq n$ agents, and we maintain the assumption that each agent is independently monitoring *all* possible elements that could lead the economic activity to cause harm. The probability of harm x^k occurring is then

$$p^k(a) = \prod_i p^k(a_i)\,,$$

where $a = (a_1, ..., a_i, ..., a_m)$ is the vector of effort choices of all m agents. We construct a ring of the m agents, such that each has exactly one predecessor and one successor. If agent i is not involved in avoiding harm x^k, her respective probability is simply $p^k(a_i) = 1$.

[4]This is easily seen by recalling that the ex ante expectated values of the ratios $\overline{p}_j \,/\, p(a_j)$ and $\overline{p}_i \,/\, p(a_i)$ are equal to one.

Agent i's cost function $c^t(a_i)$ is private knowledge and depends on her type t_i, drawn from an agent-specific and discrete set T_i, and distributed according to an agent-specific marginal probability f_i^t. Agents' types are *independently* distributed. The interim efficient choice of action for type t_i is denoted a_i^t, and minimizes social costs

$$c^t(a_i) + \sum_k p^k(a_i)\, x^k \prod_{j \neq i} \bar{p}_j^k \quad ,$$

where $\bar{p}_j^k = \sum_t f_j^t\, p^k(a_j^t)$ is the expected failure probability of agent j, i's successor in the ring of agents, with respect to harm x^k.

Let $l_i(x^k, a)$ denote i's liability payment if harm x^k occurs, and if the combined effort profile of all $m \leq n$ agents was a. We can then immediately proceed to the statement of the efficient liability rule:

Proposition 1. *Suppose $m \leq n$ agents could have contributed to prevent harm x^k. Then the following liability rule is efficient:*

$$l_i(x^k, a) = x^k \left[\alpha_i + (1 - \alpha_i) \frac{\bar{p}_i^k}{p^k(a_i)} - (1 - \alpha_h) \frac{\bar{p}_h^k}{p^k(a_j)} \right] \quad \begin{array}{l} \textit{for } h = i - 1 \mod m, \\ \textit{for } j = i + 1 \mod m, \end{array}$$

where $0 < \alpha_i < 1$, $\sum_{i=1}^m \alpha_i = 1$.

Proof. *See the Appendix.*

In fact, the rule is practically the same that we presented for the two agents case, except that we now arranged the m agents in a circle (also known as *tweed ring*) in order to match their transfers. Each agent is the sender of a transfer to his predecessor (the second component in $l_i(x^k, a)$) and the receiver of a transfer from his successor (the third component). In addition, there are now many accident scenarios, and the liability rule uses the data of only the scenario that actually occurred. Hence, x is substituted by x^k. But as far as a single agent is concerned, her expected payment $p^k(a)\, l_i(x^k, a)$ depends on her own effort choice a_i in exactly the same fashion as in the two-agents example, so we can refer to our argument there as to why efficient incentives are assured. Moreover, the rule satisfies budget balancing, $\sum_i l_i(x^k, a) = x^k$, since all side-payments cancel out and the average net contribution of each agent is $\alpha_i\, x^k$.

As to the advantages of our rule, recall that a standard application of the AGV-mechanism would require taking expectations over all possible damages and action vectors. By contrast, for our rule it is sufficient to consider only the *actual* scenario x^k when assigning liability payments to the agents. Furthermore, the rule only requires estimating the *failure ratios* $\bar{p}_i^k \,/\, p^k(a_i)$ of each agent i and not actual probabilities. The use of ratios offers two practical advantages when proceeding to fill in numbers.

1. The risk of miscalculating the *level* of the probabilities is excluded or largely diminished, and mistakes are likely to occur only with respect to the *slope* of the probabilities. For example, a monitor may detect a fraction x of all faults, if a fraction $y > x$ of all systems is checked. The actual failure rate x/y, may be hard to determine, since it depends on the competence of the monitor, etc. But it is easy to extrapolate that the same monitor would probably detect a fraction of $2x$ of faults if she controlled $2y$ of the joints.

2. Regulators or courts often have a fairly good idea of a reasonable standard for care in terms of observable *resource commitments*, like time spent monitoring, frequency of safety controls, etc. It is much harder to estimate how resource measures translate into failure probabilities. When using ratios, resource commitments can be used to proxy for probability estimates in a simple linear approximation: In our example, the ratio of the mandated frequency to the actual frequency of controls can be used to approximate the failure ratio $\bar{p}_i^k \,/\, p^k(a_i)$.

Once the failure ratios are estimated or approximated, the liability rule is very simple since all payments are linear in the failure ratios. For example, $\frac{\bar{p}_i^k}{p^k(a_j)} = 0.5$ if i's failure probability was twice as high as it was expected to be. Again, the weights α_i are indeterminate, and can be chosen so as to accommodate distributional objectives, or other judicial considerations like the equity principle.

4. Extension to General Patterns of Monitoring Interaction

Until now we assumed that the probability $p^k(a)$ for harm x^k to occur was the risk that none of the agents detected a failure, and hence it could be written as $p^k(a) = \prod_i p_i^k(a_i)$. In this Section, we demonstrate that our liability rule can be suitably extended to the most general patterns of how the monitoring efforts of the agents may interact. That is, we drop the assumption that the joint failure probability $p^k(a)$ can be separated in a multiplicative fashion into the individual failure probabilities $p^k(a)$. We maintain all other assumptions, however, notably the assumption that the problem is well-specified, i.e. $p^k(a_i)x^k + c^t(a_i)$ is strictly concave in a_i, and that the types t of each agent are independently distributed.[5]

We denote the type profile of all agents excluding agent i by $t_{-i} = (t_1, t_2, ..., t_{i-1}, t_{i+1}, ..., t_n)$. Let a_{-i}^t denote the efficient profile of actions for all agents except agent i, if their type profile

[5]An extension to correlated types is possible, though. See also Cremer and McLean (1985) and McAfee and Reny (1992).

is t_{-i}. Let $f(t)$ be the joint probability of agents having the type profile $t = (t_1, t_2, ..., t_n)$, and let $f(t_{-i})$ be the joint probability of all agents except i having the type profile t_{-i}.

Efficiency requires then that each agent i chooses her care level so as to minimize social cost

$$c_i^t(a_i) + \sum_{t_{-i}} \sum_k f(t_{-i})\, p^k(a_{-i}^t, a_i)\, x^k \ ,$$

where$\sum_{t_{-i}} \sum_k f(t_{-i})\, p^k(a_i, a_{-i}^t)\, x^k$ is agent i's expectation of the total harm, based on her own type t and her expectation that all other agents use their efficient actions. The situation is more complicated compared to independent security controls, but the basic principle of our liability rule remains the same. The only difference is that the side payments can no longer be expressed as a function of individual failure risks, since there is nothing like an "individual failure risk" in generalized multicausal damage functions.

To develop the necessary modification for the side payments, define

$$E_i \left[p^k(a) \right] = \sum_{t \in T_i} f_i^t p^k(a_{-i}, a_i^t)$$

as the expected probability for harm x^k when all other agents except agent i choose the *actual* profile a_{-i}, and agent i takes the efficient action a_i^t. In other words, the expectation is taken only with respect to agent i's action; for all other agents, their actual action choices are used. $E_j \left[p^k(a) \right]$ is defined analogously, where expectations are only taken with respect to agent j's action. The efficient liability rule can now be stated as:

Proposition 2. *The following liability rule is efficient for general patterns of interaction:*

$$l_i(x^k, a) = x^k \left[\alpha_i + (1 - \alpha_i) \frac{E_j \left[p^k(a) \right]}{p^k(a)} - (1 - \alpha_h) \frac{E_i \left[p^k(a) \right]}{p^k(a)} \right] \quad \begin{array}{l} \textit{for } h = i - 1 \mod m, \\ \textit{for } j = i + 1 \mod m, \end{array} \quad \forall x^k,$$

where $0 < \alpha_i < 1 \sum_{i=1}^n \alpha_i = 1$.

Proof. *See the Appendix.*

The intuition behind the liability rule is mainly the same as before. Again, the *expectation* of the side payment that each agent i receives (i.e. $x^k\ (1 - \alpha_h) \frac{E_i \left[p^k(a) \right]}{p^k(a)}$) is independent of agent i's behavior: First, $E_i \left[p^k(a) \right]$ is independent of a_i by definition. Second, a_i enters into the expectation over $p^k(a)$ in exactly the same way as into the expectation over x^k, and hence cancels out. Now consider the additional amount that agent i has to pay (i.e. $(1 - \alpha_i) \frac{E_j \left[p^k(a) \right]}{p^k(a)}\ x^k$): Since her actual action choice enters both into $E_j \left[p^k(a) \right]$ and into $p^k(a)$, it cancels out. It follows that her expected payment depends on a_i in the same way as the social cost function, because the expectation over x^k depends on a_i. The main difference to independent security controls is that side payments must now be defined according to the *total* (expected) probability for harm, i.e. over $E_i\left[p^k(a)\right]$ instead of $\overline{p}_k^i$. Nevertheless, the court can still remain completely agnostic about other hypothetical accident scenarios.

5. Conclusion

We have analyzed optimal liability rules for institutions or agents who independently monitor a risky economic activity. The problem is non-trivial since there are multiple agents with private knowledge of their monitoring costs, and since a widely accepted goal in safety regulation is that sanctions should not exceed total harm. Bayesian mechanisms as the theoretical answer to this problem require ex ante estimations of what accident scenarios could have occurred, how much harm they would have caused and how likely they would have been. This complexity explains why mechanism design methods still lack any significant progress as a tool in regulatory or judicial practice.

We propose a simplified version of Bayesian mechanism for an important class of situations, where damage is only caused when all of many layers in a safety net fail jointly. Our rule focuses on the information that is easily available when harm has occurred and when typically regulator and/or courts deliberate over fines and damages.

Our rule is attractive because *(i)* all *hypothetical* scenarios can be fully ignored, and *(ii)* because only estimates of *failure ratios* are required, that is of the shortfall or surplus of an agent's expected care relative to her actual care. The use of ratios offers useful advantages for application, since estimation errors in the *levels* can be avoided and the ratios can be linearly approximated by easily observable *resource commitments*.

We generalize our rule to all possible patterns of interaction in the monitoring efforts of the various agents. The essential remaining assumption is that agents' cost functions are independent. This is not a fundamental obstacle, however, as various solutions for correlated information in the Bayesian mechanism literature show. But adopting these proposals to our rule yields rather unwieldy expressions, and certainly goes at the expense of ease-to-use.

Appendix

Proof of Proposition 1: Proposition 1 is a special case of Proposition 2, hence its proof is implied by the proof of Proposition 2. ■

Proof of Proposition 2. Define

$$\Delta_j(x^k, a) = \frac{E_j\left[p^k(a)\right]}{p^k(a)}, \quad \Delta_i(x^k, a) = \frac{E_i\left[p^k(a)\right]}{p^k(a)}.$$

Hence, we prove the efficiency and balancedness of

$$l_i(x^k, a) = \alpha_i x^k + (1-\alpha_i)\Delta_j(x^k, a)x^k - (1-\alpha_h)\Delta_i(x^k, a)x^k \qquad \forall i. \tag{5.1}$$

We investigate the three parts of the liability rule (5.1) separately and call them l_i^1,l_i^2 and l_i^3, thus, rewriting (5.1) as $l_i(x^k, a) = l_i^1(x^k, a) + l_i^2(x^k, a) + l_i^3(x^k, a)$. The ex ante incentives are determined by the expectation over the damages to be paid. Let $\bar{l}_i$ be agent i's expected value of these payments, and let $\bar{l}_i^1$ be agent i's expected value of the first term l_i^1, etc..

Agent i has to take expectations over all possible realizations of types of the other agents and over all possible damages x^k. In addition to our earlier notation, denote the type profiles excluding i and j by t_{-ij}. Let a_{-ij}^t denote the efficient profile of actions for all agents except agents i and j, if their type profile is t_{-ij}. Denote by $E_{t_{-i}} = \sum_{-i} f(t_{-i})$ the expectation operator over all possible type profiles t_{-i}.

For the first term, $\bar{l}_i^1$, we get after taking probabilities over all scenarios and type profiles

$$\bar{l}_i^1 = \alpha_i E_{t_{-i}}\left[\sum_k p^k(a_{-i}^t, a_i)x^k\right].$$

For the second term, consider first only a single damage x^k and a single type vector t_{-i}. Then l_i^2 leads to:

$$l_i^2(x^k, a_{-i}^t, a_i) = (1-\alpha_i)\Delta_j(x^k, a)\, x^k = (1-\alpha_i)\,\frac{\sum\limits_{t_j\in T_j} f_j^t\; p^k\left(a_{-ij}^t, a_j^t, a_i\right)}{p^k(a_{-i}^t, a_i)}\; x^k.$$

Next, we have to take into account that agent i does not know the profile of types t_{-i} of the other injurers ex ante, and hence she takes expectations:

$$\bar{l}_i^2(x^k) = (1-\alpha_i)\sum_{t_{-i}} f(t_{-i})\;\left[\frac{\sum\limits_{t_j\in T_j} f_j^t\; p^k\left(a_{-ij}^t, a_j^t, a_i\right)}{p^k(a_{-i}^t, a_i)}\right]\; x^k.$$

Moreover, agent i will take expectations over all possible outcomes x^k. Thus:

$$\bar{l}_i^2 = (1-\alpha_i)\sum_k \left\{ \sum_{t_{-i}} f(t_{-i}) \left[\frac{\sum_{t_j \in T_j} f_j^t \; p^k\left(a_{-ij}^t, a_j^t, a_i\right)}{p^k(a_{-i}^t, a_i)} \; p^k(a_{-i}^t, a_i) \right] \right\} \; x^k.$$

The probability $p^k(a_{-i}^t, a_i)$ cancels state by state, hence

$$\begin{aligned}
\bar{l}_i^2 &= (1-\alpha_i)\sum_k \left\{ \sum_{t_{-i}} f(t_{-i}) \left[\sum_{t_j \in T_j} f_j^t \; p^k\left(a_{-ij}^t, a_j^t, a_i\right) \right] \right\} x^k \\
&= (1-\alpha_i)\sum_k \left\{ \sum_{t_{-ij}} f(t_{-ij}) \sum_{t_j \in T_j} f_j^t \left[\sum_{t_j \in T_j} f_j^t \; p^k\left(a_{-ij}^t, a_j^t, a_i\right) \right] \right\} x^k \\
&= (1-\alpha_i)\sum_k \left\{ \sum_{t_{-ij}} f(t_{-ij}) \sum_{t_j \in T_j} f_j^t \, p^k\left(a_{-ij}^t, a_j^t, a_i\right) \right\} x^k \\
&= (1-\alpha_i)\sum_k \left\{ \sum_{t_{-i}} f(t_{-i}) p^k\left(a_{-i}^t, a_i\right) \right\} x^k = (1-\alpha_i) E_{t_{-i}} \left[\sum_k p^k\left(a_{-i}^t, a_i\right) x^k \right] .
\end{aligned}$$

For the third term, the crucial point is that the denominator depends on agent i's behavior, but so does the probability p^k that scenario x^k occurs. Again, consider first only a single damage x^k and a single type vector t_{-i}. Then l_i^3 leads to:

$$\begin{aligned}
l_i^3(x^k, a_{-i}^t, a_i) &= (1-\alpha_h)\Delta_i(x^k, a) \; x^k = \\
&= (1-\alpha_h)\frac{\sum_{t_i \in T_i} f_i^t \; p^k\left(a_{-i}^t, a_i^t\right)}{p^k\left(a_{-i}^t, a_i\right)} \; x^k.
\end{aligned}$$

Next, introduce expectations over t_{-i} and over x^k:

$$\bar{l}_i^3 = (1-\alpha_h)\sum_k \left\{ \sum_{t_{-i}} f(t_{-i}) \left[\frac{\sum_{t_i \in T_i} f_i^t \; p^k\left(a_{-i}^t, a_i^t\right)}{p^k\left(a_{-i}^t, a_i\right)} \; p^k\left(a_{-i}^t, a_i\right) \right] \right\} \; x^k \; ,$$

where $p^k\left(a_{-i}^t, a_i\right)$ again cancels out:

$$\begin{aligned}
\bar{l}_i^3 &= (1-\alpha_h)\sum_k \left\{ \sum_{t_{-i}} f(t_{-i}) \left[\sum_{t_i \in T_i} f_i^t \; p^k\left(a_{-i}^t, a_i^t\right) \right] \right\} x^k \\
&= (1-\alpha_h)\sum_k \left\{ \sum_t f(t) p^k\left(a_{-i}^t, a_i^t\right) x^k \right\} = (1-\alpha_j)\bar{x}.
\end{aligned}$$

$\bar{l}_i^3$ does not depend on the choice or the revealed type of agent i. The reason is of course that we have set up $\Delta_j(x^k, a)$ in such a way as for the probability $p^k(a_{-i}^t, a_i)$ to cancel out in every state.

Putting together the three parts, the expected liability payments of every agent i in equilibrium are:

$$\begin{aligned} \bar{l}_i = \bar{l}_i^1 + \bar{l}_i^2 + \bar{l}_i^3 &= \alpha_i \; E_{t_{-i}}\left[\sum_k p^k(a_{-i}^t, a_i)x^k\right] + (1-\alpha_i)E_{t_{-i}}\left[\sum_k p^k(a_{-i}^t, a_i)x^k\right] \\ &\quad -(1-\alpha_h)\;\bar{x} \\ &= E_{t_{-i}}\left[\sum_k p^k(a_{-i}^t, a_i)x^k\right] - (1-\alpha_h)\;\bar{x}. \end{aligned}$$

Since $(1-\alpha_h)\;\bar{x}$ is independent of i's behavior and since $E_{t_{-i}}\left[\sum_{x^k} p^k(a_{-i}^t, a_i)x^k\right]$ is the total damage expectation, our liability rule is efficient.

To prove balancedness, note that the second and third term in (5.1) are simply passed around in the ring. Thus, they cancel out when the penalties are summed over all m agents, or

$$\sum_{i=1}^m \left(l_i^2(x^k, a) + l_i^3(x^k, a)\right) = 0\;.$$

Hence

$$\sum_{i=1}^m l_i(x^k, a) = \sum_{i=1}^m \alpha_i \; x^k = x^k\;.$$

■

References

ARLEN, JENNIFER H. (1992) "Liability for Physical Injury when Injurers as well as Victims Suffer Losses." *Journal of Law, Economics and Organization*, 8, 410-426.

D'ASPREMONT, CLAUDE AND LOUIS ANDRE GERARD-VARET (1979) "Incentives and Incomplete Information." *Journal of Public Economics*, 11, 25-45.

CREMER, JACQUES AND RICHARD MCLEAN (1985) "Optimal Selling Strategies under Uncertainty for a Discriminating Monopolist When Demands are Interdependent." *Econometrica*, 53, 345-361.

COOTER, ROBERT AND ULEN (1983) "Law and Economics" *Harper Collins Publishers.*

DEMSKI, JOEL S. AND DAVID SAPPINGTON (1984) "Optimal Incentive Contracts with Multiple Agents." *Journal of Economic Theory*, 33, 152-171.

EMONS, WINAND AND JOEL SOBEL (1991) "Liability Rules when Agents are not Identical." *Review of Economic Studies*, 57, 375-90.

FEESS, EBERHARD AND ULRICH HEGE (1999) "The Role of Insurance in the Adjudication of Multiparty Accidents." *International Review of Law and Economics*, 19, 69-86.

FINSINGER, JOERG AND MARK V. PAULY (1990) "The Double Liability Rule." *Geneva Papers on Risk and Insurance*, 15, 159-169.

KORNHAUSER, LEWIS A. AND RICHARD L. REVESZ (1989) "Sharing Damages Among Multiple Tortfeasors." *Yale Law Journal*, 98, 831-884.

KORNHAUSER, LEWIS A. AND RICHARD L. REVESZ (1994) "Multidefendant Settlements: The Impact of Joint and Several Liability." *Journal of Legal Studies*, 19, 617-651.

LANDES, WILLIAM S. AND RICHARD A. POSNER (1987) *The Economic Structure of Tort Law.* Cambridge MA: Harvard University Press.

MCAFEE, R. PRESTON, AND PHILIP J. RENY (1992) "Correlated Information and Mechanism Design." *Econometrica*, 60, 395-421.

POLINSKY, MITCHELL A. AND YEON-KOO CHE (1991) "Decoupling Liability: Optimal Incentives for Care and Litigation." *Rand Journal of Economics*, 22, 562-570

POLINSKY, A. MITCHELL AND STEVEN SHAVELL (1998) "Punitive Damages." *The New Palgrave Dictionary of Economics and the Law.*

SHAVELL, STEVEN (1987) *Economic Analysis of Accident Law.* Cambridge MA: Harvard University Press.