E 20340 F

Journal of Institutional and Theoretical Economics

JITE

Vol. 169, No. 4 December 2013

ARTIBUS INGENUIS J·C·B·M· 1·8·0·1

Mohr Siebeck

Damages Regimes, Precaution Incentives, and the Intensity Principle

by

Urs Schweizer*

Received September 13, 2012; in revised form June 11, 2013; accepted June 18, 2013

This paper revisits the accident model at its roots and shows that the intensity principle provides a powerful analytical tool to handle a variety of issues in a unifying frame and based on common intuition. If courts impose inefficient standards, if a cap on liability exists, or if the principal must pay an information rent to induce precaution, the exact method of quantifying damages matters. The intensity principle allows comparing the intensity of precaution incentives under different damages regimes, such as strict liability, proportional liability, and the negligence rule. Moreover, it requires less restrictive assumptions than the more traditional approach. (JEL: K13, D62)

1 Introduction

For applications of microeconomic theory, assumptions such as differentiability and convexity are commonly imposed to make use of calculus, the implicit-function theorem, and the first-order approach. Yet this comes at a price, as applications do not always fit into a setting based on a continuum of real numbers – or, as Milgrom and Shannon (1994) have phrased it, such assumptions play the role as servants to a method.

The accident model as the workhorse of the economic analysis of tort law is a prime example. Think of an accident due to an excavation pit. Under a negligence regime, courts will check whether or not the injurer has posted sufficient warning signs, whether or not he has illuminated these signs, whether or not he has built a fence around the pit, whether or not he has covered the pit over weekends, and so on. Such cases are easily captured by a finite set of available precaution measures

* University of Bonn, Germany. The author thanks Dominique Demougin, Georg Nöldeke, two referees, and Gerd Muehlheusser, as well as seminar participants in Barcelona, Hamburg, and Zurich, for helpful comments on a previous version of this paper. Support by the DFG through SFB/TR 15 is also gratefully acknowledged.

DOI: 10.1628/093245613X671247 – © 2013 Mohr Siebeck

(including combinations of them), initially without any formal structure. Continuous choice, as required by the fir t-order approach, may be more demanding to justify.

For f nite sets, the method of calculus does not apply, but Milgrom and Shannon (1994) have provided a powerful substitute that is simple to handle. The present paper refers to the substitute technique as the *intensity principle*, because comparing the intensity of precaution incentives under different damages regimes is involved.

To illustrate and to propagate the method, precaution incentives are compared for three damages regimes – strict liability, proportional liability, and the traditional negligence rule – as well as in relation to f rst best. Both the negligence rule and proportional liability are based on due-care standards. While the negligence rule adheres to an all-or-nothing compensation of the victim, proportional liability aims at capturing just that part of the loss that is actually caused by the injurer's deviation from due precaution.

The analysis will be comprehensive in that arbitrary due-care standards are combined with limited liability and wealth constraints. Moreover, the injurer may be an agent who takes precaution on behalf of his principal. Vicarious liability can be viewed as weakening the agent's wealth constraint by combining the agent's assets with those of the principal. But, due to moral hazard or adverse selection, the interaction of principal and injurer under vicarious liability may be distorted. In particular, the principal may have to leave a rent to her agent in excess of the agent's outside option. The present paper covers all these aspects in the unifying frame and intuition of the intensity principle, which rests on the purely ordinal concept of monotonicity. Convexity of rents, as would be required under the f rst-order approach, is not needed by an analysis based on the intensity principle.

New results emerge, but some of my fi dings are also reminiscent of known results. In fact, Kahan's (1989) comparison of the negligence rule with his own rule (which is closely related to proportional liability) is contained as a special case of the present fi dings. Stremitzer and Tabbach (2009) have compared precaution incentives in a setting of limited liability to show that proportional liability outperforms the other damages regimes. The present paper offers a more general version of their f ndings and points out limits of the result. Proportional liability loses its superiority whenever a principal vicariously held liable must pay an information rent to induce precaution by her agent.

The traditional literature on vicarious liability has focused on the effect of increasing the ability to cover damages claims by adding the principal's assets. Important insights have been provided by several contributions. Kornhauser (1982) and Sykes (1984) examine wealth constraints of the agent as the basic condition favoring vicarious liability. Again, f ndings of the present paper based on the intensity principle can be seen as generalizing some of theirs.

Demougin and Fluet (1999) have compared strict liability with the negligence rule for vicarious liability where the principal–agent relationship is plagued by moral hazard and adverse selection. The present paper adds proportional liability. Moreover, in the adverse-selection case, Demougin and Fluet have courts imposing the ex post eff cient precaution expenditures as due-care standards, while the present paper allows for a wider variety of standards, including those that are not type-contingent and thus may come closer to what courts actually do impose.

Under both moral hazard and adverse selection, the principal is conf ned to precaution levels that are implementable by appropriate contracts. The sets of implementable precaution levels are endogenous and may contain holes even if the set of feasible care levels consists of all nonnegative real numbers. Applying calculus to sets with holes becomes tedious. The intensity principle, in contrast, does not face any comparable diff culties.

The paper is organized as follows. In section 2, a version of the accident model is introduced that distinguishes between precaution measures, precaution costs, and the accident probability. Some precaution measures can be ruled out, as they are dominated (from the injurer's perspective) by other measures. Rational injurers will never choose dominated precaution measures. Ruling dominated measures out leaves a set of actions that can be ordered in accordance with the level of precaution spending.

Section 3 compares the intensity of precaution incentives under three damages regimes: the negligence rule, proportional liability, and strict liability. There may be a cap on damages awards, which either reflect legally imposed limited liability or the wealth constraint of the injurer. The intensity principle is established, which shows that one regime provides more intensive precaution incentives than another one if the difference of the injurer's objective functions under the two regimes is monotonically increasing. At a f xed due-care level, proportional liability leads more closely to the f rst-best solution than does the negligence rule.

In substance, the f ndings of this section are known. Establishing them again by making use of the intensity principle serves two purposes. First, it illustrates the intensity principle at work in a surrounding familiar to many readers of this paper. Second, it prepares for the insight gained in later sections that the logic and intuition of the intensity principle easily extends to precaution decisions reached by principal–agent relationships. The f rst-order approach, in contrast, would require new proofs.

The next section, in fact, considers an agent taking precaution measures on behalf of a principal. In case of an accident, the principal may be held vicariously liable. Due to the added wealth of the principal, vicarious liability can be interpreted as raising the cap. In the presence of moral hazard, however, the interaction between principal and agent may be distorted. To induce sufficien precaution, the principal must pay to her agent an information rent increasing with precaution. In the intensity of precaution incentives, not much changes. The information rent simply cancels as far as the difference of the objective function under different regimes is concerned.

The superiority of proportional liability on eff ciency grounds, however, may be lost because the rent, which is of purely redistributive nature, does not affect welfare. Yet, the wider range of precaution measures implementable under the negligence rule allows outperforming proportional liability even if the relationship between principal and agent suffers from moral hazard.

The next section investigates a setting where the principal–agent relationship is governed by adverse selection. Compared with previous sections, quite restrictive assumptions of the single-crossing type are needed to keep the principal–agent problem tractable. Monotonicity of rents, however, emerges directly from these incentive constraints. Convexity of rents, as would be needed under the f rst-order approach, can be dispensed with by relying on the intensity principle instead. Moreover, the same intuition as under moral hazard supports the f ndings, because rents cancel out whenever differences of the principal's objective function under different regimes are under consideration.

Insights from mechanism design are needed for either approach. To set off the issues, I have relegated them to the appendix.

2 *The Accident Model*

The traditional accident model specifie the probability $\varepsilon(x)$ of an accident as a monotonically decreasing function of precaution expenditures x from the set $\mathbb{R}_+$ of nonnegative real numbers. If an accident occurs, the victim suffers a loss of fi ed size L, and, depending on the damages regime in place and possibly the precaution expenditures actually made, the injurer may owe damages to the victim.

Under strict liability S, if an accident has occurred, damages $D_S = L$ are due, no matter what precaution the injurer has actually taken. The other two rules considered by the present paper are based on a due-care standard x^o. Suppose the injurer has actually spent x. Then the negligence rule awards damages $D_N(x, x^o) = L$ equal to the total loss if $x < x^o$ but $D_N(x, x^o) = 0$ otherwise, whereas proportional liability awards

$$D_P(x, x^o) = \frac{\varepsilon(x) - \varepsilon(x^o)}{\varepsilon(x)} \cdot L$$

if the injurer is found negligent ($x < x^o$) and $D_P(x, x^o) = 0$ otherwise. Proportional liability is in the spirit of Kahan's (1989) rule, as it attempts to award only that part of the loss that is caused by the deviation from the standard. Shavell (1985) has investigated proportional liability to exempt the injurer if the loss has actually been caused by nature (but-for test).

In legal practice, courts are unlikely to focus solely on the actual amount x spent on precaution. Think of the example from the introduction, where courts would check whether or not the injurer has posted a warning sign and whether or not he has built a protective fence around the pit. Or think of a car accident. The driver may or may not have obeyed the speed limit. The tires of his car may or may not have been in proper condition. The brakes may or may not have been well maintained, and so on. The set of all conceivable precaution measures (including combinations of them) is denoted by A. This set may be envisaged as a subset of the n-dimensional real vector space, i.e., $A \subset \mathbb{R}^n$, where some dimensions capture binary decisions while others involve continuous choice. It could even be appropriate to think of A as a f nite set without much formal structure.

In any case, precaution measures impose costs to be borne by the injurer. Let $c(a)$ denote the costs (expressed in monetary equivalent) associated with the precaution measure $a \in A$. These costs may ref ect disutility of effort or, alternatively, may correspond to precaution expenditures. The probability $p(a)$ of an accident also depends on the precaution measure a that the injurer has taken ex ante.

In this setting, it might be more appropriate to have courts focusing on the accident probability $p(a)$ than on the costs $c(a)$ actually incurred. A tiny fence offering little protection will not impress courts just because it was made out of unreasonably expensive material. For this reason, I rather specify a due accident probability p^o such that a precaution action a is negligent if it causes an accident with a probability $p(a)$ higher than this standard, i.e., $p(a) > p^o$.

In the same spirit, let me assume that damages awarded would not increase if a precaution measure were chosen with a lower accident probability. It then follows that the damages $d(p(a), p^o) \geq 0$ due in case of an accident are a function of the actual accident probability $p(a)$ and the probability standard p^o. In any case, I assume that damages due in case of an accident are a (weakly or strictly) decreasing function of the accident probability that comes with the chosen precaution measure, i.e.,

$$d(p(a), p^o) \geq d(p(a'), p^o)$$

is assumed to hold for any due accident probability p^o and for any two measures a and a' from A with $p(a) \geq p(a')$. Notice that, as a degenerate case, this condition would also be met in the case of strict liability, where the damages $d(p(a), p^o) = d(p(a'), p^o) = L$ are independent of the true and the due accident probability.

In this setting, it proves useful to introduce the following notion of dominance between precaution measures. Measure a is said to dominate measure a' if one of the following two conditions is met. First, the two measures a and a' induce the same accident probability $p(a) = p(a')$ but a' causes higher costs, i.e., $c(a') > c(a)$. Second, measure a comes with a strictly lower accident probability (i.e., $p(a) < p(a')$) but does not cost more than a' (i.e., $c(a) \leq c(a')$). Let A^u denote the set of undominated measures that are left after all dominated measures have been eliminated.

Notice that this elimination procedure does not depend on the damages regime in place. Yet, as long as damages due in case of an accident are a (weakly or strictly) decreasing function of the accident probability, the rational injurer will not take any precaution measure that is dominated by another one in the above sense.

In fact, the (risk-neutral) injurer maximizes an objective function of the form

$$\phi(a, p^o) = G - c(a) - p(a) \cdot d(p(a), p^o)\,,$$

where G capturers the injurer's private gain from running the activity. Therefore, if measure a dominates a' in the sense of the f rst condition, it follows that

$$\begin{aligned}\phi(a, p^o) &= G - c(a) - p(a) \cdot d(p(a), p^o)\\ &> G - c(a') - p(a) \cdot d(p(a), p^o) = \phi(a', p^o)\,,\end{aligned}$$

and hence the expected payoff is lower under a' than under a. Similarly, if a dominates a' in the sense of the second condition, then at least

$$\phi(a, p^o) \geq \phi(a', p^o)$$

must be true. Notice that this inequality holds in the strict sense again if either $c(a) < c(a')$ or $d(p(a), p^o) < d(p(a'), p^o)$.

In any case, the injurer's payoff under a measure a' that is dominated by a cannot exceed the one under measure a, and hence he at least weakly (if not strictly) prefers measure a over measure a'. Notice that this claim is valid for any legal regime awarding damages that are a (weakly or strictly) decreasing function of the accident probability and for any due accident probability, eff cient or not.

Therefore, if I no longer take dominated precaution measures into account, it is not because some invisible hand has removed them, but because an endogenous party, namely the injurer, never has the incentive to choose them.

The crucial feature of the set A^u of undominated precaution measures rests on the fact that its complete ordering based on the accident probability is equivalent to the one based on precaution expenditures. More precisely, for any two measures a and a', both from the set A^u of undominated precaution measures, $p(a) < p(a')$ holds if and only if $c(a) > c(a')$, and $p(a) = p(a')$ holds if and only if $c(a) = c(a')$. From a formal perspective, specifying a due accident probability becomes equivalent to specifying a due level of spending. But notice: such a claim is valid only after dominated precaution measures have been eliminated.

Let

$$X = \{x \in \mathbb{R}_+ : \exists\, a \in A^u \text{ such that } x = c(a)\}$$

denote the set of precaution spendings arising from undominated precaution measures. Associated with any given level x of spending from X, an accident probability

$$\varepsilon(x) = \min_{a \in A} p(a) \quad \text{subject to} \quad c(a) \leq x$$

can unambiguously be defined Notice that any undominated precaution measure a for which $c(a) = x$ holds solves the above problem, i.e., $\varepsilon(x) = p(a)$.

Moreover, by construction of the set X, the probability $\varepsilon(x)$ is a strongly monotonically decreasing function of precaution spendings x from X.

If the idea that courts focus on precaution measures, not just on spending, is taken seriously, then the set X of levels of spending arising from undominated precaution measures must be derived from the primitives of the model. Since the set A of all precaution measures has little formal structure, the set X of spending levels arising from undominated precaution measures may also have little formal structure except for being a (possibly strict) subset of real numbers. Fortunately, for the method propagated by the present paper, the exact shape of the set X does not matter.

A more subtle point remains to be settled. If courts impose a due accident probability p^o, it cannot be taken for granted that there exists an undominated precaution measure $a^o \in A^u$ that comes with an accident probability $p(a^o)$ exactly equal to the due care probability, i.e., $p(a^o) = p^o$. Yet, to avoid lengthy case distinctions, let me simply assume that courts voluntarily impose due accident probabilities that injurers can exactly meet with some undominated precaution measure. This means that a due-care level $x^o \in X$ is assumed to exist, such that $\varepsilon(x^o) = p^o$ exactly holds.

Except for X possibly not being a continuum, we are back to the setting of the traditional accident model if dominated precaution measures are ruled out. The accident probability emerges as a monotonically decreasing function of precaution costs by construction. Convexity (and differentiability), however, as would be needed for the fir t-order approach, remains more demanding to justify.

3 Comparing Precaution Incentives

In the following, precaution incentives will be compared under strict liability S, the negligence rule N, and proportional liability P as introduced at the beginning of the previous section. The injurer chooses a precaution spending from the set X that captures spending levels arising from undominated precaution measures. The accident probability $\varepsilon(x)$ is a strongly monotonically decreasing function of such spendings.

If the injurer has suffic ent wealth to cover damages claims and if the standard is specifie at its effic ent level, then all three regimes provide eff cient precaution incentives. If, however, standards are ineff cient or if there is a cap $H \leq L$ on liability, the three rules provide precaution incentives of different intensities.

Think of H as a legally imposed cap on liability. Alternatively, the cap may be due to the injurer's wealth constraint. If precaution costs are of the type of disutility of effort, the cap will remain constant. If, however, precaution expenditures have to be covered from the injurer's assets, the cap will decrease with increasing precaution spending (unless an upper bound on liability is set by the legislator, in which case it will remain constant again). To begin with, I assume a constant cap. At the end of the present section, the analysis is extended to nonconstant caps.

Under strict liability, the injurer's objective function is $\phi_S(x) = G - x - \varepsilon(x) \cdot H$, whereas under the other two rules $J \in \{N, P\}$ based on a due-care standard x^o, his objective function is

$$\phi_J(x, x^o) = G - x - \varepsilon(x) \cdot \min[D_J(x, x^o), H] .$$

Notice that, under the negligence rule as well as under proportional liability, the injurer will never choose precaution in excess of the standard, that is, his choice will be from the range $I(x^o) = \{x \in X : x \leq x^o\}$. His optimal choice need not be unique. Let

$$M(J, x^o) = \arg \max_{x \in I(x^o)} \phi_J(x, x^o)$$

denote the set of all precaution costs that maximize the corresponding objective function.

The injurer under strict liability, as well as the social planner with expected welfare $\phi_W(x) = G^s - x - \varepsilon(x) \cdot L$ as objective function, may choose precaution in excess of the standard (G^s denotes the social gain from running the activity, which may differ from the injurer's private gain G). Yet, it proves analytically convenient to consider the set of precaution costs

$$M(J, x^o) = \arg\max_{x \in I(x^o)} \phi_J(x)$$

that maximize these objective functions artificially constrained to the range $I(x^o)$ nonetheless. For regimes $J \in \{S, W\}$, the actual choice would be from $M(J, \infty)$.

If precaution costs x_J and x_K maximizing the objective functions under regimes J and K are generally unique, and if $x_J \leq x_K$ holds, then the regime K provides more intensive precaution incentives than the regime J, in an obvious sense. Given the few assumptions imposed, however, such uniqueness cannot be taken for granted. To continue to rank incentives, I follow Milgrom and Shannon by making use of the strong set order. Accordingly, regime K is said to provide *more intensive precaution incentives* than regime J at standard x^o – for short, $M(J, x^o) \leq_s M(K, x^o)$ – if, for any $x_J \in M(J, x^o)$ and $x_K \in M(K, x^o)$, it follows that

$$x_J \wedge x_K = \min[x_J, x_K] \in M(J, x^o) \quad \text{as well as} \quad x_J \vee x_K = \max[x_J, x_K] \in M(K, x^o)$$

also hold. Notice that the binary operators $\wedge$ and $\vee$ impose a lattice structure on the set X of precaution spendings that are undominated in the sense of the previous section.

Suppose the intensity relation $M(J, x^o) \leq_s M(K, x^o)$ holds and x_J and x_K are maximizers as above. If $x_J \leq x_K$, then, under regime K, more is actually spent on precaution than under regime J. If, however, $x_K < x_J$, then $x_K = x_J \wedge x_K \in M(J, x^o)$ and $x_J = x_J \vee x_K \in M(K, x^o)$ will both hold as well. Therefore, under regime J, the injurer would be equally well off by choosing $x_J \wedge x_K$, and under regime K he would be equally well off by choosing $x_J \vee x_K$. As a consequence, the injurer would raise no objection against a selection of maximizers where he spends more on precaution under regime K than under regime J.

The *intensity principle* as summarized by the following proposition provides a simple criterion for the intensity relationship to hold. The intensity principle is a special case of Theorem 4 in Milgrom and Shannon (1994). Proving the proposition directly, however, turns out to be less demanding than showing in detail why it is a special case of Theorem 4.

PROPOSITION 1 *If the difference of the injurer's objective functions, $\phi_K(x, x^o) - \phi_J(x, x^o)$, is monotonically increasing in x from the range $I(x^o)$, then the intensity relation $M(J, x^o) \leq_s M(K, x^o)$ holds.*

PROOF Suppose $x_J \in M(J, x^o)$ and $x_K \in M(K, x^o)$. By definition, it follows that $x_J \wedge x_K \leq x_J$, and, since the difference of the objective functions is monotonically increasing, it follows that

$$\phi_K(x_J \wedge x_K, x^o) - \phi_J(x_J \wedge x_K, x^o) \leq \phi_K(x_J, x^o) - \phi_J(x_J, x^o)$$

must hold.

Notice further that either $x_J \wedge x_K = x_J$ and $x_J \vee x_K = x_K$ or $x_J \wedge x_K = x_K$ and $x_J \vee x_K = x_J$ must be true. In either case, it follows that

$$\phi_K(x_J \wedge x_K, x^o) - \phi_K(x_J, x^o) = \phi_K(x_K, x^o) - \phi_K(x_J \vee x_K, x^o) \geq 0\,,$$

where the nonnegativity is due to the fact that x_K maximizes ϕ_K.

Combining the above inequalities immediately leads to

$$\phi_J(x_J, x^o) \leq \phi_J(x_J \wedge x_K, x^o) - [\phi_K(x_J \wedge x_K, x^o) - \phi_K(x_J, x^o)]$$

and hence to $\phi_J(x_J, x^o) \leq \phi_J(x_J \wedge x_K, x^o)$. Since x_J maximizes ϕ_J, this inequality cannot hold in the strict sense, and $x_J \wedge x_K$ must maximize ϕ_J as well.

The remaining claim that $x_J \vee x_K$ maximizes ϕ_K follows analogously. *Q.E.D.*

Notice that if one of the objective functions has a unique maximizer or if the difference of the objective functions is strongly monotonically increasing, then $x_J \leq x_K$ necessarily holds for any pair x_J and x_K of optimizers.

For regimes $J \in \{S, N, P, W\}$, it is easy to verify that the following differences of objective functions are monotonically increasing in the range $I(x^o)$:

$$\phi_P(x, x^o) - \phi_S(x) = \max[\varepsilon(x^o) \cdot L - \varepsilon(x) \cdot (L - H), 0],$$

$$\phi_N(x, x^o) - \phi_S(x) = \begin{cases} 0 & \text{if } x < x^o, \\ \varepsilon(x^o) \cdot H & \text{if } x = x^o, \end{cases}$$

$$\phi_W(x) - \phi_S(x) = S - G - \varepsilon(x) \cdot (L - H),$$

$$\phi_W(x) - \phi_P(x, x^o) = S - G + \min[-\varepsilon(x^o) \cdot L, -\varepsilon(x) \cdot (L - H)]\,.$$

The following proposition combines such monotonicity with the intensity principle.

PROPOSITION 2 *For any given standard $x^o \in X$, the following claims are valid:*

(i) $M(S, x^o) \leq_s M(P, x^o) \leq_s M(W, x^o) \leq_s M(W, \infty)$ *and* $M(S, x^o) \leq_s M(N, x^o)$.

(ii) If $x^o \notin M(N, x^o)$ *then* $M(N, x^o) = M(S, x^o)$, *whereas if* $x^o \in M(N, x^o)$ *then* $M(W, x^o) \subset M(P, x^o)$.

PROOF Except for $M(W, x^o) \leq_s M(W, \infty)$, claim (i) follows directly from the intensity principle and the monotonicity of the corresponding differences of objective functions. The relation $M(W, x^o) \leq_s M(W, \infty)$ holds for the following reason. If the standard is high enough, then $M(W, x^o)$ is even a subset of $M(W, \infty)$; otherwise, $x^o < x_W$ holds for all $x_W \in M(W, \infty)$. In both cases, the intensity relation $M(W, x^o) \leq_s M(W, \infty)$ will obviously be met.

As for claim (ii), if $x^o \notin M(N, x^o)$ then $M(N, x^o) = M(S, x^o)$, as the two objective functions are the same in the relevant range, whereas if $x^o \in M(N, x^o)$ then

$$\phi_S(x) \leq \phi_N(x, x^o) \leq \phi_N(x^o, x^o) = \phi_P(x^o, x^o)$$

holds for all $x \in I(x^o)$. Therefore, since

$$\phi_P(x, x^o) = \max[G - S + \varepsilon(x^o) \cdot L + \phi_W(x), \phi_S(x)]$$

holds for $x \in I(x^o)$, any maximizer of $G - G^s + \varepsilon(x^o) \cdot L + \phi_W(x)$ or, what is the same, of $\phi_W(x)$ must also maximize $\phi_P(x, x^o)$ in the range $I(x^o)$. Claim (ii) is established. *Q.E.D.*

While the intensity principle refers to the intensity of incentives, under additional assumptions on the shape of expected welfare as a function of precaution, the above proposition also provides normative insights with respect to the comparison of the negligence rule and proportional liability. Proportional liability always generates (weakly) less intensive precaution incentives than what would be fir t best (confi ed to the range $I(x^o)$). But either the negligence rule generates even less intensive precaution incentives than proportional liability, or else all welfare-maximizing solutions constrained to the interval $I(x^o)$ maximize the injurer's payoff function under proportional liability. Loosely speaking, proportional liability leads closer to the welfare optimum constrained to the range $I(x^o)$ than the negligence rule. If, as traditionally assumed, social welfare is a concave function of the care level (single-peakedness as the purely ordinal counterpart would be enough), then incentives leading closer to the welfare optimum are always welfare-enhancing. Under such circumstances, proportional liability is superior in eff ciency to the negligence rule at any common standard x^o.

A similar result has been established by Stremitzer and Tabbach (2009) for the traditional frame where the set X of precaution measures coincides with the set of nonnegative real numbers and where the probability of an accident is a differentiable and convex function of precaution expenditures. In contrast, the above proof makes use of ordinal properties only.

So far, the cap H on liability has been assumed constant. Under a plausible assumption, all results of the present section can easily be extended to a monotonically decreasing cap $H = H(x)$. The additional assumption can be expressed in terms of the victim's expected payoff function

$$\psi_S(x) = \varepsilon(x) \cdot [H(x) - L]$$

under strict liability, which is assumed (weakly) monotonically increasing with precaution, that is, $\psi_S(x) \leq \psi_S(x')$ holds for $x < x'$. Under this assumption, all the above proofs remain valid. In fact, replacing H by $H(x)$ and $L - H$ by $L - H(x)$ does not affect the monotonicity of differences of objective functions, and hence, by the intensity principle, intensity relations remain the same.

Beard (1990) has pointed out that the precaution incentives may possibly be counterintuitive if precaution expenditures lower the injurer's ability to compensate the victim. Yet, if the victim's payoff under strict liability remains weakly increasing with precaution spending, intensity relations remain qualitatively the same.

4 *Vicarious Liability under Moral Hazard*

From now on, the injurer is an agent who takes precaution on behalf of his principal. To compensate for the wealth constraint of her agent, the principal may be held vicariously liable. Due to the combined wealth of principal and agent, under vicarious liability, the cap rises from H to H^v where $0 \leq H \leq H^v \leq L$. For the rest of the paper, these caps are assumed constant, but see the comment at the end of the previous section on possible extensions to nonconstant caps.

Much of the literature on vicarious liability assumes that the principal has full control over the agent's precaution decision at zero cost. If, however, their relationship operates under moral hazard, the principal may have to control the agent's precaution choice indirectly by offering a suitable bonus contract. The agent receives a f xed payment t plus the bonus $b \geq 0$ if the accident is avoided.

Due to the agent's wealth constraint, the principal possibly has to pay a rent on top of the agent's outside option, as is well known from the principal–agent literature. Not surprisingly, this rent $r(x)$ is a monotonically increasing function of the precaution x that the principal wants to implement. Moreover, depending on the exact structure of the underlying model, only a subset X^B of all levels can possibly be implemented by a bonus contract. Section A.1 of the appendix provides a formal proof of these statements, showing that no further assumptions are needed to establish monotonicity of the rent as a function of precaution spending.

It is a virtue of the intensity principle that, except for the monotonicity of the rent, details do not matter. In fact, since the rent is independent of the damages regime in place, it cancels whenever the difference of the principal's objective function under different regimes is examined to make use of the intensity principle. For that reason, all results of the previous section continue to hold as long as no comparison with f rst best is involved. But comparing with fir t best remains easy as well, because the only difference to the previous section concerns the rent, which is purely redistributive and hence does not affect welfare. Yet, since the rent is monotonically increasing, the intensity principle allows us immediately to rank precaution incentives relative to f rst best.

For the sake of completeness, I present a formal account of these extensions, even though the intuition gained in the previous section remains essentially the same. Suppose regime J is in place at standard x^o. Then the principal's objective function under vicarious liability v amounts to

$$\phi_J^v(x, x^o) = G - u - x - \varepsilon(x) \cdot \min[D_J(x, x^o), H^v] - r(x) .$$

In fact, in order to induce the agent to take the job, the principal must cover the agent's outside option u and, in addition, must compensate the agent for his precaution effort, on top of which comes the rent required to induce x. Damages claims are borne by the principal as well. The principal's choice is conf ned to the set X^B of precautions implementable with a bonus contract.

To avoid tedious subcases, let me assume that courts impose implementable standards only, that is, $x^o \in X^B$. Then, under the negligence rule as well as under proportional liability, the principal will implement precaution from the range $I^B(x^o) = \{x \in X^B : x \leq x^o\}$ only. For regimes $J \in \{N, P\}$, let

$$M^B(J, x^o) = \arg \max_{x \in I^B(x^o)} \phi_J^v(x, x^o)$$

be the set of precaution costs that maximize the corresponding objective function.

The information rent constitutes mere redistribution and, as such, does not affect welfare. For analytical convenience, however, it proves useful to include the additional regime Wr with objective function

$$\phi_{Wr}(x) = G^s - x - \varepsilon(x) \cdot L - r(x) ,$$

where the rent is deducted from welfare. Notice that the difference between welfare and the above objective function,

$$\phi_W(x) - \phi_{Wr}(x) = r(x)\,,$$

is monotonically increasing in the whole range X.

Under regimes $J \in \{S, Wr, W\}$ and for the same reason as in the previous section, consider the set

$$M^B(J, x^o) = \arg\max_{x \in I^B(x^o)} \phi_J^v(x)$$

of precaution costs that maximize the corresponding objective function,[1] but artif - cially conf ned to the range $I^B(x^o)$. In the absence of this constraint, optimal choices would be from $M^B(J, \infty)$.

The following proposition summarizes the adaptation needed to cover vicarious liability under moral hazard.

PROPOSITION 3 *For any given standard $x^o \in X^B$, the following claims are valid:*

(i) $M^B(S, x^o) \leq_s M^B(P, x^o) \leq_s M^B(Wr, x^o) \leq_s M^B(W, x^o) \leq_s M^B(W, \infty)$ *and* $M^v(S, x^o) \leq_s M^v(N, x^o)$.

(ii) If $x^o \notin M^B(N, x^o)$ then $M^B(N, x^o) = M^B(S, x^o)$, whereas if $x^o \in M^B(N, x^o)$ then $M^B(Wr, x^o) \subset M^B(P, x^o)$.

The above proposition compares precaution incentives under the different regimes for a setting of vicarious liability involving moral hazard. The results look very similar to the ones where the injurer was acting on his own. The superiority of proportional liability, however, is no longer supported. In fact, due to the information rent, precaution incentives under proportional liability are generally insuff cient. As a consequence, the potentially more intensive precaution incentives under the negligence rule may lead closer to fir t best. In such cases, the negligence rule may well outperform proportional liability.

More generally, it can be shown that any precaution chosen under strict liability as well as any precaution chosen under proportional liability would also be chosen under the negligence rule, though possibly at a different standard. The following proposition establishes this claim formally.

PROPOSITION 4 *(i) If $x_S \in M^B(S, \infty)$ then $x_S \in M^B(N, x_S)$. (ii) If $x_P \in M^B(P, x^o)$ then $x_P \in M^B(N, x_P)$.*

PROOF (i) If $x_S \in M^B(S, \infty)$ then $x_S \in M^B(S, x_S)$. Moreover, if $x < x_S$ then $\phi_N^v(x, x_S) = \phi_S^v(x) \leq \phi_S^v(x_S) \leq \phi_N^v(x_S, x_S)$ and hence $x_S \in M^B(N, x_S)$. Claim (i) is established.

(ii) If $x_P \in M^B(P, x^o)$ then

$$\phi_S^v(x) \leq \phi_P^v(x, x^o) \leq \phi_P^v(x_P, x^o) \leq \phi_P^v(x_P, x_P) = \phi_N^v(x_P, x_P)$$

for all $x \in I^B(x^o)$ and hence $x_P \in M^B(N, x_P)$, as was to be shown. *Q.E.D.*

---

[1] To simplify notation, I use $\phi_W^v = \phi_W$ and $\phi_{Wr}^v = \phi_{Wr}$ even though these functions are independent of the cap under vicarious liability v.

Notice that a similar proposition could have been established for the setting of the previous section with an injurer acting on his own. In any case, if the standard of care is adapted to the damage regime in an appropriate way, the negligence rule outperforms all the other rules. Only at one and the same standard is it possible that proportional liability outperforms the negligence rule.

5 *Precaution under Adverse Selection*

If the relationship between a principal and an injurer (her agent) is characterized by adverse selection, the principal must be prepared to pay an information rent on top of the agent's outside option as well.

The principal expects the agent she faces to be of type $i = 1, ..., n$ with probability f_i, where $f_1 + \cdots + f_n = 1$. The agent knows his type. If an agent of type i chooses precaution $x \in X$, his effort costs are type-contingent, amounting to $c_i(x)$. The set X is still understood as a (finite subset of real numbers. Yet, since precaution costs are now type-contingent, elements of X will be referred to as precaution levels. Except for effort costs, the parameters of the model do not depend on type.

The principal may now wish to implement some type-contingent precaution prof le $y = (x_1, ..., x_n)$ with the intention that an agent of type i should choose precaution x_i at subjective costs $c_i(x_i)$ and receive payment t_i for it. Given adverse selection, the principal cannot directly identify the type of his agent, but instead must ensure that an agent knowing himself to be of type i has the incentive to choose precaution x_i.

Let $Y = X \times \cdots \times X$ denote the set of all precaution profi es. Not all of them will be implementable. Under the appropriate single-crossing property (for further details, the reader is referred to section A.2 of the appendix), it is the subset Y^m of monotonic prof les $x_1 \leq \cdots \leq x_i \leq \cdots \leq x_n$ that can be implemented, a well-known result from mechanism design.

To implement the prof le $y \in Y^m$, the principal must pay an expected rent $r(y) = \sum_{i=1}^{n} f_i \cdot \rho_i$, where $\rho_i = t_i - c_i(x_i) - u \geq 0$ denotes the agent's rent on top of his outside option u if he is of type i. Moreover, it follows from the single-crossing property (see appendix A.2) that the expected rent is additively separable and can be written as

$$r(y) = \sum_{i=1}^{n} f_i \cdot r_i(x_i)\,,$$

where the ith component is a monotonically increasing function $r_i(x_i)$ of the ith agent's precaution x_i.

Given regime J, the principal's objective function $\Phi_J(y, y^o)$ depends on the precaution profi e y he wishes to implement and on the standards $y^o = (x_1^o, ..., x_n^o)$ of due care, which may or may not be type-contingent. To avoid lengthy subcases, however, I assume that these due-care standards are implementable, i.e., $y^o \in Y^m$. Notice that noncontingent standards would always be implementable.

The objective function $\Phi_J(y, y^o)$ of the principal inherits additive separability from the rent function and can be written as

$$\Phi_J(y, y^o) = \sum_{i=1}^{n} f_i \cdot \phi_{Ji}(x_i, x_i^o) ,$$

where $\phi_{Ji}(x_i, x_i^o)$ is essentially of the same form as in the previous section. In fact,

$$\phi_{Ji}(x_i, x_i^o) = G - u - c_i(x_i) - \varepsilon(x_i) \cdot \min[D_J(x_i, x_i^o), H^v] - r_i(x_i) ,$$

where J denotes any of the regimes S, N, or P. Since the rent constitutes mere redistribution, it does not enter social welfare, i.e., for regime $J = W$, the components of the objective function are

$$\phi_{Wi}(x_i, x_i^o) = G^s - u - c_i(x_i) - \varepsilon(x_i) \cdot L .$$

As in the case of moral hazard, it proves analytically useful again to consider the additional regime $J = Wr$ that deducts rents from welfare, i.e., $\phi_{Wri}(x_i, x_i^o) = \phi_{Wi}(x_i, x_i^o) - r_i(x_i)$.

For the two regimes N and P involving a prof le $y^o = (x_1^o, ..., x_n^o)$ of possibly type-contingent due-care standards, the principal cannot be better off by outperforming any of these standards, i.e., his optimal choice will be from the subset

$$I^m(y^o) = \{y \in Y^m : x_i \leq x_i^o \text{ for all } i\}$$

of all implementable profiles For the other regimes, I impose this constraint under the same premises as in previous sections. Given regime J, profi es from the set

$$M^m(J, y^o) = \arg \max_{y \in I^m(y^o)} \Phi_J(y, y^o)$$

will be implemented that maximize the principal's expected payoff $\Phi_J(y, y^o)$.

To extend the intensity principle, it proves useful to endow the set Y^m of monotonic (i.e., implementable) precaution profile with the following lattice structure. For any two precaution profi es $y = (x_1, ..., x_n)$ and $y' = (x_1', ..., x_n')$, the precaution profi es $y \wedge y'$ and $y \vee y'$ are def ned as those with the ith components

$$(y \wedge y')_i = x_i \wedge x_i' = \min[x_i, x_i'] \quad \text{and} \quad (y \vee y')_i = x_i \vee x_i' = \max[x_i, x_i'] ,$$

respectively. Notice that Y^m is indeed a lattice because, for any two monotonic prof les y and y', both $y \wedge y'$ and $y \vee y'$ will be monotonic as well.

At prof le y^o of standards, regime K is said to provide more intense precaution incentives than regime J if the strong set relationship $M^m(J, y^o) \leq_s M^m(K, y^o)$ is met. The definitio remains the same as in previous sections. It holds if and only if, for any pair $y_J \in M^m(J, y^o)$ and $y_K \in M^m(K, y^o)$ of maximizers, it follows that

$$y_J \wedge y_K \in M^m(J, y^o) \quad \text{and} \quad y_J \vee y_K \in M^m(K, y^o)$$

will hold as well. If this relationship is met, the principal would be willing to implement the precaution profi e $y_J \wedge y_K$ under regime J and prof le under K such that each type of agent would spend (at least weakly) more under K than under J. In this sense, the intensity of precaution incentives increases type by type.

The following proposition adapts the intensity principle to the above setting of adverse selection.

PROPOSITION 5 *If, for all types i of agents, the difference of objective functions $\phi_{Ki}(x_i, x_i^o) - \phi_{Ji}(x_i, x_i^o)$ is a monotonically increasing function of agent i's precaution x_i (in the range $x_i \leq x_i^o$), then the intensity relation $M^m(J, y^o) \leq_s M^m(K, y^o)$ must hold.*

PROOF The proof makes use of the additive separability of the objective functions and parallels the one of Proposition 1.

In fact, suppose $y_J = (x_{J1}, ..., x_{Jn}) \in M^m(J, y^o)$ and $y_K = (x_{K1}, ..., x_{Kn}) \in M^m(K, y^o)$. By def nition, it follows that $x_{Ji} \wedge x_{Ki} \leq x_{Ji}$ holds for all types. Since the differences of the objective functions are monotonically increasing type by type, it follows from additive separability that

$$\Phi_K(y_J \wedge y_K, y^o) - \Phi_J(y_J \wedge y_K, y^o) \leq \Phi_K(y_J, y^o) - \Phi_J(y_J, y^o)$$

must hold.

Notice further that for each type i, either $x_{Ji} \wedge x_{Ki} = x_{Ji}$ and $x_{Ji} \vee x_{Ki} = x_{Ki}$ or $x_{Ji} \wedge x_{Ki} = x_K$ and $x_{Ji} \vee x_{Ki} = x_{Ji}$ must be true. In either case, it follows again from additive separability that

$$\Phi_K(x_J \wedge x_K, x^o) - \Phi_K(x_J, x^o) = \Phi_K(x_K, x^o) - \Phi_K(x_J \vee x_K, x^o) \geq 0\,,$$

where the nonnegativity is due to the fact that x_K maximizes Φ_K.

Combining the above terms immediately leads to

$$\Phi_J(x_J, x^o) \leq \Phi_J(x_J \wedge x_K, x^o) - [\Phi_K(x_J \wedge x_K, x^o) - \Phi_K(x_J, x^o)]$$

and hence to $\Phi_J(x_J, x^o) \leq \Phi_J(x_J \wedge x_K, x^o)$. Since x_J maximizes Φ_J, this inequality cannot hold in the strict sense, and $x_J \wedge x_K$ must maximize Φ_J as well.

The remaining claim that $y_J \vee y_K$ maximizes Φ_K follows analogously. *Q.E.D.*

By comparing the above proof with the one of Proposition 1, it becomes obvious that the extension of the intensity principle to the present setting is based on exactly the same logic and intuition.

The same holds true for the monotonicity of differences. In fact, the differences

$$\phi_{Ki}(x_i, x_i^o) - \phi_{Ji}(x_i, x_i^o)$$

of objective functions remain monotonically increasing for the following pairs (J, K) of regimes:

$$(S, P),\ (P, Wr),\ (Wr, W),\ \text{and}\ (S, N)\,,$$

as follows from the corresponding f ndings of the previous sections. The f nal proposition is an immediate consequence of the intensity principle extended to the setting of adverse selection.

PROPOSITION 6 *For any profil y^o of standards, the following claims are valid:*

$$M^m(S, y^o) \leq_s M^m(P, y^o) \leq_s M^m(Wr, y^o) \leq_s M^m(W, y^o)$$

and

$$M^m(S, y^o) \leq_s M^m(N, y^o)\,.$$

Except for the extension of the intensity principle as established in the previous proposition, no further proof is required to validate the above proposition, and hence no extra intuition is needed to support the result.

Demougin and Fluet have confined their comparison of strict liability versus the negligence rule to the case with the first-best profile as standard. In the second-best world of adverse selection, such standards need not be optimal. The above proposition applies to any monotonic profile of due-care standards, including noncontingent ones.

6 Concluding Remarks

This paper revisits the accident model at its roots. The injurer selects his precaution measure from a possibly finite set of alternatives. By eliminating dominated measures, the remaining set can be endowed with an order relationship. The intensity principle is then used as a unifying method to compare the intensity of precaution incentives under different legal regimes and under a variety of distortions due to inefficient due care levels, limited liability, wealth constraints, and information rents.

The proofs are both simpler and more general, as they do not rely on assumptions such as differentiability and convexity. No doubt, by imposing the more restrictive assumptions required for the first-order approach, similar results could also be established based on calculus. Yet, the results would be less general though harder to prove. The reader may try: take an accident model involving moral hazard and a possibly nonconstant cap on liability that remains insufficient to fully cover damages claims. Comparing the intensity of precaution incentives with calculus as method would certainly be a lengthy and painful exercise. The intensity principle, in contrast, can easily cope with such a setting.

On the legal side, the findings of the present paper are more difficult to digest. The economic analysis of tort law has argued in favor of due-care standards at their efficient level. Whether courts have followed such advice remains a matter of dispute. Suppose courts actually specify due-care standards in accordance with other criteria. In this case, the present paper still provides a positive theory of how the intensity of precaution incentives would be affected by inefficient standards.

Alternatively, suppose courts are actually aiming at raising welfare. If caps and information rents are involved, their task becomes quite complicated. The negligence rule awarding damages on an all-or-nothing basis may well be in line with court practice. The present paper has shown that the negligence rule offers a wide range of possible precaution incentives. Yet, increasing welfare would require some delicate fine tuning. To induce precaution beyond what can be achieved under proportional liability, the due-care standard would have to be raised to the point where it is still kept under the negligence rule. The efficient level (first best) could well be too high. Such fine tuning remains subtle, as the actual wealth constraints as well as the shape of the information rent would have to be taken into account, a difficult task indeed.

From an analytical perspective, however, the intensity principle has proven to be the appropriate method for handling even complicated versions of the accident model. As a matter of fact, the intensity principle provides a unifying setting that may be useful far beyond tort law. Quite generally, examining the monotonicity of the difference of objective functions proves a convenient approach if the intensity of incentives under different legal arrangements or relative to first best is at stake. In Schweizer (2012), I have examined the effects of breach remedies and performance excuses on investment decisions along similar lines. It seems promising to look for still further applications of the intensity principle in future research.

Appendix

A.1 Moral Hazard

At a bonus contract with fixed payment t and bonus $b \geq 0$, the agent's objective function amounts to $t + \psi(x, b)$ with $\psi(x, b) = b \cdot (1 - \varepsilon(x)) - x$ such that he chooses precaution from the set

$$m(b) = \arg\max_{x \in X} t + \psi(x, b) = \arg\max_{x \in X} \psi(x, b)\,,$$

which is independent of the fixed payment t. Precaution $x \in X$ is implementable by a bonus contract (i.e., $x \in X^B$) if and only if there exists a bonus b such that $x \in m(b)$. To be feasible the fixed payment must satisfy two conditions, $t + H \geq 0$ and $t + \psi(x, b) \geq u$. The first condition reflects the wealth constraint of the agent, and the second one is his participation constraint. In fact, $t + \psi(x, b) - u \geq 0$ is the rent he receives on top of his outside option u.

For any given $x \in m(b)$, the principal goes for the lowest t that satisfies both conditions, viz., $t = \max[-H, u - \psi(x, b)]$, which gives rise to rent

$$t + \psi(x, b) - u = \max[\psi(x, b) - u - H, 0]\,.$$

Of all bonuses that implement x (if there are several), the principal offers the lowest one, denoted as $b = \beta(x)$, which can be determined as follows.

The bonus β implements precaution x if and only if, for all $x' < x < x''$, it holds that

$$\psi(x', \beta) \leq \psi(x, \beta) \quad \text{and} \quad \psi(x'', \beta) \leq \psi(x, \beta)\,,$$

which is easily seen to be equivalent to

$$\frac{x - x'}{\varepsilon(x') - \varepsilon(x)} \leq \beta \leq \frac{x'' - x}{\varepsilon(x) - \varepsilon(x'')}\,.$$

Therefore, the lowest bonus $\beta = \beta(x)$ that still implements x (if x is implementable at all) amounts to

$$\beta(x) = \sup_{x' < x} \frac{x - x'}{\varepsilon(x') - \varepsilon(x)}\,.$$

With this bonus, the rent $r(x)$ is at its minimum and amounts to

$$r(x) = \max[\psi(x, \beta(x)) - u, 0]\,.$$

To establish the monotonicity of this minimum rent, consider two precautions $x_1 < x_2$ where x_i is implemented with bonus β_i, i.e., $x_i \in m(\beta_i)$. Hence,

$$\psi(x_2, \beta_1) \leq \psi(x_1, \beta_1) \quad \text{and} \quad \psi(x_1, \beta_2) \leq \psi(x_2, \beta_2)$$

must both hold and are equivalent to

$$\beta_1 \cdot [\varepsilon(x_1) - \varepsilon(x_2)] \leq x_2 - x_1 \leq \beta_2 \cdot [\varepsilon(x_1) - \varepsilon(x_2)]\,,$$

from which

$$0 \leq [\beta_2 - \beta_1] \cdot [\varepsilon(x_1) - \varepsilon(x_2)]\,,$$

and hence $\beta_1 \leq \beta_2$ follows immediately. As a consequence,

$$\max[\psi(x, \beta_1) - u, 0] \leq \max[\psi(x, \beta_2 - u, 0]\,,$$

and hence $r(x_1) \leq r(x_2)$ must hold. This completes the proof that the rent is a monotonically increasing function of precaution.

Readers preferring the traditional f rst-order approach would now have to introduce suff ciently many assumptions on the primitives of the model to ensure that any x is implementable by a bonus contract and that the rent is a differentiable function of precaution. If rigorously done, this is an unpleasant task that, fortunately, can be dispensed with by relying on the intensity principle. On top of that, since no further assumptions are needed, the result is of greater generality.

A.2 Adverse Selection

In the case of adverse selection, the intensity principle has been extended to the situation where information rents are additively separable and where exactly those precaution profi es can be implemented that are monotonic. For these properties to hold, the following three assumptions turn out to be suff cient.

ASSUMPTION A1 *If $x \in X$ then $c_{i+1}(x) \leq c_i(x)$ (for $i = 1, \ldots, n-1$).*

ASSUMPTION A2 *If $x, x' \in X$ and $x < x'$ then $c_{i+1}(x') - c_{i+1}(x) < c_i(x') - c_i(x)$ (for $i = 1, \ldots, n-1$).*

ASSUMPTION A3 *If $x, x' \in X$ and $x < x'$ then $\varepsilon(x) > \varepsilon(x')$ and $c_i(x) < c_i(x')$ (for $i = 1, \ldots, n$).*

Assumptions A1 and A2 – types with a higher index have lower costs and lower marginal costs – correspond to the single-crossing property well known from mechanism design. Assumption A3 is the familiar assumption traditionally imposed on the accident model.

For the reader's convenience, I briefl sketch the argument why the above three assumptions are indeed suff cient. Most of it is well known from mechanism design.

Under the three assumptions imposed, it is suff cient to check the local incentive constraints. Moreover, to minimize the rents, the principal will choose payments such that the *downward* constraints

$$t_{i+1} - c_{i+1}(x_{i+1}) = t_i - c_{i+1}(x_i)$$

are binding, i.e., an agent of type $i + 1$ would just be indifferent between his contract (t_{i+1}, x_{i+1}) and the one chosen by an agent of the next lower type i. Due to the single-crossing property, all the other incentive constraints will then be fulfilled a fortiori.

Participation constraints must be obeyed as well. Again due to the single-crossing property, it is sufficient to ensure participation of the lowest type, from which all the other participation constraints will follow. Moreover, to minimize expected rent payments, the principal offers payments such that the participation constraint of the lowest type is binding, i.e.,

$$\rho_1 = t_1 - c_1(x_1) - u = 0$$

will hold (no rent at the bottom).

For $i = 2, ..., n$, the payments t_i as well as the rents

$$\rho_i = t_i - c_i(x_i) - u$$

that must be granted to an agent of type i can be calculated recursively from the binding downward incentive constraints.

In fact,

$$\rho_2 = t_2 - c_2(x_2) - u = c_1(x_1) - c_2(x_1) \geq 0$$

is a function of x_1 and x_2, and similarly for ρ_i, which is a function of $x_1, ..., x_i$. The expected rent amounts to

$$r(y) = \sum_{i=1}^{n} f_i \cdot \rho_i(x_1, ..., x_i)\,.$$

Since $\rho_i(x_1, ..., x_i)$ is additively separable, by rearranging terms we find functions $r_i(x_i)$ such that

$$r(y) = \sum_{i=1}^{n} f_i \cdot r_i(x_i)$$

holds. Moreover, by making use of the single-crossing property again, it can be shown that the functions $r_i(x_i)$ must in fact be monotonically increasing.

Under the traditional approach, further restrictive assumptions on second-order derivatives of cost functions would be needed to ensure the convexity of the rent functions, which, in turn, would be needed for reliance on first-order conditions. The intensity principle, in contrast, does not need any further assumptions. Not even differentiability is required.

References

Beard, T. R. (1990), "Bankruptcy and Care Choice," *The RAND Journal of Economics*, 21(4), 626–634.

Demougin, D., and C. Fluet (1999), "A Further Justification for the Negligence Rule," *International Review of Law and Economics*, 19(1), 33–45.

Kahan, M. (1989), "Causation and Incentives to Take Care under the Negligence Rule," *The Journal of Legal Studies*, 18(2), 427–447.

Kornhauser, L. A. (1982), "An Economic Analysis of the Choice between Enterprise and Personal Liability for Accidents," *California Law Review*, 70(6), 1345–1392.
Milgrom, P., and C. Shannon (1994), "Monotone Comparative Statics," *Econometrica*, 62(1), 157–180.
Schweizer, U. (2012), "Breach Remedies, Performance Excuses, and Investment Incentives," *The Journal of Law, Economics, & Organization*, 28(4), 717–733.
Shavell, S. (1985), "Uncertainty over Causation and the Determination of Civil Liability," *The Journal of Law & Economics*, 28(3), 587–609.
Stremitzer, A., and A. Tabbach (2009), "Insolvency and Biased Standards – The Case for Proportional Liability," Yale Law School Faculty and Affliate Scholarship Series Paper 24, New Haven (CT).
Sykes, A. O. (1984), "The Economics of Vicarious Liability," *The Yale Law Journal*, 93(7), 1231–1280.

Urs Schweizer
Department of Economics
University of Bonn
Kaiserstraße 1
53113 Bonn
Germany
schweizer@uni-bonn.de

Escalating Penalties for Repeat Offenders: Why are they So Hard to Explain?

by

Thomas J. Miceli*

Received July 12, 2012; in revised form April 1, 2013; accepted April 4, 2013

Escalating penalties for repeat offenders are a pervasive feature of punishment schemes in various contexts, but economic theory has had a hard time rationalizing the practice. This paper reviews the literature on escalating penalties, and then develops a theory based on uncertainty on the part of enforcers about offenders' gains from committing socially undesirable acts. The analysis derives the conditions under which escalating penalties are both optimal (cost-minimizing) and subgame-perfect. It goes on to discuss several extensions and qualifications. (JEL: K14, K42)

1 Introduction

The imposition of increasingly harsh penalties for repeat offenders is a pervasive feature of punishment schemes in a variety of contexts, including criminal sentencing, regulatory enforcement, and penalization of sports infractions, to name just a few examples. As intuitively appealing as such schemes appear, however, it has proven surprisingly difficult to show that they are consistent with an optimal (cost-minimizing) enforcement policy. One difficulty lies in the notion of "efficient offenses," which is based on the idea that optimal punishments should reflect the social cost of an offense and nothing more, thus inducing only those offenders who value the act more than its cost to society to commit it. This approach, which characterizes the standard economic theory of law enforcement since Becker (1968),[1] offers no basis for increasing penalties regardless of how many times an offender commits an illegal act. Repeat crimes are no different (or less desirable) than repeat purchases of a consumer good.

* Professor of Economics, University of Connecticut, Storrs. In writing this paper, I have benefited from conversations with Kathy Segerson and the very helpful comments of three reviewers and Gerd Muehlheusser.

1 See Polinsky and Shavell (2000) for a modern treatment of that model.

Journal of Institutional and Theoretical Economics 169, 587–604 – ISSN 0932-4569
DOI: 10.1628/093245613X671193

This logic suggests that a policy of increasing penalties for repeat offenders is relevant only for those offenses that should be prevented altogether (Posner, 2003, p. 228). But even in this context, if the social harm from an act exceeds the benefit then a policy of setting the initial penalty equal to (or above) the harm should deter all "rational" offenders from *ever* committing the offense. And if some irrational offenders still commit the act, there is no social gain from threatening to punish them more harshly in the future (or at all!). One possible explanation is that harsher penalties are aimed at incapacitating recidivists,[2] which could be an effective way of preventing harm caused by undeterrable offenders, but this logic does not explain increasing f nes, and more generally, it does not provide a very satisfying explanation for the common-sense appeal of escalating penalty schemes. Another, more likely obstacle to setting the sanction high enough to deter all offenses is that there is a maximum feasible sanction (e.g., life imprisonment, or, in the case of a fi e, the offender's wealth).

Given the above arguments, this paper offers an explanation for escalating penalties based on the following logic. Suppose an enforcer wishes to deter a particular undesirable act but is unsure about what level of punishment it will take because offenders vary in their gains from committing the act, with the highest gain being larger than the maximal sanction. One approach would be to set a high initial punishment to deter as many crimes as possible, but this policy would be costly to implement because some offenders cannot be deterred. A cheaper strategy might therefore be to set a low initial punishment, and then to raise it only for those who commit the act, thereby revealing their higher valuation. Under this escalating scheme, some early crime is tolerated in order to save on initial punishment costs. I will show that the optimality of this latter strategy turns on the following factors: (1) the social undesirability of the act in question, (2) the costliness of punishment, and (3) the existence of a suff ciently large number of undeterrable offenders. Before proceeding with the analysis, I will review the existing literature on escalating penalties, which itself has recently been escalating.

2 *Literature Review*

As noted, the existing literature has had mixed success in explaining escalating penalties. Most notably, models of criminal enforcement based on the standard economic approach, extended to allow offenders to commit crimes over multiple periods, generally f nd that the optimal penalty structure is either f at or declining (Dana, 2001). For example, in a model where offender gains are counted in social welfare and punishment is by a f ne, Polinsky and Shavell (1998) show that although fir t-time offenders face less severe fi es than repeat offenders in the second of two periods, repeaters face the same (maximal) f ne in both periods. Thus, f nes never actually escalate for a given offender. In a model where criminal acts are strictly undesirable, Emons (2003, 2004) shows that the optimal f nes are actually

[2] For economic analyses of incapacitation, see Shavell (1987) and Miceli (2010).

declining over a two-period time horizon. Burnovski and Safra (1994) reach a similar conclusion in a setting where offenders can choose (in advance) to commit repeat crimes over any f nite number of periods. Although Rubinstein (1980) is able to show that there exists an offender utility function that makes an escalating penalty scheme optimal within the context of Becker's model, this special case can hardly account for the pervasiveness of the practice.

Evidently, some departure from the standard framework is necessary to show that escalating punishments are optimal. One possibility, f rst noted by Stigler (1970, pp. 528f.), is that f rst-time offenders may have committed their crimes "accidentally." Following this suggestion, several authors have shown that this situation may indeed lead to an increasing penalty structure (Rubinstein, 1979, and Emons, 2007). Intuitively, because repeat offenders are more likely to have committed their acts deliberately, they need to be punished more severely to be deterred. In a related analysis, Chu, Hu, and Huang (2000) show that escalating penalties are optimal when the legal system sometimes convicts innocent defendants.

Another approach, f rst proposed by Polinsky and Rubinfeld (1991), is based on the idea that offenders differ in their propensities to commit socially undesirable acts. An escalating f ne may be useful in this context as a sorting device, aimed at differentially deterring those offenders with high offense propensities. In a similar vein, McCannon (2009) shows that if some offenders are "experimenters" whose acts may turn out to be socially desirable, while others are habitual offenders whose acts are def nitely undesirable, then a rising penalty scheme can again serve as a screening device to punish the latter type of offenders more harshly.

Several authors have argued that an escalating penalty scheme is necessary to offset the learning-by-doing effect of repeat crime, which raises the cost of apprehending experienced offenders (Baik and Kim, 2001; Posner, 2003, p. 229). While this may justify increasing sanctions, an offsetting effect is that enforcers will likely pursue repeat offenders more vigorously, and if this latter effect dominates, it will make declining punishments optimal (Dana, 2001; Mungan, 2010). Garoupa and Jellal (2004) note, however, that enforcers can also learn from experience – for example, about more eff cient apprehension technologies or the tendencies of offenders. When this is true, low initial penalties may be optimal in order to allow enforcers to learn from early crimes.

A f nal strand of literature has focused on the stigma effect of criminal conviction, which acts as a supplement to formal criminal penalties in deterring some offenders (Rasmusen, 1996; Funk, 2004; Miceli and Bucci, 2005; Dana, 2001, pp. 772–776). Although stigma may help to deter some would-be offenders from committing crimes in the f rst place (thereby lowering the need for high formal sanctions), it also reduces the deterrent effect of criminal punishment for repeat offenders because, for example, their legal employment opportunities are diminished. Thus, higher actual penalties are necessary to deter them. A related explanation for escalating penalties is based on the notion that some (if not most) people obey the law not because of some cost–benefi calculation, but because it is the right thing to do. In this view, legal sanctions serve an "expressive" function by def ning what kinds of conduct are right and what kinds are wrong. An escalating scheme may be especially useful for this purpose, and may therefore induce compliance with the law at a lower cost (Dana, 2001, pp. 776–783).

As is evident from the foregoing survey, several arguments can be used to justify escalating penalties. The fact that none seems entirely satisfactory on its own suggests that some basic insight is missing. The following analysis seeks to further close the gap between theory and practice.

3 The Model

This section develops the simplest possible model of repeat offenders in an effort to derive the minimal conditions under which escalating penalties are consistent with an optimal law enforcement strategy.

3.1 Assumptions

The following assumptions describe the structure of the model. (The impact of relaxing several of these assumptions will be examined in section 4 below.)

ASSUMPTION 1 *There are three types of offenders, who differ according to their gains from committing an offense:* g_1*,* g_2*, and* g_3*, where* $0 < g_1 < g_2 < g_3$*. Let* α *be the proportion of* g_1*'s in the population,* β *the fraction of* g_2*'s, and* $1 - \alpha - \beta$ *the fraction of* g_3*'s.*

ASSUMPTION 2 *An offense imposes social harm of* h*, and is considered socially undesirable in the sense that it should ideally be completely deterred. In other words, the gain to offenders is not counted in social welfare.*

This is contrary to most versions of the standard economic model of crime,[3] but is probably descriptive of many criminal acts. As noted, this assumption seems to be crucial for showing the optimality of escalating penalties, since if offenders' gains counted, then repeat offenses, if "eff cient" in the sense that the gain to the offenders exceeded the social harm, would not be undesirable.

ASSUMPTION 3 *To conserve on notation, the probability of apprehension and conviction,* p*, is f xed and set at one.*[4]

ASSUMPTION 4 *Punishments are costly to impose.*[5] *This is obviously true of imprisonment, but may also be true of fin s, either because there could be administrative*

[3] There are, however, competing views of whether the gain to offenders should be counted. See, for example, Stigler (1970, p. 527), Lewin and Trumbull (1990), and Friedman (2000, p. 230).

[4] This follows the approach in Polinsky and Rubinfeld (1991).

[5] See, for example, Kaplow (1990), who studies the optimal use of nonmonetary sanctions for single (one-time) offenses when p is fi ed.

costs of collecting fin s, or because society simply has an aversion to imposing criminal sanctions of any sort beyond what is deemed "appropriate" (or proportional) to the offense in question, or is minimally necessary to deter the offense. (This aversion could be due to the stigma associated with criminal conviction, regardless of the nature of the sanction.) Let the cost of a sanction to an offender be s and the cost of imposing that sanction to society be cs, where c is the (fi ed) marginal cost.

The costliness of punishment is key for the model because, given the existence of undeterrable offenders, it implies that high initial penalties aimed at deterring f rst-timers are costly. The assumption of a fi ed marginal punishment cost is standard in the literature, though one can easily imagine the case of increasing marginal costs and/or the existence of f xed costs of punishment. I leave the analysis of these cases to future work, but conjecture here that a f xed punishment cost would bias the results *away* from escalating penalties because, as will be shown, the advantage of low initial penalties in the current model is the savings in up-front punishment cost, an advantage that would be greatly eroded by high f xed costs.

ASSUMPTION 5 *When indifferent between committing an offense and not, an offender will not commit it. Thus, $s \geq g_i$ will deter an offender of type i. The maximum possible sanction in a given period is $\bar{s}$, where $g_3 > \bar{s} > g_2 > g_1$. This assumption effectively makes g_3's undeterrable.*[6] *Let s_1 denote the sanction for a f rst-time offender (whether in period 1 or 2), and s_2 the sanction for a repeat offender. The only a priori restrictions on these sanctions are that $s_j \geq 0$ and $s_j \leq \bar{s}$ ($j = 1, 2$).*

ASSUMPTION 6 *Offenders live for two periods and can commit at most one criminal act in each period. They are forward-looking and thus commit acts in period 1 based on the expected gain over both periods. For example, even if $g_i \leq s_1$ for an offender of type g_i in period 1 (a f rst-timer), he will commit the act if the sanction for a repeat offender in period 2 is such that $2g_i > s_1 + s_2$.*[7] *Note that an offender can be a fi st-timer in period 1 or 2, but he can only be a repeat offender in period 2.*

ASSUMPTION 7 *An optimal enforcement strategy will be chosen to minimize the expected present value of social harm from an offense plus punishment costs over the two-period time horizon. In addition, the strategy must satisfy subgame perfection. In other words, noncredible penalty schemes will not be allowed (Emons, 2004). Thus, the enforcer will use backward induction based on the observed behavior of offenders in period 1.*

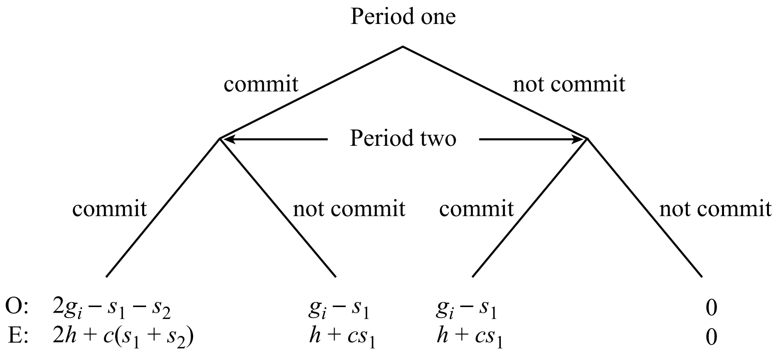
[6] This is not an overly restrictive assumption. In standard models with a continuum of types, it is usually assumed that g is distributed on $[0, \infty)$, which likewise implies that some offenders are effectively undeterrable.

[7] For simplicity I ignore discounting.

3.2 Analysis of the Model

The decision tree for an offender of type g_i is shown in Figure 1. At each of the four terminal nodes, the payoffs for the offender (O) and the enforcer (E) are shown. Consider f rst the optimal strategy of offenders.

Figure 1
Decision Tree for Offender g_i over the Two-Period Time Horizon



Using backward induction, we begin with the two subgames corresponding to each of the offender's period 1 strategies ("commit" or "not commit"). If the offender committed an offense in period 1, for which he incurred punishment of s_1, he will commit a second offense in period 2 if and only if $2g_i - s_1 - s_2 > g_i - s_1$, or if and only if $g_i > s_2$. Likewise, if he did not commit a period 1 offense, he will commit a f rst offense in period 2 if and only if $g_i > s_1$.

Moving back to period 1, we need to consider four cases depending on the magnitudes of s_1 and s_2 relative to offender i's gain from committing an offense:

Case 1: $s_1, s_2 < g_i$. In this case, an offender of type g_i will commit a period 2 offense regardless of his prior history. Given this, he will commit a period 1 offense if and only if $2g_i - s_1 - s_2 > g_i - s_1$, which holds because $s_2 < g_i$. Thus, he will be a repeat offender.

Case 2: $s_1 < g_i \leq s_2$. In this case, offender g_i will commit a period 2 offense if and only if he *did not* commit a period 1 offense. He will thus be indifferent between committing an offense in period 1 only and in period 2 only. We will therefore assume in this case (without loss of generality) that he commits his single crime in period 1.

Case 3: $s_2 < g_i \leq s_1$. In this case, offender g_i will commit a period 2 offense if and only if he committed a period 1 offense. He will therefore be deterred from committing an offense in period 1 (and hence from committing *any* offenses) if $2g_i - s_1 - s_2 \leq 0$, that is, if

$$g_i - s_2 \leq s_1 - g_i\,, \tag{1}$$

whereas if the opposite holds, he will commit offenses in both periods.

Case 4: $g_i \leq s_1, s_2$. In this case, offender g_i will never commit a period 2 offense, nor will he commit a period 1 offense.[8]

This completes the description of the rational behavior of offenders. We now turn to the optimal strategy of the enforcer in light of these cases. As with the offender, we use backward induction to ensure subgame perfection. Observe f rst that the optimal sanction for a period 2 offender, whether a repeater (s_2) or a fir t-timer (s_1), will equal 0, g_1, or g_2. To prove this, note that $s_i > g_2$ can never be optimal, because it contributes nothing to deterrence but involves additional costs. By similar logic, neither $g_1 < s_i < g_2$ nor $0 < s_i < g_1$ can be optimal. The optimal choice among these three options depends on the behavior of offenders in period 1. There are three possible scenarios.

Scenario 1: No Offenders were Deterred in Period 1. In this case, all types are potential repeat offenders, and anyone who commits a period 2 offense will be subject to s_2. If the enforcer sets $s_2 = 0$, all offenders will repeat their crimes, and total period 2 costs will be $TC_2^0 = h$. Alternatively, if the enforcer sets $s_2 = g_1$, only g_2's and g_3's will commit crimes in period 2, and period 2 costs will be $TC_2^1 = (1-\alpha)(h + cg_1)$. Finally, if the enforcer sets $s_2 = g_2$, only g_3's will commit repeat offenses, and period 2 costs will be $TC_2^2 = (1-\alpha-\beta)(h + cg_2)$. In this scenario, we will rule out $s_2 = 0$ as uninteresting because it involves no deterrence in either period. Thus, the choice is between $s_2 = g_1$ and g_2.[9] The cost-minimizing choice is g_2 if and only if $TC_2^2 < TC_2^1$, or if and only if

$$(1-\alpha)c(g_2 - g_1) < \beta(h + cg_2)\,. \tag{2}$$

Scenario 2: Only g_2's and g_3's Committed Period 1 Crimes, while g_1's were Deterred. If the enforcer sets s_2 equal to either 0 or g_1 in this case, both types will commit the offense again, yielding period 2 costs of $TC_2^0 = (1-\alpha)h$ and $TC_2^1 = (1-\alpha)(h + cg_1)$, respectively. Clearly, $s_2 = 0$ strictly dominates $s_2 = g_1$ in this case because it results in the same level of crime but lower punishment costs in period 2, so the latter choice cannot be subgame-perfect. Finally, if the enforcer sets $s_2 = g_2$, only g_3's will commit a further crime. The expected period 2 cost of this option is

$$TC_2^2 = (1-\alpha-\beta)(h + cg_2)\,.$$

---

[8] Note that Cases 2, 3, and 4 can never apply to type g_3 offenders, given the assumption that $g_3 > \bar{s} \geq s_j$, $j = 1, 2$.

[9] The assumption that $s_2 = 0$ is not optimal in this case therefore amounts to assuming that $h > \min\{(1-\alpha)(h + cg_1), (1-\alpha-\beta)(h + cg_2)\}$.

The enforcer will prefer $s_2 = g_2$ over $s_2 = 0$ if and only if $TC_2^2 < TC_2^0$, or if and only if

$$(3) \qquad (1-\alpha)cg_2 < \beta(h + cg_2)\,.$$

Scenario 3: Only g_3's Committed Period 1 Crimes. In this scenario, s_2 will have no deterrence effect, since it cannot be set high enough to deter g_3's given $\bar{s} < g_3$. Thus, subgame perfection again requires that $s_2 = 0$. This completes the analysis of the possible period 2 scenarios.

We now move back to period 1 to derive the optimal choice of s_1, conditional on the preceding results. In the fir t scenario, we conjectured that all offenders committed an offense in period 1, and then showed that $s_2 = g_2$ is optimal if (2) holds, and $s_2 = g_1$ if it does not. (Recall that we ruled out $s_2 = 0$ in this scenario.) In either case, $s_1 < g_1$ is consistent with all types committing a crime in period 1, but $s_1 \geq g_1$ is not, for in that case g_1's will be deterred in period 1.[10] Given this, suppose f rst that the condition (2) holds, so that $s_2 = g_2$ is cost-minimizing. This corresponds to Case 2 for both g_1 and g_2, neither of whom therefore commits a period 2 offense. Given that only g_3's will be repeat offenders, the period 1 problem for the enforcer is to

$$\min_{s_1 < g_1} (h + cs_1) + (1-\alpha-\beta)(h + cg_2)\,,$$

which has as its solution $s_1 = 0$. The optimal (s_1, s_2) pair in this case is thus $(0, g_2)$, and the resulting expression for social costs is[11]

$$(4) \qquad TC_1^* = h + (1-\alpha-\beta)(h + cg_2)\,.$$

Now suppose that (2) does not hold, so that $s_2 = g_1$. This corresponds to Case 2 for g_1's, who therefore do not commit period 2 offenses, and Case 1 for g_2's, who do. Thus, given that g_2's and g_3's will be repeat offenders, the enforcer's period 1 problem is to

$$\min_{s_1 < g_1} (h + cs_1) + (1-\alpha)(h + cg_1)\,,$$

which again has as its solution $s_1 = 0$. The optimal (s_1, s_2) pair in this case is $(0, g_1)$, and the resulting expression for social costs is

$$(5) \qquad TC_2^* = h + (1-\alpha)(h + cg_1)\,.$$

(Comparing (4) and (5) verif es that the cost-minimizing choice indeed depends on (2).) Note that both of these schemes involve an escalating penalty for repeat offenders, with a zero penalty for f rst-timers and a positive penalty for repeaters.

10 Note that because g_1's know they will be deterred in period 2 under both possible choices of s_2, there is no future gain to justify committing a crime in period 1 that yields no immediate net benefit

11 In this and all subsequent cost expressions, we omit the fi ed cost of apprehension.

In the second scenario above, g_2's and g_3's committed period 1 crimes but g_1's were deterred. This requires $g_2 > s_1 \geq g_1$, for otherwise, g_1's will also commit a crime in period 1 regardless of what s_2 is (i.e., even if they have no intention of committing a second offense). Given this situation in period 1, we showed that if the condition (3) holds, it is optimal to set $s_2 = g_2$ in order to deter g_2's from becoming repeat offenders. Thus, Case 4 is relevant for g_1's (total deterrence), and Case 2 is relevant for g_2's (commit an offense in period 1 only). The optimal choice of s_1 therefore solves

$$\min_{g_1 \leq s_1 < g_1} (1-\alpha)(h + cs_1) + (1-\alpha-\beta)(h + cg_2)\,,$$

which has as its solution $s_1 = g_1$. Thus, the optimal penalty structure is (g_1, g_2), and the resulting total cost is

$$TC_3^* = (1-\alpha)(h + cg_1) + (1-\alpha-\beta)(h + cg_2)\,. \tag{6}$$

Note that this is also an escalating scheme with a positive (but low) initial penalty.

In this same scenario, when (3) does not hold, $s_2 = 0$. Thus, Case 3 is relevant for g_1's, so there is a risk of them committing an offense in period 1, even with $s_1 \geq g_1$, in order to essentially have a "free crime" in period 2 as a repeat offender. To prevent this from happening – that is, to completely deter type g_1's – the condition (1) must hold for them. But since $s_2 = 0$ in this scenario, (1) reduces to $s_1 \geq 2g_1$. As for g_2's, the opposite of (1) must hold in order for them to be *willing* to commit period 1 crimes given $s_2 = 0$; that is, $s_1 < 2g_2$. The optimal choice of s_1 therefore solves

$$\min_{2g_1 \leq s_1 < 2g_2} (1-\alpha)(h + cs_1) + (1-\alpha)h\,,$$

which has as its solution $s_1 = 2g_1$. The resulting penalty structure is $(2g_1, 0)$, which is a declining scheme, and the resulting total cost is

$$TC_4^* = (1-\alpha)(h + c2g_1) + (1-\alpha)h\,. \tag{7}$$

The final scenario involved only g_3's committing period 1 offenses, which implied that $s_2 = 0$ is the subgame-perfect solution. In this case, s_1 must be set to deter both g_1's and g_2's in period 1. Case 3 is therefore relevant for both types, which implies that $s_1 \geq 2g_2$. The corresponding first-period problem is to

$$\min_{s_1 \geq 2g_2} (1-\alpha-\beta)(h + cs_1) + (1-\alpha-\beta)h\,,$$

which has as its solution $s_1 = 2g_2$. The resulting penalty scheme is $(2g_2, 0)$, which is again a declining scheme, and the associated total-cost expression is

$$TC_5^* = (1-\alpha-\beta)(h + c2g_2) + (1-\alpha-\beta)h\,. \tag{8}$$

This exhausts the description of the penalty schemes that are potentially optimal in the model. The table summarizes the possibilities: note that three are escalating and two are declining.[12] The next task is to compare the corresponding total-cost

[12] Notably, a flat penalty scheme is never optimal in the current model (save the uninteresting scheme (0, 0), which is ruled out by assumption).

Table
Candidates for an Optimal Penalty Scheme

Scheme	(s_1, s_2) pair	Total cost	Type
1	$(0, g_2)$	$h + (1 - \alpha - \beta)(h + cg_2)$	escalating
2	$(0, g_1)$	$h + (1 - \alpha)(h + cg_1)$	escalating
3	(g_1, g_2)	$(1 - \alpha)(h + cg_1) + (1 - \alpha - \beta)(h + cg_2)$	escalating
4	$(2g_1, 0)$	$(1 - \alpha)(h + c2g_1) + (1 - \alpha)h$	declining
5	$(2g_2, 0)$	$(1 - \alpha - \beta)(h + c2g_2) + (1 - \alpha - \beta)h$	declining

expressions, given by (4), (5), (6), (7), and (8), to determine the conditions under which each is optimal.

The first step is to show that scheme 3 is dominated by the lower-cost of schemes 4 and 5. Comparing costs under 4 and 5 shows that 5 has lower costs when (2) holds and 4 has lower costs when the reverse holds. Thus, supposing first that (2) holds, we compare 5 and 3, and find that 5 involves lower costs. Next, supposing that the reverse of (2) holds, we compare 4 and 3, and find that 4 involves lower costs. These conclusions rule out 3 as an optimal scheme. To understand why this is true, note that scheme 3 is a combination of schemes 4 and 5 in terms of its implications for the behavior of g_2 types (g_1's and g_3's behave the same under all three schemes). Specifically, g_2's commit crimes in both periods under scheme 4, and are deterred in both periods under scheme 5, but they commit crimes in period 1 and are deterred in period 2 under scheme 3. Given the linearity of costs, it turns out that either complete or no deterrence of this offender type is cost-minimizing, depending on whether or not the condition (2) holds. Thus, the combination of the two strategies can never be optimal.

It is not possible to rule out any of the other schemes, which implies that each may be optimal under different conditions. We have already seen that the condition (2) determines whether 4 or 5 is optimal among the two declining schemes. It turns out that (2) also determines which of the two remaining escalating schemes is preferred: scheme 1 involves lower costs when (2) holds, and scheme 2 involves lower costs when the reverse of (2) holds. Thus, supposing first that (2) holds, we compare schemes 1 and 5, and find that scheme 1 is preferred when $TC_1^* < TC_5^*$, or when

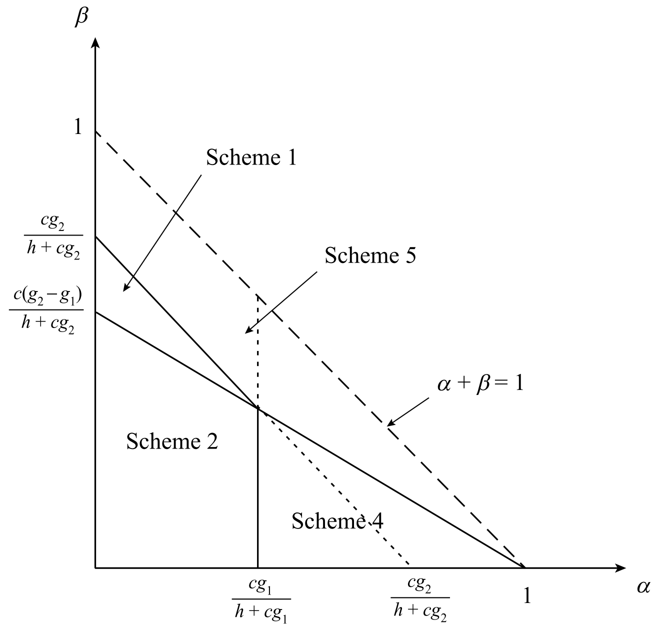
(9) $$(\alpha + \beta)h < [1 - (\alpha + \beta)]cg_2 .$$

Conversely, when (2) does not hold, we compare schemes 2 and 4 and find that scheme 2 is preferred when $TC_2^* < TC_4^*$, or when

(10) $$\alpha h < (1 - \alpha)cg_1 .$$

The globally optimal outcome can now be determined by using the conditions (2), (9), and (10). Figure 2 graphically depicts the ranges over which the four schemes are optimal for different combinations of the parameters α and β, subject to the constraint that $\alpha + \beta \leq 1$. (Recall that α is the proportion of g_1's and β is the proportion of g_2's – the "deterrable" offender types – in the population of all

Figure 2
Regions where the Various Punishment Schemes Are Optimal



offenders.) The fully solid, negatively sloped diagonal line is the locus of points for which the condition (2) holds as an equality; for points to the northeast of this line (and below the dashed line), (2) holds, and thus scheme 1 or 5 may be optimal, while for points to the southwest of this line, the reverse of (2) holds, and scheme 2 or 4 may be optimal.

The ranges are further partitioned by using the conditions (9) and (10). Graphing the locus defined by (9) written as an equality yields the half solid, half dashed diagonal line. As shown, it has a higher intercept and is steeper than the fully solid line, and the two lines intersect at $\alpha = cg_1/(h + cg_1)$. The only portion of this line that is relevant, however, is the segment in the range where (2) holds, and so that portion of the line is drawn solid. In the triangle bounded by the two solid diagonal lines, the conditions (2) and (9) simultaneously hold, so scheme 1 is optimal in this region. In the remaining region above the two diagonal lines, (2) holds but (9) does not, so scheme 5 is optimal.

Now consider the condition (10). Note that it defines a vertical line at $\alpha = cg_1/(h + cg_1)$, and so it intersects the solid line at the same point as the intersection of the two diagonal lines defined above. The relevant portion of this vertical line, however, is the solid segment below the fully solid diagonal line in the region where schemes 2 and 4 may be optimal. To the left of this vertical segment, scheme 2 is preferred, while to the right, scheme 4 is preferred.

We have now characterized the optimal penalty scheme over the entire parameter space. The results show that the two escalating schemes (schemes 1 and 2) are optimal if the deterrable offenders (types g_1 and g_2) do not comprise too large a percentage of the population of all offenders. The next section gives the intuition for this conclusion.

3.3 Discussion

Recall that the basic trade-off in the model is between the benefi s of deterrence (the prevented social harm from crime) and the costs of imposing punishment. The declining schemes, because they deter one or both of the deterrable offender types over both periods by setting high initial penalties, therefore become more desirable when the gains from deterrence increase, which will happen when the fraction of deterrable offenders in the population increases. In the limit where all offenders are deterrable, these schemes will theoretically be able to deter all crimes, and hence will involve no costs. This shows the importance of the assumption that some offenders are undeterrable, which rules out this f rst-best, zero-cost outcome. The advantage of the escalating schemes, in contrast, is the savings in initial punishment costs from setting a low (zero) penalty for f rst-time offenders, and then raising the penalty for repeat offenders so as to selectively deter one or both of the deterrable types from committing further offenses in period 2. In other words, the escalating schemes balance the benefi s of some deterrence against low initial punishment costs. These schemes will therefore be more desirable as the fraction of undeterrable offenders in the population of all offenders becomes high. (Recall that we have ruled out schemes that deter no crimes over both periods but avoid all punishment costs.)

This argument suggests that the use of escalating penalties is, in a sense, a concession to the ineffectiveness of criminal punishment.[13] Although this may seem like an odd justif cation, on ref ection the idea of low penalties for f rst-time offenders is a perfectly rational response to the possibility that some crime is unavoidable, whether because many f rst-time offenders are inadvertent, experimenting, or truly undeterrable. Given this observation, it makes sense not to punish them too harshly, to save on punishment costs. Thus, the conclusion follows logically from the model. The question of whether it comports with the general appeal of escalating penalty schemes, however, is another matter.

Because the advantage of escalating penalties comes from the costliness of punishment, one might suppose that the current explanation applies only when punishment takes the form of imprisonment. It is worth noting, however, that society's dislike of high penalties could ref ect factors other than (or in addition to) monetary costs. One reason is an aversion to disproportionate penalties. As Hart (1982, Ch. VII) notes, actual punishment schemes embody a vestige of retributive motives, reflectin a de-

[13] As one reviewer noted, it suggests that a society populated with hardcore criminals should use an escalating scheme, but one populated with law-abiding citizens should not.

sire for proportionality (or “just deserts”) in the setting of criminal penalties. By this logic, harsh punishment schemes aimed primarily at deterrence, whether in the form of prison or fines, may be viewed as noncredible by offenders, especially given the sequential nature of criminal procedure, which leaves considerable discretion in the hands of prosecutors, judges, and juries after an offender has been apprehended. Often, for example, judges and juries are reluctant to impose harsh punishments, even if prescribed by legislatures, and prosecutors may be unwilling to pursue cases if the punishment on conviction seems disproportionate.[14] This reluctance, however, likely wanes as offenders continue to commit crimes, thereby displaying resistance to (if not outright defiance of) previous “reasonable” efforts to deter them. For these habitual offenders, harsh punishments will eventually become acceptable as the only way to prevent them from committing further undesirable acts.

Another reason society might be averse to imposing high fines is the possibility of rent-seeking law enforcers (Garoupa and Klerman, 2002). If enforcers value the revenue from fines, they might be more aggressive in apprehending offenders than would be socially optimal. Thus, one way to control such an enforcer might be to limit the size of fines, at least to the extent that this does not conflict with other goals.

4 Extensions and Qualifications

This section examines the implications of relaxing several of the assumptions underlying the basic model in the previous section. As will be seen, some assumptions are crucial for escalating schemes to be optimal, and others are not.

4.1 Punishment Is Costless and Unbounded

As discussed above, costly punishment is key for the optimality of escalating penalties, given the existence of some offenders who, for whatever reason, cannot be deterred. Recall in particular that the advantage of low initial penalties was the savings in punishment costs for those offenders who were not deterred by the maximum feasible penalty. If $c = 0$, however, it is easy to see that scheme 5, which imposes the highest sanction on first-timers, is the lowest-cost option. More generally, when punishment is truly costless there is no reason not to set the punishment as high as possible for all offenses, regardless of an offender’s criminal history, in order to deter the maximum number of crimes. In this case, there is no economic justification for escalating penalties.

4.2 Offenders’ Gains Count in Social Welfare

Another assumption in the basic model was that offenders’ gains were not counted in social welfare. Although this assumption does not seem unreasonable, especially for

[14] See, for example, Adelstein (1981), Andreoni (1991), and Miceli (2008) for discussion of some of these aspects of the criminal punishment process.

violent crimes, it is not the usual assumption in standard economic models of crime. It turns out, however, that relaxing this assumption does not alter the qualitative conclusions of the model as long as we retain the assumption that all crimes are socially undesirable; that is, as long as we assume that $h > g_3 > g_2 > g_1$. Given this relationship between harms and benef ts, the analysis in the basic model goes through, the only change being that the gains to offenders who commit crimes are subtracted from total social costs.

The conclusion is different if some offenses are “eff cient” in the sense that $g_i > h$ for some i. If punishment is costless and unbounded, the optimal (f rst-best) policy would simply be to set $s = h$ so that eff cient crimes are committed and ineff cient ones are deterred. Offense history plays no role in determining optimal punishment in this case, because repeat crimes are in fact socially desirable. If punishment is costly, the optimal scheme is more complicated, but it again seems unlikely, based on this same logic, that escalating penalties will be optimal.[15] As discussed in the introduction, a policy of escalating penalties therefore seems only applicable to acts that are def nitely socially undesirable.

4.3 Probability of Apprehension Less than One

Consider next the implications of relaxing the assumption that the probability of apprehension, p, equals one. Suppose initially that $p < 1$ but it remains fi ed. The f rst change this requires is that the minimally deterring sanction for an offender of type g_i is now g_i/p. The maximal possible sanction, $\bar{s}$, must therefore satisfy $g_3/p > \bar{s} > g_2/p > g_1/p$ so as to ensure that type g_3’s continue to be “undeterrable.” The other effect of setting $p < 1$ is that some f rst-period offenders are not caught, and so if they commit repeat offenses in period 2 and are caught, they are treated like f rst-timers (i.e., they face a sanction of s_1). Figure 3 shows the game tree that applies to this situation. Note that, in contrast with the game tree in Figure 1, there are now *three* subgames for offenders in period 2, corresponding to the following period 1 outcomes for an offender: “commit and caught,” “commit and not caught,” and “not commit.”[16] Note, however, that the optimal period 2 behavior of offenders who committed period 1 crimes but were not caught is the same as for offenders who did not commit period 1 crimes.

After making suitable adjustments in the total-cost expressions, it is possible to show that the qualitative conclusions from the basic model carry over to the current setting.[17] It follows that the explanation for escalating penalties derived above does

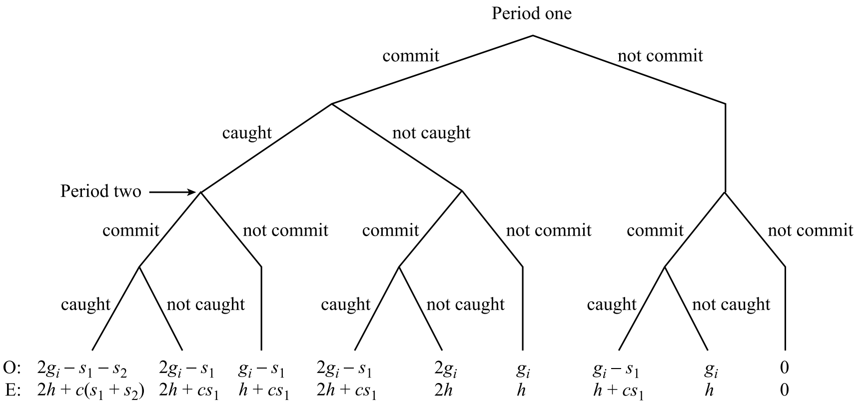
---

[15] Working out the full details of such a model is beyond the scope of this paper.

[16] The enforcer, however, cannot distinguish between offenders who did not commit a period 1 crime, and those who did but were not caught.

[17] One reviewer pointed out that the quantitative effect of allowing $p < 1$ is that the minimally deterring penalty, g_i/p, is larger, which increases punishment costs relative to the harm from criminal acts. Punishment will be imposed proportionately less often, however, making the net effect unclear.

Figure 3
Decision Tree when $p < 1$



not depend on the assumption that $p = 1$.[18] It apparently does depend, however, on the assumption that p is *fixed*, as discussed in the next subsection.

4.4 Endogenous Probability of Apprehension

Polinsky and Shavell (1998) and Emons (2004) have examined the optimal punishment of repeat offenders when the probability of apprehension is endogenous. Although their models are quite different from each other, neither found that escalating penalties was optimal. In a traditional model in which some acts are socially desirable, Polinsky and Shavell (1998) showed that sanctions are maximal for first-time offenders in period 1, and for repeat offenders who were caught in period 1, but sanctions may be less than maximal for second-period offenders without a record – namely, those who did not commit first-period crimes and those who did but were not caught. In other words, second-period offenders without a record may face a lower sanction than second-period offenders with a record. It is important to note, however, that a *given* offender who commits a crime and is caught in both periods faces the same (maximal) sanction for both offenses. In this crucial sense, the sanction does not escalate.[19]

In a model where all offenses are undesirable, Emons (2004) showed that when the probability of apprehension is endogenous and sanctions take the form of fines,

---

[18] The literature review above noted that several authors have examined the possibility that p will differ for first-time and repeat offenders, with p possibly being higher for repeaters (due to greater police effort) or lower for repeaters (as a result of learning by offenders). As indicated there, these differences will affect (in reverse directions) the desirability of escalating penalties.

[19] Polinsky and Shavell derived the main results for a fine, but conjectured that the results would carry over to the case of imprisonment.

the optimal fine schedule is declining, assuming that the enforcer can commit to a punishment scheme. If the enforcer cannot commit (i.e., if subgame perfection is imposed), the optimal sanction is either declining or flat. In either case, the total fine collected over the two-period horizon is maximal.[20]

These models show the difficulty of deriving an escalating penalty scheme when the probability of apprehension is endogenous. The logic reflects the well-known result from standard one-period enforcement models that when p is endogenous, the optimal sanction is maximal, whether it takes the form of a fine or imprisonment (Polinsky and Shavell, 2000). The intuition is obvious for a fine, which it is costless to raise. To see why it is also true for prison, suppose s is less than maximal, and then raise s and lower p so as to hold ps fixed. Since the expected sanction is not changed, the amount of crime is held constant, but expected punishment and apprehension costs, equal to $pcs + c(p)$, fall. Thus, welfare is increased.

The analysis of Polinsky and Shavell (1998) shows that this logic extends to a two-period setting, even when the enforcer can commit in period 1 to carry out the optimal enforcement plan. The requirement of subgame perfection on the part of enforcers only makes it harder to sustain a less-than-maximal sanction, because once period 2 arrives, previous decisions are sunk. Thus, optimal enforcement becomes a one-shot decision in period 2 and, by backward induction, in period 1 as well. As a result, two-period punishment schemes chosen optimally in period 1 but that do not satisfy subgame perfection are not credible.

5 *Conclusion*

Criminal penalties are aimed at deterring two kinds of offenders: those who have never committed a crime but may, and those who have already committed crimes in the past. Common sense suggests, however, that different penalties are appropriate for these two groups, with lighter punishments being set for the former and harsher punishments being reserved for the latter. While this prescription is routinely followed in a wide range of real-world punishment settings, the economic theory of law enforcement has had trouble justifying it as an optimal policy. The simple logic of the economic approach suggests that by setting the punishment equal to the social cost of an act, the efficient level of deterrence (whether complete or partial) will be achieved. Such a policy prescribes no relationship between the punishment imposed on an offender and his record. While this departure of theory from practice is not the only one that has emerged from the economics-of-crime literature, it has for some reason been seized upon by authors seeking to reconcile the two.

[20] Emons's model differs from the current one in two ways. First, he assumes that the fine constraint applies to the offender's entire asset holdings over the two-period horizon, whereas here it is imposed period by period (as in Polinsky and Shavell, 1998). Second, in the noncommitment case, Emons assumes that the enforcer is a rent-seeker in the sense of Garoupa and Klerman (2002) – that is, its objective function includes the fine revenue.

This paper has contributed to the scholarship on escalating penalties by f rst reviewing the literature on the subject, and then proposing a new theory based on the inability of enforcers to observe the gains to offenders from committing criminal acts. The analysis showed that escalating penalties can be optimal under the following assumptions: (1) all criminal acts are socially undesirable; (2) punishment is costly to impose; and (3) some offenders are undeterrable. The advantage of escalating penalties in this setting is that they balance the gains from low initial punishment costs against the reduction in deterrence, a trade-off that is cost-justif ed when the fraction of undeterrable offenders in the population is high enough. The paper f nally discussed several extensions of the basic model as a robustness check.

The overall conclusion of this study is that no one theory (including the present one) seems to offer a convincing explanation for the pervasiveness of escalating penalties. This does not necessarily reflec a failure of the theory, however, for sometimes a theory can be as useful for what is does not explain (or at least for what it has diff culty explaining) as for what it does. The criminal justice system is a complex social institution that serves many goals, only one of which is economic eff ciency. Also important are concepts of fairness and proportionality, which seem to be hardwired into human thinking about punishment in a way that theory cannot easily rationalize. After reading the extensive literature on escalating penalties and thinking hard about the issue for a long time, I am convinced that purely economic models miss an important reason for their pervasiveness – they just “feel right.”

References

Adelstein, R. (1981), “Institutional Function and Evolution in the Criminal Process,” *Northwestern University Law Review*, 76(1), 1–99.

Andreoni, J. (1991), “Reasonable Doubt and the Optimal Magnitude of Fines: Should the Penalty Fit the Crime?” *The RAND Journal of Economics*, 22(3), 385–395.

Baik, K. H., and I.-G. Kim (2001), “Optimal Punishment when Individuals May Learn Deviant Values,” *International Review of Law and Economics*, 21(3), 271–285.

Becker, G. S. (1968), “Crime and Punishment: An Economic Approach,” *Journal of Political Economy*, 76(2), 169–217.

Burnovski, M., and Z. Safra (1994), “Deterrence Effects of Sequential Punishment Policies: Should Repeat Offenders be More Severely Punished?” *International Review of Law and Economics*, 14(3), 341–350.

Chu, C. Y. C., S. c. Hu, and T. y. Huang (2000), “Punishing Repeat Offenders More Severely,” *International Review of Law and Economics*, 20(1), 127–140.

Dana, D. A. (2001), “Rethinking the Puzzle of Escalating Penalties for Repeat Offenders,” *The Yale Law Journal*, 110(5), 733–783.

Emons, W. (2003), “A Note on the Optimal Punishment for Repeat Offenders,” *International Review of Law and Economics*, 23(3), 253–259.

— (2004), “Subgame-Perfect Punishment for Repeat Offenders,” *Economic Inquiry*, 42(3), 496–502.

— (2007), “Escalating Penalties for Repeat Offenders,” *International Review of Law and Economics*, 27(2), 170–178.

Friedman, D. D. (2000), *Law’s Order: What Economics has to Do with Law and Why it Matters*, Princeton University Press, Princeton (NJ).

Funk, P. (2004), "On the Effective Use of Stigma as a Crime-Deterrent," *European Economic Review*, 48(4), 715–728.
Garoupa, N., and M. Jellal (2004), "Dynamic Law Enforcement with Learning," *The Journal of Law, Economcs, & Organization*, 20(1), 192–206.
— and D. Klerman (2002), "Optimal Law Enforcement with a Rent-Seeking Government," *American Law and Economics Review*, 4(1), 116–140.
Hart, H. L. A. (1982), *Punishment and Responsibility: Essays in the Philosophy of Law*, Clarendon Press, Oxford.
Kaplow, L. (1990), "A Note on the Optimal Use of Nonmonetary Sanctions," *Journal of Public Economics*, 42(2), 245–247.
Lewin, J. L. , and W. N. Trumbull (1990), "The Social Value of Crime?" *International Review of Law and Economics*, 10(3), 271–284.
McCannon, B. C. (2009), "Differentiating between First and Repeat Offenders," *Contemporary Economic Policy*, 27(1), 76–85.
Miceli, T. J. (2008), "Criminal Sentencing Guidelines and Judicial Discretion," *Contemporary Economic Policy*, 26(2), 207–215.
— (2010), "A Model of Criminal Sanctions that Incorporates Both Deterrence and Incapacitation," *Economics Letters,* 107(2), 205–207.
— and C. Bucci (2005), "A Simple Theory of Increasing Penalties for Repeat Offenders," *Review of Law and Economics*, 1(1), 71–80.
Mungan, M. (2010), "Repeat Offenders: If they Learn, we Punish them More Severely," *International Review of Law and Economics*, 30(2), 173–177.
Polinsky, A. M., and D. L. Rubinfeld (1991), "A Model of Fines for Repeat Offenders," *Journal of Public Economics*, 46(3), 291–306.
— and S. Shavell (1998), "On Offense History and the Theory of Deterrence," *International Review of Law and Economics*, 18(3), 305–324.
— and — (2000), "The Economic Theory of Public Law Enforcement of Law," *Journal of Economic Literature*, 38(1), 45–76.
Posner, R. (2003), *Economic Analysis of Law*, 6th edn., Aspen Publishers, New York.
Rasmusen, E. (1996), "Stigma and Self-Fulfillin Expectations of Criminality," *The Journal of Law & Economics*, 39(2), 519–543.
Rubinstein, A. (1979), "An Optimal Conviction Policy for Offenses that May have been Committed by Accident," in: S. J. Brams, A. Schotter, and G. Schwödiauer (eds.), *Applied Game Theory*, Physica-Verlag, Würzburg, pp. 406–413 .
— (1980), "On an Anomaly of the Deterrent Effect of Punishment," *Economics Letters*, 6(1), 89–94.
Shavell, S. (1987), "A Model of Optimal Incapacitation," *The American Economic Review*, Papers and Proceedings, 77(2), 107–110.
Stigler, G. J. (1970), "The Optimum Enforcement of Laws," *Journal of Political Economy*, 78(3), 526–536.

Thomas J. Miceli
Department of Economics
University of Connecticut
365 Fairfiel Way
Storrs, CT 06269-1063
U.S.A.
Thomas.Miceli@UConn.edu

Incentive-Compatible Reimbursement Schemes for Physicians

by

Winand Emons*

Received November 25, 2012; in revised form May 1, 2013; accepted July 18, 2013

Physicians choose capacity before demand materializes; actual demand may be higher or lower than capacity. If a physician's capacity exceeds demand, she may have an incentive to overtreat, i.e., she may provide unnecessary treatments to use up idle capacity. By contrast, with excess demand she may undertreat, i.e., she may not provide necessary treatments because other activities are f nancially more attractive. We firs show that simple fee-for-service reimbursement schemes do not provide proper incentives for all demand realizations. If, however, insurers use fee-for-service schemes with quantity restrictions, they solve the fraudulent-physician problem. (JEL: D82, I11)

1 Introduction

The U.S. spends between one-f fth and one-third of its health-care expenditures, that is, between 500 and 700 billion dollars, on care that does not improve anybody's health. These unnecessary tests and treatments are not just expensive; they can also harm patients.[1]

One factor contributing to this enormous waste is that medical services constitute *credence goods*: a physician not only provides the medical services; at the same time she also acts as the expert who determines how much treatment is necessary, because her patient is unfamiliar with the medical condition. Furthermore, ex post the patient typically cannot determine which treatment was required ex ante. It is often impossible to f nd out whether provided treatments were necessary or whether necessary treatments were not provided. From ex post observations the patient can

* University of Bern, CEPR, CIRPÉE. I thank Uwe Dulleck, Rudi Kerschbamer, Albert Ma, Fridolin Marty, Pius Matter, Gerd Muehlheusser, Klaus Neusser, and two referees for helpful comments.

[1] See, e.g., Brownlee (2007, p. 5) or Reilly and Evans (2009). Likewise, the Canadian Association of Radiologists estimates that 30% of imaging in the Canadian health-care system is unnecessary; see http://www.car.ca/uploads/news%20publications/car_cat_scan_eng.pdf.

Journal of Institutional and Theoretical Economics 169, 605–620 – ISSN 0932-4569
DOI: 10.1628/093245613X671869

never be certain of the quality of the treatments he obtained; that is why such services have been termed credence goods (Darby and Karni, 1973).

This information advantage may induce physicians to behave opportunistically: they may recommend unnecessary yet profi able treatments, or they may not perform urgently needed yet unprof table treatments.

For example, in the Swiss canton of Ticino the population average underwent 33% more of the seven most important operations than medical doctors and their families. Interestingly enough, lawyers and their loved ones have about the same operation frequency as the families of medical doctors (Domenighetti et al., 1993). Marty (1998) shows, using 8000 bills of Swiss general practitioners, that doctors with suff cient demand charge significa tly less per patient than doctors with excess capacity. For Germany Jürges (2009) find no evidence for demand inducement for statutorily insured patients; there is, however, demand inducement among the patients with private insurance, which reimburses medical doctors at higher rates than their statutory counterparts. Even in China overprescription is routine; hospitals use the resulting prof ts to subsidize underfunded operations (*The Economist*, July 11, 1998). Gruber, Kim, and Mayzlin (1999) show that in the U.S. the frequency of cesarean deliveries compared to vaginal deliveries reacts positively to fee differentials of health insurance programs. Primary-care physicians are squeezed financiall , so that their numbers dwindle; at the same time, the number of the highly prof table specialists continues to rise, leading Brownlee (2007, p. 265) to sigh: "[...] sometimes what we really need is not a doctor who delivers more care but one who seems to care more [...]."

Brownlee (2007, p. 8) mentions a couple of other reasons for overtreatment: doctors simply do not know which treatments are most effective, they want to help patients even when they do not know the right thing to do, malpractice fears drive defensive medicine, medical custom varies from region to region, one doctor often does not know that another one has already ordered a battery of tests, and patients, being insured, ask for fancy treatments (demand-induced supply). Yet, as she elaborates in the book, the most powerful reason for overtreatment is that doctors and hospitals get paid more for doing more. Interestingly, in a 2009 survey among 627 U.S. primary-care physicians, only 3% of the respondents said money inf uences their practice, but most think money does inf uence the practice of other physicians: 62% said there would be fewer diagnostic tests if tests did not create revenue for subspecialists, and 39% think the same of primary-care doctors.[2]

In this paper we analyze whether health insurers can design reimbursement schemes so that physicians have no incentives to behave fraudulently; by fraudulent behavior we mean that a physician performs unnecessary treatments or does not perform necessary treatments.[3] We fir t show that simple fee-for-service reimbursement

[2] http://dartmed.dartmouth.edu/winter11/html/disc_study/.

[3] Mark Twain (1885, p. 128) describes this behavior as follows: "Well, then, says I, what's the use you learning to do right, when it's troublesome to do right and ain't no trouble to do wrong, and the wages is just the same?" Bernard Shaw (1922, p. xiii) writes: "That any sane nation having observed that you could provide for the supply

schemes do not provide proper incentives. If, however, insurers use fee-for-service schemes with quantity restrictions, they solve the fraudulent-physician problem.

As a workhorse we use the basic model of Emons (1997, 2001). Patients are in need of a checkup. Some patients are in good condition and require no further treatment; the rest are in bad condition and need treatment. After the diagnosis the physician knows which condition the patient is in. She can then treat him. The physician can only perform the treatment after a diagnosis. We thus have economies of scope between diagnosis and treatment, making the separation of diagnosis and treatment inefficient.[4]

We consider a set of physicians, each of whom chooses a fixed capacity: when actual demand realizes, a physician may have to ration her patients due to insufficient capacity, or she may also end up with idle capacity. If a physician has excess demand, she may undertreat patients, i.e., she may not provide necessary treatments if diagnosis is financially more attractive than treatment. By contrast, with excess capacity the physician may start to overtreat, i.e., provide unnecessary treatments to use up idle capacity.

An insurer sets reimbursement terms. Then physicians choose their capacity levels; we focus on symmetric strategies. Nature then determines total demand; each physician gets an equal share thereof. This modeling implies that all doctors have either excess capacity or excess demand. Doctors decide how many of their patients they want to diagnose; patients who obtain no diagnosis end up with no service. Having diagnosed her patients, a doctor then decides whom to treat.

Physicians choose their capacity and their diagnosis and treatment policy so as to maximize profits. The insurer wants to induce nonfraudulent services. Moreover, he wants to implement some average capacity level, which in turn implies that with positive probability demand may exceed capacity and vice versa.

We first analyze simple fee-for-service reimbursement schemes: the physician is paid per diagnosis and per treatment she performs. We show that there exists no fee-for-service scheme under which patients get nonfraudulent services for all possible demand realizations. Consider, for example, equal-compensation prices equalizing the profit per diagnosis with the profit per treatment. With these prices doctors are indifferent between diagnosis and treatment and, accordingly, provide honest services if demand exceeds capacity. Yet if doctors have excess capacity, they overtreat to use up their idle capacity.[5] By contrast, with a fully capitated scheme where the treatment price is zero, the incentive to overtreat disappears and

---

of bread by giving bakers a pecuniary interest in baking for you, should go on to give a surgeon a pecuniary interest in cutting off your leg, is enough to make one despair of political humanity."

[4] This separation mechanism is often encountered in the prescription and preparation of drugs: the physician prescribes the drugs, and the pharmacist may only sell only what has been prescribed by the doctor.

[5] "Fee-for-service is especially inflationary in the context of physician oversupply; there is nothing more expensive than an underemployed specialist" (Robinson, 2001, p. 156).

doctors behave nonfraudulently when they have excess capacity. But now doctors undertreat when they have excess demand, because diagnosis is much more attractive than treatment. It is thus impossible to find fee-for-service schemes that give proper incentives for all possible demand realizations, i.e., when physicians have excess demand or when they have excess capacity.

In the next step we use the fact that the insurer has more information than the individual patient. Whereas the patient has only one observation of the physician's behavior, the insurance company has the set of observations for its entire clientele. In particular, the insurer knows how many of its policy holders actually underwent treatment with a particular doctor. In addition to the fees for services, the insurer can thus use a quota that states the maximum fraction of diagnosed patients per physician for which the insurer pays for treatment.

Obviously, this quota needs to be equal to the fraction of patients actually in need of treatment. If the quota is lower, it enforces undertreatment; if it is higher, it opens the door for overtreatment. It turns out that a quota equal to the fraction of patients in need of treatment curbs overtreatment. If a doctor wishes to overtreat to use up idle capacity, she is not reimbursed for such treatments. We are thus only left with the problem of undertreatment if a doctor has excess demand. This problem is solved by prices making diagnosis not more attractive than treatment. With these prices a doctor prefers providing necessary treatment to diagnosing another patient. The level of the prices determines the physicians' capacity choice: the higher the revenue per patient, the higher their capacity choice. Physicians make positive expected profits.

The literature on credence goods as surveyed by Dulleck and Kerschbamer (2006) looks at one-shot relationships between the expert and her customer. The customer has only one observation of the expert's actions. This information together with the outcome of his case does not allow the customer to draw inferences about the appropriateness of the treatment he has received.[6] Most of this literature considers experts operating in a market environment. The only model we are aware of incorporating insurance in a credence-good setup is that of Sülzle and Wambach (2005). They take prices as given and analyze the influence of coinsurance on the physician's incentives to cheat and on the patients' incentive to search for a second opinion. They do not attempt to find contracts inducing nonfraudulent behavior.

In Ely and Välimäki (2003) short-lived motorists play a repeated game with long-lived mechanics. Good mechanics prefer to act truthfully, while bad mechanics prefer to always change the engine. Each motorist observes the repairs performed for preceding customers but has no idea whether these repairs were appropriate.[7]

6 Typically, this literature assumes the undertreatment problem away and deals only with the overtreatment one; see our discussion below.

7 Ely and Välimäki assume that a motorist finds out ex post whether or not he received the appropriate service. Strictly speaking, they do not analyze a credence good, but a horizontally differentiated experience good. Yet, the motorist takes the information about the appropriateness of the repair with him to his grave. Thus, the following motorists know which repair he got, but do not know whether it was appropriate.

Good mechanics may refrain from doing necessary engine replacements early on in the game, to separate themselves from the bad mechanics and signal their good type to future motorists. Motorists anticipate this incentive to undertreat to build up a good reputation and may not visit the mechanic in the f rst place. Like us, Ely and Välimäki use the information of the expert's treatment history. For them prices are exogenously given; they are such that the bad mechanic always wants to change the engine. Ely and Välimäki do not analyze how the bad mechanic's incentives can be aligned with prices. By contrast, we also determine reimbursement prices that, together with the quota, give experts proper incentives so that the outcome is effic ent.

In the health economics literature physician-induced demand has been studied in a variety of models. Farley (1986) and De Jaegher and Jegers (2000) are models based on demand-setting and altruism. In Dranove (1988) patients make rational decisions about whether or not to accept a doctor's recommendation; informed patients will be subject to less inducement than less informed patients. Calcott (1999) and De Jaegher and Jegers (2001) model demand inducement as cheap-talk games and derive equilibria with or without demand inducement. See McGuire (2000) for a survey of the earlier literature. The more recent literature stresses the gate-keeping role of general practitioners: besides diagnosis and treatment, the general practitioner also refers patients to specialists; see, e.g., Brekke, Nuscheler, and Straume (2007) or Allard, Jelovac, and Léger (2011).

The rest of the paper is organized as follows. The next section introduces the basic model. In section 3 we look at fee-for-service reimbursement schemes. In the section after, we extend fee-for-service reimbursement with quantity restrictions. Section 5 concludes.

2 *The Model*

An agent needs a medical checkup. During the period to come, the individual may fall ill or he may stay healthy. At the time of diagnosis the agent may be in good or bad condition. If the patient is in good condition, the probability of staying healthy is $q_h \in (0, 1)$; if the patient is in bad condition, the probability of staying healthy is $q_\ell \in (0, q_h)$, i.e., lower than when the consumer is in good condition. Let $p \in (0, 1)$ be the probability that the patient is in bad condition. The patient does not know which of the two conditions he is in, nor can he infer it ex post, since he may fall ill or stay healthy under either condition.

The patient visits one of n medical doctors, indexed by $i = 1, \ldots, n$; in what follows we will suppress the index i wherever possible. By diagnosing the agent, the physician detects his true condition. When the patient is in good condition, he needs no further treatment. When the consumer is in bad condition, the doctor should treat him; after the treatment the consumer is in good condition. A treatment is only possible after diagnosis.

Each physician makes a prior sunk capacity decision determining L units of time that she devotes to her practice. Since we normalize the doctors' reservation wage to 1, L also measures a physician's sunk cost. The capacity L can only be allocated between diagnosis and treatment: $d > 0$ is the time a doctor needs per diagnosis, and $t > 0$ the time per treatment; given our normalization, d and t also measure the minimum average costs of diagnosis and treatment. Note that marginal costs are different from average costs. A doctor has a fxed capacity, the cost of which is sunk. Therefore, her marginal costs are 0 except for the capacity margin, where marginal costs are "$+\infty$." When, in the following, we talk about minimum average costs, we mean d and t.[8] If there are additional variable costs per diagnosis and per treatment, the fees for service we introduce in the next section are simply the doctor's remuneration net of these variable costs.

There is a continuum of identical consumers, the mass X of which is random; it has continuous density $g(X)$ over the support $[\underline{X}, \overline{X}]$. In units of time a capacity of $X(d + pt)$ is necessary to serve the entire market. Each physician gets an equal share of consumers. A doctor's demand is thus $x := X/n$, which is distributed on $[\underline{x}, \overline{x}]$ with density $f(x) := ng(X)$; denote the c.d.f. by $F(x)$. Patients' risks are independent and identically distributed. We assume that a continuum of such random variables sums to a nonrandom variable. The size of a physician's demand is thus random, whereas the fraction of her patients in need of treatment is nonrandom (see Judd, 1985).[9] Defne $\lambda = L/(d + pt)$ as a doctor's capacity, in terms of customers, given nonfraudulent behavior. Since we look for symmetric strategies, the total capacity equals $n\lambda$. According as $n\lambda \lesseqgtr X$, there is too little, suffcient, or excess capacity in the market. Due to our symmetry assumption, market conditions translate to the individual physician level: either all doctors have excess capacity or they all face excess demand.

Let us now look at a doctor's incentives. After diagnosis the physician knows the patient's condition. When the patient is in bad condition, she can perform a treatment that changes him to a good condition. Yet she can also treat a patient in good condition; in this case the physician wastes t units of time, leaving the patient at least in good condition. This kind of behavior has been termed overtreatment (Dulleck and Kerschbamer, 2006) or supplier-induced demand in health economics (Labelle, Stoddart, and Rice, 1994).

If the patient is in good condition, the medical doctor can recommend no treatment. Nevertheless, the same recommendation is also possible when the patient is in bad condition. We will refer to this type of fraud as undertreatment. Most of

---

8 We assume that diagnosing and treating a patient if necessary is efficient Our diagnosis corresponds to Dulleck and Kerschbamer's (2006) cheap treatment; their expensive treatment corresponds to our "diagnosis cum treatment."

9 We make the continuum assumption not only for notational convenience. With a fnite number of consumers we run into the following problem. Suppose the physician expects a clientele with $1 - p$ patients in good and p patients in bad condition. With a fnite number of customers, however, the actual realization of her clientele will typically be different from the expected one. Accordingly, at the end of the day she has either insuffcient or excess capacity, and she will start behaving fraudulently (which suggests that it is better to see a doctor in the morning than in the late afternoon).

the credence-goods literature assumes the undertreatment problem away by setting $q_h = 1$. Under this assumption a patient knows for sure that he did not get the necessary treatment when he falls ill. Moreover, the patient's health status is verif able, and a legal rule holds the physician liable if the patient becomes sick; see, e.g., Dulleck and Kerschbamer (2006).

Ex post the patient cannot fin out whether he was treated unnecessarily or whether necessary treatment was not provided. The physician's services thus constitute credence goods as distinct from search and experience goods – from ex post observations the consumer can never be certain of the quality of the services he got.

Note that we assume diagnosis and treatment to be verif able. This assumption seems appropriate for physicians, whose patients necessarily take part in any treatment. It is not appropriate for, e.g., a consumer who sends his gadget to a service center. When the gadget is returned, the customer is unable to tell whether somebody in the repair center has actually worked on it. Here the expert has yet another possibility to defraud her customers. She can claim to have f xed the widget without having touched it, thus collecting repair fees from an unlimited number of customers. See, e.g., Emons (2001) or Dulleck and Kerschbamer (2006) for setups where the expert's actions are not verif able.

All patients have full insurance from one insurance company. The insurer reimburses the physicians. The sequence of events is as follows. The insurance company chooses the reimbursement terms. Next, physicians choose their capacity. Then Nature chooses the total mass X or, equivalently, a physician's mass x of patients. A physician then decides how many patients $\mu \leq x$ she diagnoses; $x - \mu$ patients get no service. After having diagnosed her μ patients, the doctor then decides whom to treat.

Medical doctors maximize profi s. The insurer wants to f nd reimbursement terms that induce nonfraudulent behavior. Moreover, the insurer wants to implement an average aggregate capacity level, meaning for the physician a capacity $\lambda \in (\underline{x}, \overline{x})$. We thus assume that neither the aggregate capacity $\underline{X}$ nor the capacity $\overline{X}$ is optimal.[10] Accordingly, with positive probability there will be excess capacity, and with positive probability there will be excess demand. We will now look at different reimbursement schemes.

3 *Fee for Service*

Under a simple fee-for-service reimbursement scheme the physician gets D per performed diagnosis and T per performed treatment. Consider the subgame starting after the physician has chosen a capacity of L units of time, which by then is sunk. In terms of patients the physician has capacity $\lambda < \overline{x}$, given honest behavior.

[10] Diagnosis and treatment are eff cient; see footnote 8. This does not, however, imply that the maximum capacity $\overline{X}$, which is needed with probability zero, is effi cient. Rather than specifying $g(\cdot)$ and deriving the optimal capacity level, we show that any interior capacity level can be implemented.

Apparently, her behavior depends on the size of her clientele x relative to her capacity λ. According as $x \gtreqless \lambda$, we will say that the physician has too many, enough, or not enough patients, given nonfraudulent behavior. If, say, the doctor does not have enough patients, she may start treating patients in good condition to utilize her otherwise idle capacities. If she has too many patients, she may, e.g., be tempted not to treat all patients in bad condition, given that diagnosis is more profitable than treatment.

The physician's incentives also depend on the profitability of diagnosis relative to treatment, which, in turn, is determined by the prices D and T. If the doctor has too many patients, she only faces (at the margin) her time constraint. She compares the profit per hour of treatment, T/t, with the profit per hour of diagnosis, D/d.[11] If the former exceeds the latter, she will overtreat, whereas she will undertreat if diagnosis is more profitable than treatment. We specify these ideas more precisely in the following lemma; here we assume that if the doctor is indifferent between nonfraudulent and fraudulent behavior, she opts for the former.[12]

LEMMA 1

(i) If $x > \lambda$, the physician is honest if and only if $T = tD/d$;
(ii) if $x = \lambda$, the doctor is honest if and only if $T \leq tD/d$;
(iii) if $x < \lambda$, the doctor is honest if and only if $T = 0$.

PROOF (i) If $x > \lambda$, the doctor has more patients than she can handle with honest behavior. Given her time constraint, she is only interested in the profit per hour of treatment, T/t, compared to the profit per hour of diagnosis, D/d. If $T = tD/d$, she is indifferent between diagnosis and treatment and therefore behaves honestly. If $T > tD/d$, she prefers treatment to diagnosis and thus overtreats; she undertreats if $T < tD/d$.

(ii) If $x = \lambda$, the physician fully utilizes her capacity with nonfraudulent behavior. If $T < tD/d$, she strictly prefers diagnosis to treatment; yet she makes diagnoses for her entire clientele. She has to perform treatments to use up her remaining time $L - xd$; honestly treating the patients in bad condition of her clientele just exhausts her capacity. If $T = tD/d$, the argument is along similar lines to that for (i). If $T > tD/d$, the physician strongly prefers treatment to diagnosis. Hence, she will treat all patients she diagnoses.

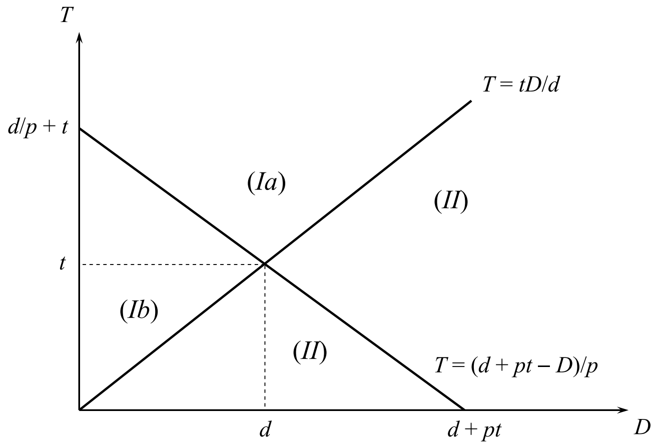
(iii) If $x < \lambda$, the doctor has idle capacity with honest behavior. As long as $T > 0$, she makes money by treating some more patients to use her idle capacity. Only when $T = 0$ does the incentive for overtreatment disappear. *Q.E.D.*

To explain Lemma 1 consider the Figure. Along the line $T = tD/d$ the profit per hour of diagnosis equals the profit per hour of treatment: on this equal-compensation line the physician is indifferent between diagnosis and treatment, so that with too

[11] Recall that the capacity cost is sunk; our results, however, do not change if we defin profit per hour as $(D - d)/d$ and $(T - t)/t$.

[12] This result corresponds to Lemma 1 in Emons (1997).

Figure
The Equal-Compensation and the Zero-Profi Line



many patients she picks eff cient treatment. She diagnoses $\mu = \lambda$ patients and treats a fraction p thereof; the remaining $x - \lambda$ patients get no treatment.

In regions (*Ia*) and (*Ib*), where $T > tD/d$, the doctor prefers treatment to diagnosis. Whatever the number of patients, she will treat everybody she diagnoses, i.e., she will overtreat. In region (*II*), in which $T < tD/d$, the physician prefers diagnosis to treatment, so she wishes to increase the number of diagnoses at the expense of treatments. If the physician has too many patients, we will observe undertreatment. With enough patients, however, she cannot diagnose more patients; she treats eff ciently to make some money out of her otherwise unused capacity.

When the physician does not have enough patients, she will overtreat as long as $T > 0$. Only when $T = 0$ does the physician have proper incentives if she does not have enough patients. She does not overtreat to utilize her idle capacity, because there is no money in treatment. When we rule out the prices $D = T = 0$, which provide no incentives whatsoever, we can summarize our fi dings as follows:

PROPOSITION 1 *If the insurer uses simple fee-for-service reimbursement schemes* (D, T)*, there exists no set of prices under which for all demand realizations patients get nonfraudulent services.*

In the f gure we have also depicted the line $T = (d + pt - D)/p$. All prices along this line generate zero prof ts when a physician serves λ patients nonfraudulently. Suppose, for example, the insurance company reimburses equal-compensation prices (d, t). Then if physicians have enough or too many customers, they have proper incentives and they make zero prof ts. Yet, if a physician has excess capacity, she will overtreat.

If insurers use, say, the fully capitated reimbursement $(d + pt, 0)$, physicians have proper incentives with enough or too few patients.[13] Yet, if there is excess demand, doctors will undertreat.

Note that this negative result is driven by the physician's f xed capacity. In a setup where experts only incur variable costs, equal-compensation prices always induce honest behavior; see, e.g., Dulleck and Kerschbamer (2006).[14]

4 Fee for Service with Quantity Restrictions

Let us now use the fact that the insurance company has more information than an individual patient. Whereas the patient has only one observation of the physician's behavior, the insurance company has the set of observations for its entire clientele. In particular, the insurer knows how many of its policy holders actually underwent treatment.

The insurer offers reimbursement schemes (D, T, z), where z denotes the maximum fraction of diagnosed patients for whom the insurer actually pays for the treatment. It turns out that the quota z is a powerful instrument to curb overtreatment.

First note that to implement eff cient behavior we need $z = p$. If $z < p$, the insurer enforces undertreatment. If, by contrast, $z > p$, we run into the problems described by Lemma 1.

LEMMA 2 *Let* $z = p$.

(i) If $x > \lambda$, *the physician is honest if and only if* $T \geq tD/d$;
(ii) if $x \leq \lambda$, *the doctor is honest for all prices* (D, T).

PROOF (i) If $x > \lambda$, the physician has more patients than she can handle with honest behavior. If $T < tD/d$, she prefers diagnosis to treatment. She diagnoses all x patients (or the number exhausting her capacity) and uses her remaining capacity (if any) to treat a few patients. We thus have undertreatment.

If $T = tD/d$, the physician is indifferent between diagnosis and treatment and therefore deals honestly with λ patients.

If $T > tD/d$, the physician prefers treatment to diagnosis. She would like to treat all patients she diagnoses. Yet she can bill for treatments only for the fraction p of the patients she diagnoses. To use up her capacity, she diagnoses λ patients and treats the fraction p thereof who are in bad condition.

(ii) If $x = \lambda$, the physician fully uses her capacity with nonfraudulent behavior. If $T = tD/d$, she has proper incentives and uses up her capacity by honestly serving all patients. If $T < tD/d$, she prefers diagnosis to treatment. She diagnoses all patients;

[13] With enough patients the physician makes zero profit ; with excess capacity, however, she suffers losses.

[14] In Dulleck and Kerschbamer (2006) doctors have no capacity constraint and t and d are are the marginal costs, so that equal-compensation (equal-markup) prices satisfy $T - t = D - d$. Another setup with capacity-constrained experts can be found in Richardson (1999).

to use up her remaining time $L - xd$ she has to treat. Honestly treating the patients in bad condition just exhausts her capacity. If $T > tD/d$, the doctor prefers treatment to diagnosis. She would like to treat all patients but is curbed by the quota p. Hence, she behaves honestly.

If $x < \lambda$, the physician has unused capacity with nonfraudulent behavior. As long as $D > 0$, she will diagnose all x patients. As long as $T > 0$, she would like to treat more than px patients to use her idle capacity. Yet she cannot bill more than px patients for treatment. *Q.E.D.*

The quota $z = p$ induces honest behavior for all prices if the physician has enough or not enough demand. With excess capacity the physician diagnoses all patients. As long as $T > 0$, she would like to overtreat to use her idle capacity. Yet the reimbursement quota prevents her from doing so. By contrast, if the physician has excess demand, diagnosis may not be more attractive than treatment. If diagnosis is more prof table than treatment, the physician will diagnose all patients she can get hold of and treat less than the fraction p thereof, i.e., we have undertreatment.

Def ne region (*Ia*) as all prices (D, T) on and above the equal-compensation line and *above* the zero-prof t line. Lemma 2 implies the following result:

PROPOSITION 2 *Under the reimbursement schemes* (D, T, p) *with* (D, T) *in region (Ia), all patients get nonfraudulent service. Each physician chooses the capacity* $\lambda^* \in (\underline{x}, \overline{x})$ *solving* $F(\lambda^*) = 1 - (d + pt)/(D + pT)$*; a physician's equilibrium profi is* $(D + pT)\int_{\underline{x}}^{\lambda^*} xf(x)dx > 0$.

PROOF Lemma 2 implies that for reimbursement schemes (D, T, p) with (D, T) in region (*Ia*) physicians have proper incentives whatever their demand. Given nonfraudulent behavior, physicians choose their capacity λ so as to maximize $(D + pT)[\int_{\underline{x}}^{\lambda} xf(x)dx + \lambda(1 - F(\lambda))] - \lambda(d + pt)$. Solving the f rst-order condition yields $F(\lambda^*) = 1 - (d + pt)/(D + pT)$. *Q.E.D.*

Let us f rst comment on the capacity choice. If the insurer picks prices on the zero-prof t line, then $D + pT = d + pt$, so that $\lambda = \underline{x}$. Doctors pick the capacity level they can sell for sure and break even. For prices above the zero-prof t line (region (*Ia*)), then the revenue per customer $D + pT > d + pt$, and physicians choose capacity $\lambda > \underline{x}$; with positive probability a doctor has idle capacity. Physicians make positive expected prof ts. The capacity level increases with the revenue per customer and is below $\overline{x}$.

There thus exist reimbursement schemes (D, T, z) inducing nonfraudulent behavior for all realizations of demand. With the level of prices (D, T) the insurer controls the capacity that is provided by physicians. The higher the prices, the more capacity they provide. Physicians' prof ts increase with the price level.

In our setup the insurer knows the equilibrium capacity level of a physician. Nevertheless, for our incentive scheme to work once capacity is chosen, the insurer need *not* know the physician's actual capacity level.[15]

[15] Assessing a physician's capacity is a tricky task. For example, in Switzerland a lot of (in particular female) physicians prefer to work part- rather than full-time,

A few qualifying remarks are in order. In our model the number of patients is a continuum. Each physician serves a fraction of the market. Therefore, a doctor's clientele is also a continuum. We assume that a continuum of independent and identically distributed random variables sums to a nonrandom variable. To be more specif c, a physician has a continuum of patients, the fraction p of which is in need of treatment; see also the discussion in footnote 9. The reimbursement quota $z = p$, therefore, coincides with the actual number of patients in need of treatment.

With a f nite population the actual number of patients in need of treatment will typically be different from the expected value. This creates problems at the aggregate-insurance and at the individual-physician level. If at the insurance level the actual fraction of patients in need of treatment is above p, our quota $z = p$ leads to undertreatment. If the actual fraction is below p, doctors will overtreat. This problem becomes smaller, the more clients the insurance company has.[16]

At the physician level our results change as follows. Assume, for the sake the argument, that at the insurance level the actual realization equals the expected value p. When deciding on how many patients to diagnose, the doctor bases her decision on the expected value p, as in our setup; in particular, a physician with excess demand diagnoses λ patients. Nevertheless, when it comes to the treatment decision, the physician treats the fraction $z = p$ independently of the actual needs of her clientele. Thus, if the actual fraction of patients in need of treatment is below z, the physician overtreats, and if it is above z, she undertreats. To sum up: with a f nite population our quota system works best for large insurance companies the clientele of which are treated by relatively few physicians.

Another diff culty arises if patients are not identical as in our setup. Suppose the probability of being in need of the treatment is distributed in the population on [0, 1] with mean p, the density having full support. As long as each physician gets a random sample of the population, our results continue to hold. If, however, there is a selection bias such that some physicians get on average less healthy (i.e., higher-p) patients than others, our one-size-f ts-all quota no longer gives proper incentives for all physicians. The quota then has to be adjusted to the group of patients seeing the doctor. With excess capacity a doctor with high-p patients does better than her colleague with a low-p group; she uses up more of her idle capacity. With excess demand, if treatment is more attractive than diagnosis, a physician also prefers high-p patients; with equal-compensation prices doctors facing excess demand are indifferent to the health status of their clientele. Fee for services with an adjusted quota thus makes less healthy patients in expectation attractive for physicians. This is in stark contrast to capitation, where physicians try to skim the healthy and avoid the ill.

making her capacity level her private information. Any reimbursement scheme that builds on a physician's capacity level, therefore, has to deal with the issue of how this information is revealed.

16 In Switzerland, Santésuisse, the association of Swiss health insurers, collects and aggregates the data of the individual insurers on a physician's behavior, to check, e.g., that the doctor does not bill more than 24 hours a day.

We have assumed that only one treatment is available and thus that the fraction of patients in need of treatment is well def ned. Often there are, however, professional disagreements covering the diagnosis and treatment of illness. For example, Wennberg, Bames, and Zubkoff (1982) show that the wide range of acceptable diagnoses and therapies is a major factor in the wide variation in rates of utilization and costs of medical services among neighboring medical markets. Our analysis, therefore, applies to diseases where there are no professional disagreements, or to cases where insurers enforce the most effective way of dealing with the illness. If there are several treatment options and the physician has private information on what the best treatment is for the patient, prof ts per hour of treatment have to be equalized across all treatments, and quotas for each treatment have to be implemented.

Despite these shortcomings of our simple model, we think that treatment quotas should be a useful instrument for insurers to curb overtreatment incentives. To our knowledge, insurers tend to make little use of this instrument. In the U.S. physicians are paid bonuses to restrict the percentage of patients who are given referrals (Grumbach et al., 1998).[17] Such bonuses may give incentives not to overtreat at the margin. If, however, excess capacity is suff ciently large, overtreating is more profi able than cashing in on the bonus. In Switzerland insurers start an investigation if a physician's actual billing per patient is 30% higher than the average for her physician group.[18] Here it is unclear what the average actually measures: ineff ciencies may be being compared with ineffic encies.

In Germany physicians are endowed each quarter with a so-called *budget*. Once they exceed this budget, they get paid less per treatment. At first insurers did not pay at all for any service provided outside the budget; the quota (*Praxisbudget*) was strict, like our quota z. Poorly set quotas made it, however, diff cult for statutorily insured patients to see their doctor at the end of a quarter, because her budget was exhausted. Therefore, in 2009 the quota was softened (*Regelleistungsvolumen*); now the doctor's reimbursement goes down from 100% to 75% to 50% to 25% once she exceeds the budget. It is not entirely clear how the budget is determined; since it is def ned for the entire practice, it is perhaps too broad. Furthermore, the budget is based on past behavior. It may thus lead to ratcheting: staying within the budget today may lead the regulator to lower the budget tomorrow. Nevertheless, the *Praxisbudget* seems to have curbed overtreatment, albeit at the expense of some undertreatment. The new *Regelleistungsvolumina* try to deal with the undertreatment problem while maintaining the virtues of quotas concerning overtreatment. Our results tend to support the German approach. Yet, a more sophisticated use of treatment records seems warranted.

Besides physician-induced demand, the classic moral-hazard problem that consumers with full insurance tend to overconsume health care is a major driver of overtreatment. We have not addressed consumer moral hazard in our model. It is,

[17] Consumers, however, seem to disapprove of cost-control bonuses (Gallagher et al., 2001).

[18] For more on this so-called *ANOVA method* see, e.g., Roth and Stahel (2005).

however, obvious that our treatment quota also curbs excessive treatment demanded by fully insured consumers. An appropriately designed treatment quota, therefore, not only curbs sellers' incentives to overtreat, but at the same time also curbs patients' incentives to overconsume. Consumer-based instruments like copayments solve the consumer moral-hazard problem but are not effective in dealing with physician-induced demand.

A last remark concerns the relation of our results here to our earlier results where the market solves the expert problem without quantity restrictions. In Emons (1997) experts set prices after they have chosen capacity. If capacity falls short of demand, experts charge high equal-compensation prices and make positive profits; if capacity exceeds demand, prices are zero and experts suffer losses. In either case experts provide honest services. The experts' capacity choices (entry decisions) are mixed so that on average they break even. The important difference to the paper at hand is thus when prices are actually set. Whereas in Emons (1997) prices adjust to the capacity and demand realizations and induce proper behavior, this is not possible in the current setup where prices are set beforehand. This is typically the case in health-care markets.[19]

If in our scenario a patient pays the doctor himself, he has only one observation of her treatment policy. Quantity restrictions are not feasible, and we encounter fraudulent behavior. We then use the special feature of health-care markets, namely that patients have insurance and doctors are reimbursed by the insurance companies. The insurance companies get information about the medical doctor's overall treatment behavior that enables them to employ quantity restrictions.[20]

5 Conclusions

The purpose of this paper is to develop incentive-compatible reimbursement schemes for physicians. We have chosen a framework where, due to the physicians' fixed capacity levels, both the problem of undertreatment and that of overtreatment arise with positive probability. Simple fee-for-service schemes do not solve the incentive problems. Either physicians with excess capacity or physicians with excess demand have the wrong incentives.

We then use the fact that the insurer observes a physician's actions for the entire set of his policy holders. This allows the insurer to set a quota that states the maximum fraction of diagnosed patients for whom he actually provides the treatment. If the insurer sets this quota equal to the fraction of patients in need of treatment, he curbs overtreatment. Therefore, only the undertreatment problem remains, which is solved by prices making diagnosis not more attractive than treatment.

[19] In Emons (2001) the monopolistic expert chooses prices and capacity together, so that the issue of over- or undercapacity does not arise in equilibrium.

[20] Our approach thus seems to be applicable to other markets where insurers reimburse experts, such as, e.g., the market for legal services.

References

Allard, M., I. Jelovac, and P. Léger (2011), "Treatment and Referral Decisions under Different Physician Payment Mechanisms," *Journal of Health Economics*, 30(5), 880–893.

Brekke, K. R., R. Nuscheler, and O. R. Straume (2007), "Gatekeeping in Health Care," *Journal of Health Economics*, 26(1), 149–170.

Brownlee, S. (2007), *Overtreated: Why too Much Medicine is Making us Sicker and Poorer*, Bloomsbury, New York.

Calcott, P. (1999), "Demand Inducement as Cheap Talk," *Health Economics*, 8(8), 721–733.

Darby, M. R., and E. Karni (1973), "Free Competition and the Optimal Amount of Fraud," *The Journal of Law & Economics*, 16(1), 67–88.

De Jaegher, K., and M. Jegers (2000), "A Model of Physician Behaviour with Demand Inducement," *Journal of Health Economics*, 19(2), 231–258.

— and — (2001), "The Physician–Patient Relationship as a Game of Strategic Information Transmission," *Health Economics*, 10(7), 651–668.

Domenighetti, G., A. Casabianca, F. Gutzwiller, and S. Martinoli (1993), "Revisiting the Most Informed Consumer of Surgical Services," *International Journal of Technology Assessment in Health Care*, 9(04), 505–513.

Dranove, D. (1988), "Demand Inducement and the Physician/Patient Relationship," *Economic Inquiry*, 26(2), 281–298.

Dulleck, U., and R. Kerschbamer (2006), "On Doctors, Mechanics, and Computer Specialists: The Economics of Credence Goods," *Journal of Economic Literature*, 44(1), 5–42.

Ely, J. C., and J. Välimäki (2003), "Bad Reputation," *The Quarterly Journal of Economics*, 118(3), 785–814.

Emons, W. (1997), "Credence Goods and Fraudulent Experts," *The RAND Journal of Economics*, 28(1), 107–119.

— (2001), "Credence Goods Monopolists," *International Journal of Industrial Organization*, 19(1), 375–389.

Farley, P. J. (1986), "Theories of the Price and Quantity of Physician Services: A Synthesis and Critique," *Journal of Health Economics*, 5(4), 315–333.

Gallagher, T. H., R. F. St. Peter, M. Chesney, and B. Lo (2001), "Patients Attitudes toward Cost Control Bonuses for Managed Care Physicians," *Health Affairs*, 20(2), 186–192.

Gruber, J., J. Kim, and D. Mayzlin (1999), "Physician Fees and Procedure Intensity: The Case of Cesarian Delivery," *Journal of Health Economics*, 18(4), 473–490.

Grumbach, K., D. Osmond, K. Vranizan, D. Jaffe, and A. B. Bindman (1998), "Primary Care Physicians Experience of Financial Incentives in Managed-Care Systems," *The New England Journal of Medicine*, 339(21), 1516–1521.

Jürges, H. (2009), "Health Insurance Status and Physician Behavior in Germany," *Schmollers Jahrbuch (Journal of Applied Social Science Studies)*, 129(2), 297–308.

Judd, K. L. (1985), "The Law of Large Numbers with a Continuum of IID Random Variables," *Journal of Economic Theory*, 35(1), 19–25.

Labelle, R., G. Stoddart, and T. Rice (1994), "A Re-Examination of the Meaning and Importance of Supplier-Induced Demand," *Journal of Health Economics*, 13(3), 347–368.

Marty, F. E. (1998), "Capacity as a Determinant of the Supply for Physicians' Services," Discussion Paper, University of Bern, http://staff.vwi.unibe.ch/emons/downloads/phy_0598_eng.pdf, last access August 1, 2013.

McGuire, T. G. (2000), "Physician Agency," in: A. J. Culyer and J. P. Newhouse (eds.), *Handbook of Health Economics, Vol. 1*, Elsevier Science B.V., Amsterdam, pp. 461–536.

Reilly, B. M., and A. T. Evans (2009), "Much ado about (doing) Nothing," *Annals of Internal Medicine*, 150(4), 270–271.

Richardson, H. (1999), "The Credence Good Problem and the Organization of Health Care Markets," Discussion Paper, Texas A&M University.

Robinson J. C. (2001), "Theory and Practice in the Design of Physician Payment Incentives," *The Milbank Quarterly*, 79(2), 149–177.

Roth, H.-R., and W. Stahel (2005), "Die ANOVA-Methode zur Prüfung der Wirtschaftlichkeit von Leistungserbringern nach Artikel 56 KVG: Gutachten zu Handen santésuisse," http://www.physicianprofiling.ch/CONGutachtenAN VADrRoth.pdf, last access August 1, 2013.

Shaw, B. (1922), *The Doctor's Dilemma: A Tragedy*, http://archive.org/stream/doctorsdilemmatr00shawuoft#page/n5/mode/2up.

Sülzle, K., and A. Wambach (2005), "Insurance in a Market for Credence Goods," *The Journal of Risk & Insurance*, 72(1), 159–176.

Twain, M. (1885), *Adventures of Huckleberry Finn (Tom Sawyers Comrade)*, http://archive.org/stream/adventureshuckle00twaiiala#page/n9/mode/2up%7C.

Wennberg, J. E., B. A. Barnes, and M. Zubkoff (1982), "Professional Uncertainty and the Problem of Supplier-Induced Demand," *Social Science & Medicine*, 16(7), 811–824.

Winand Emons
Departement Volkswirtschaftslehre
Universität Bern
Schanzeneckstrasse 1
Postfach 8573
3001 Bern
Switzerland
winand.emons@vwi.unibe.ch

Yes Men in Tournaments

by

Jason G. Cummins and Ingmar Nyman*

Received March 28, 2008; in revised form December 21, 2012;
accepted April 29, 2013

We study a rank-order tournament in which employees acquire and use private information for an investment decision. In this environment, competition can turn employees into *yes men* who make investment decisions that excessively agree with preconceived notions. The specter of yes-man behavior may drive the tournament incentive intensity and the employees' information-collection effort either to zero or above the first-best efficient levels. We also show that yes-man problems are alleviated by a stronger correlation between the employees' sources of uncertainty and by the use of individual compensation contracts rather than a tournament. (JEL: D82, J33, M51)

1 Introduction

Firms often use competition for promotion or other rewards to motivate their employees. To answer the question how this incentive mechanism shapes the way employees use information, we study a rank-order tournament model in which two employees choose how much effort to expend to collect private information about a choice between two investment projects. The employees' supervisor then receives additional, but noisy, information about the two chosen investment projects and, based on her assessment of their profitability, subsequently rewards one of the employees. We reach a number of conclusions. First, the competitive pressure of a tournament can turn employees into *yes men* who make decisions that excessively

* Jason Cummins: Brevan Howard Asset Management; Ingmar Nyman (corresponding author): Department of Economics, Hunter College, CUNY, and Department of Economics, The Graduate Center, CUNY. We are grateful for feedback from Jonathan Conning, Partha Deb, Ken McLaughlin, Matt Baker, who also helped us with the figures, and seminar audiences at Hunter College, the Stockholm School of Economics, and the "Tournaments, Contests, and Relative Performance Evaluation" conference at NCSU in March 2008. The suggestions of three anonymous referees greatly improved the paper. Nyman thanks SITE at the SSE for its kind hospitality.

DOI: 10.1628/093245613X13806312325779

agree with the prevailing prior opinion. This inefficient behavior arises because it is easier to convince the supervisor that your choice is a good one if you have her prior beliefs working for you rather than against you. On the other hand, a countervailing incentive arises because, everything else equal, a more profitable investment project is more likely to generate a signal for the supervisor to that effect, especially if the employee's and the supervisor's signals are both accurate. Therefore, yes-man behavior occurs when the supervisor's additional information is noisy compared to the prior information or when a high marginal cost of collecting information prevents the employees from creating an accurate signal for themselves. Second, we find that the specter of yes-man behavior may alter the design of the tournament and the supply of effort by the employees. In general, when employees become yes men, the information that they collect is wasted, the firm therefore chooses to abandon the tournament, and the employees supply no effort. But when more accurate information for the employees could induce them to use it efficiently, then the firm may find it profitable to implement this outcome by increasing the incentive intensity in the tournament – and with it the effort supplied by the employees – above and beyond the first-best efficient level. Third, we show that the yes-man problem is more severe the weaker is the correlation between the uncertainty of the employees' decisions. Hence, competition among employees that perform similar tasks is desirable. Finally, we demonstrate that when it comes to discouraging yes-man behavior, individual wage contracts are superior to a tournament. The reason is that while a tournament bases compensation only on whether you win or lose, an individual contract can base it on the margin of victory or defeat as well. This allows individual contracts to recognize when you lose by not catering to prior opinion and to increase your compensation accordingly.

Our analysis takes as given the use of a rank-order tournament, even though it may not be the theoretically optimal incentive mechanism. The reason for this is empirical: in light of the prevalence of tournaments in real life, it is interesting and important to study how they affect the use of private information. Needless to say, the topic of whether the yes-man problem that we highlight can be eliminated through the design of alternative mechanisms more generally is of great interest. However, we leave that to be addressed in future research.

The literature on rank-order tournaments is vast (see Gibbons, 1997, Gibbons and Waldman, 1999, McLaughlin, 1988, and Prendergast, 1999, for surveys). Early contributions by Lazear and Rosen (1981) and Nalebuff and Stiglitz (1983) confirmed that rank-order tournaments can implement a first-best efficient outcome under risk neutrality, but that the tension between incentive provision and risk sharing prevents first-best efficiency when employees are risk-averse. This risk-exposure cost of incentive provision is something that a tournament has in common with the main alternative incentive mechanism, individual compensation contracts. But a tournament also has specific features that typically are absent, or at least less prominent, in individual contracts. The most striking of these is that a tournament uses only relative performance as a basis for the distribution of rewards. This can lower the risk-sharing cost of incentive provision by filtering out common risk, especially when employees have highly correlated uncertainty (Holmström, 1982).[1] But this exclusive reliance on relative performance also has the drawback of discouraging cooperation among employees. The reason is, of course, that the tournament can be won not only by improving one's own performance, but also by compromising that of one's competitors (Drago and Garvey, 1998, Dye, 1984, and Lazear, 1989).

The interplay between information asymmetries, on the one hand, and promotion and task assignment, on the other, has been a popular topic of inquiry (see, for example, Bernhardt, 1995, Koch and Peyrache, 2005, Ricart i Costa, 1988, Waldman, 1984, and Zábojník and Bernhardt, 2001). Most of the research in this area has focused on precontractual information asymmetries across the boundary of the f rm: typically, the current employer knows more about a worker's ability than potential alternative employers. We study instead postcontractual information asymmetries inside the f rm, between the employee and his current employer. Zábojník (2002) does consider a situation that is similar to the one we have in mind. However, while he is interested in the effect of private information on incentives, we investigate the effect of a particular type of incentives on the use of private information.

This paper builds on our previous research that demonstrates that competitive pressure makes it costly to act with integrity and therefore can breed ineffic ent conformity (Cummins and Nyman, 2005). Here, we apply this insight to the tournament setting and extend our previous analysis in several important directions. Our work also naturally complements Prendergast (1993b) and Gentzkow and Shapiro (2006), who show, respectively, that individual compensation contracts and reputation concerns also can create yes men.

The rest of the paper is organized as follows. In the next section, we introduce the model. We then demonstrate that when information is symmetric, the equilibrium outcome is eff cient. This serves as a benchmark for the main analysis, in which we assume that employees have private information about their decisions. Finally, we investigate how the correlation of the employees' investment uncertainty affects their incentive to use information eff ciently and how the yes-man problems in a tournament compare to those under individual compensation contracts. A short discussion of our f ndings concludes. The appendix contains the proofs of the propositions.

2 *Model Setup*

The model extends the classic rank-order tournament setup in Lazear and Rosen (1981) with a more intricate production decision in the form of an investment

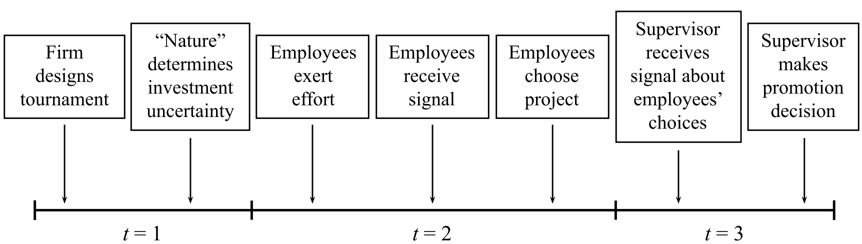
[1] Other reasons for the use of tournaments have also been proposed. First, they place modest demands on the ability to evaluate employee performance. Second, they may solve the problem of assigning workers to jobs (Baker, Jensen, and Murphy, 1988, Fairburn and Malcomson, 2001, and Prendergast, 1993a). Third, they can discourage ineff cient influenc activities (Fairburn and Malcomson, 2001). Finally, with its fi ed total reward, a tournament removes the benefi to the f rm of denying any one employee the reward she has been promised (Bhattacharya, 1983, Carmichael, 1983, and Malcomson, 1984).

problem, which is based on Brandenburger and Polak (1996). There are two ex ante identical employees, indexed by i and j, and one supervisor.[2] Each employee is charged with the task of picking one of two mutually exclusive uncertain alternatives, which we will think of as investment projects. To guide this choice, the employees exert effort to collect information. Their incentive to do so is provided by a rank-order tournament.

2.1 *Dynamic Structure*

There are three time periods. In the fir t period, the fir designs the tournament and Nature determines the state of the world for the investment uncertainty. In the second period, the employees f rst simultaneously decide how much costly effort to put into learning about the prof tability of their projects. The effort then generates information in the form of one signal per employee, and based on this signal each employee simultaneously chooses one of his two projects. In the third period, the supervisor receives signals about the profitabilit of the choices that the employees have made, one for each project, and subsequently allocates two tournament prizes, which we will think of as being promoted or not, to the employees.

Figure 1
Dynamic Structure of the Model



2.2 *The Investment Decision*

The employees face the identical problem in terms of project choice. The two projects are referred to as the *green* and the *red*. Employee i's project choice is denoted by $z_i \in \{g, r\}$. There are two possible states of the world that determine the prof t from the projects, the *Green* and the *Red.* This uncertainty will be denoted by $Z \in \{G, R\}$. A project generates a higher profi (net of investment) if it matches the state. For simplicity, both projects have the same high and low payoff realizations, normalized to unity and zero, respectively. The profi of employee i's project, π_i, is thus equal to

$$\pi_i(z_i, Z) = \begin{cases} 1 & \text{if } (g, G) \text{ or } (r, R)\,, \\ 0 & \text{if } (g, R) \text{ or } (r, G)\,. \end{cases}$$

[2] The employees will be male and the supervisor female.

The prior belief about the uncertainty, which is shared by everyone and is common knowledge, weakly favors the Green state: $\Pr(Z = G) = \mu \in [1/2, 1)$.

2.3 *The Effort Decision*

The employees can acquire additional information, in the form of a signal, to guide their choice of project. The signals are conditionally independent across employees. Employee i's signal is denoted by $e_i \in \{\gamma, \rho\}$, with γ and ρ indicating, with equal accuracy, that the state is Green and Red, respectively. The signal accuracy is denoted by $\varepsilon_i = \Pr(e_i = \gamma \mid G) = \Pr(e_i = \rho \mid R)$.

The employees' information is not free, but requires effort. Employee i's information collection effort is denoted by λ_i and amounts to the information content in his signal above and beyond the public information: $\lambda_i = \varepsilon_i - \mu$. Therefore, employee i's choice set when it comes to information collection is $\lambda_i \in [0, 1 - \mu]$. To create a conf ict of interest between employee and employer, effort is costly. Both employees have the same cost of effort, which for expositional simplicity is assumed to be quadratic: $C(\lambda_i) = 0.5c\lambda_i^2$, where c is the parameterized slope of the marginal cost of effort, $MC(\lambda_i) = C'(\lambda_i) = c\lambda_i$.

2.4 *The Promotion Decision*

The employees' incentive to invest in effort is provided by a rank-order tournament. For concreteness, we will think of the winning prize in the tournament as a promotion, but it could be any type of reward.[3] In period 3, the supervisor promotes one of the employees based, in part, on their choice of project. For both employees, the value of receiving and being denied the promotion is equal to W_1 and W_2, respectively. Promotion is desirable, so $W_1 > W_2$.

The supervisor evaluates each employee's project choice by her expectation of its profitabilit , which represents a summary measure of how the employee has managed this task on the shareholders' behalf. It is denoted by $\hat{\pi}_i = \mathrm{E}[\pi_i \mid$ Supervisor's information]. The information upon which the supervisor forms her posterior beliefs about the expected profi for each project contains (a) the prior information, (b) whatever she knows – depending on the informational assumptions – about the employees' two signals, and (c) two conditionally independent signals that indicate the profitabilit of each of the chosen projects. The signal about employee i's project can be either high or low and is denoted by $s_i \in \{h, l\}$. A high (low) signal points to the prof t from the project being high (low). Both signal realizations are equally informative, and their accuracy is parameterized by

3 In this paper, we focus on the tournament as an incentive device. Another important role that it can also serve is as a means of selecting employees based on talent. By making the simplifying assumption that the employees are identical in talent, we abstract away from this potential benef t of a tournament. In the end, the cost of using a tournament that we point out would have to be traded off against the benefit that we have left out in order to make the analysis more transparent.

$\sigma = \Pr(s_i = h \mid \pi_i = 1) = \Pr(s_i = l \mid \pi_i = 0) \in (1/2, 1)$.[4] The supervisor's signal can be thought of as an early indication of the success of a project, perhaps in the form of an early cash f ow, which it would be natural to assume is positively – but imperfectly – correlated with the total cash f ow.

The evaluation measure upon which the promotion is based is also subject to uncertainty unrelated to the investment decision, due, for example, to uncertainty about the supervisor's state of mind or about the evaluation of other decisions, past and future, for which the employees are held responsible at the time of promotion. Hence, the evaluation measure for employee i is $V_i = \hat{\pi}_i + \eta_i$ where η_i is a random variable with an expected value equal to zero. η_i and η_j are mutually iid as well as independent of the investment uncertainty, Z.

2.5 *Strategies and Objectives*

All the players in the game are risk-neutral and update their information using Bayes's rule whenever possible. The supervisor has only one decision to make and simply carries out the standard promotion rule in a tournament. Hence, she promotes employee i (employee j) if $V_i > V_j$ ($V_i < V_j$), and each employee with probability one-half if $V_i = V_j$.

Each employee makes two decisions: information collection effort and project choice. Since project choice is conditioned on the preceding effort choices as well as on the realization of his signal, each employee has three distinct decisions to make: effort choice, investment with a γ-signal, and investment with a ρ-signal.[5] Hence, employee i's strategy specif es these three choices: $(\lambda_i, z_i(\lambda_i, \lambda_j, e_i = \gamma), z_i(\lambda_i, \lambda_j, e_i = \rho))$.

The employees are risk-neutral, and the outcome of the promotion decision is the only compensation they receive. They therefore choose their strategy with the objective of maximizing the expected value of their net utility, consisting of the value of the outcome of the promotion decision net of the cost of effort:

$$\mathrm{E}[U_i(\cdot)] = W_2 + \mathrm{E}[P_i(\cdot)][W_1 - W_2] - C(\lambda_i)\,,$$

where $P_i(\cdot)$ is the probability that employee i gets promoted,

$$\begin{aligned} P_i(\cdot) &= \Pr(V_i(\cdot) > V_j(\cdot)) \\ &= \Pr(\hat{\pi}_i(\cdot) + \eta_i > \hat{\pi}_j(\cdot) + \eta_j) \\ &= \Pr(\hat{\pi}_i(\cdot) - \hat{\pi}_j(\cdot) > \eta_j - \eta_i)\,. \end{aligned}$$

Hence, in order to result in promotion, the supervisor's impression of employee i's management of the investment project must at least offset whatever advantage employee j may have due to other aspects of the performance evaluation – including sheer luck – which we will denote by $\xi = \eta_j - \eta_i$.

4 Notice that these signals provide information only about whether the state of the world is Green or Red.

5 Since the employees make their investment decisions simultaneously, each knows his own signal realization, but not that of the other employee.

$P_i(\cdot)$ depends upon two independent sources of uncertainty. First, for any given level of $\hat{\pi}_i - \hat{\pi}_j$, we have $P_i(\cdot) = \Pr(\xi < \hat{\pi}_i - \hat{\pi}_j) = F_\xi(\hat{\pi}_i - \hat{\pi}_j)$, where $F_\xi(\cdot)$ is the distribution function of ξ. Second, in contrast to a standard tournament model, for a given distribution of ξ the threshold prof tability advantage $[\hat{\pi}_i(\cdot) - \hat{\pi}_j(\cdot)]$ – and with it $P_i(\cdot)$ – is random. This is because both $\hat{\pi}_i$ and $\hat{\pi}_j$ are determined by the realizations of the supervisor's signals, which in turn depend on the investment uncertainty and are unknown to the employees when they make their decisions. It greatly simplif es the analysis if the employees can make their decisions with the objective of maximizing the expected value of the prof tability of their project in the eyes of the supervisor. This requires that $\mathrm{E}[P_i(\hat{\pi}_i, \hat{\pi}_j)] = \mathrm{E}[F_\xi(\hat{\pi}_i - \hat{\pi}_j)] = F_\xi(\mathrm{E}[\hat{\pi}_i - \hat{\pi}_j]) = F_\xi(\mathrm{E}[\hat{\pi}_i] - \mathrm{E}[\hat{\pi}_j])$, which, in turn, holds only if $F_\xi(\cdot)$ is linear in $\hat{\pi}_i - \hat{\pi}_j$. Therefore, in addition to the standard restriction in the literature that $F_\xi(\cdot)$ is such that a unique symmetric equilibrium choice in effort exists, we assume specific ally that on the interval $[-1/2, 1/2]$, ξ is uniformly distributed with $dF_\xi(x)/dx = \phi$, where ϕ is the probability that ξ falls between $-1/2$ and $1/2$. We believe, however, that our qualitative conclusions would stand if this assumption were dropped so that the employees were forced to make decisions using the more cumbersome evaluation criterion of f rst-order stochastic dominance of the distribution of outcomes of V_i.

Finally, before the employment relationship starts, the f rm and the employees can credibly commit to an agreement about the rules for the promotion tournament. In this negotiation, the f rm has all the bargaining power, except that each employee has access to employment elsewhere that yields a net utility that is normalized to zero. Therefore, before the employees make any decision, the fir can unilaterally choose W_1 and W_2 to maximize its expected prof t, subject to the participation constraints that it must afford each employee an expected net utility that is nonnegative.[6]

2.6 Information Structure

The structure of the game is common knowledge. All actions – the choice of tournament payoffs by the f rm, the choice of effort by the employees, the choice of investment project by the employees, and the choice of employee to promote by the supervisor – are publicly observable. However, even though they are observable, the employees' effort levels are not verifia le. This is to prevent the fir from basing compensation on employee input, thus creating a nontrivial incentive problem.[7] The processes that generate information – the employees' signals, the supervisor's

[6] To keep the incentive problem simple and allow for a f rst-best efficien benchmark, we rule out the two most often studied obstacles to incentive provision: risk exposure, and limited-liability or wealth constraints. Hence, in addition to risk neutrality, we also assume that there is no lower bound on the losing prize W_2.

[7] The assumption that effort is observable is unimportant, because even if it were not directly observable, as a rational choice it would be unambiguously determined by the design of the tournament.

signals, and the exogenous uncertainty in the supervisor's evaluation measures for the employees, η_i and η_j – are all common knowledge. In the main analysis, the realization of each employee's signal is private information and therefore unobservable to everyone else. The realizations of the supervisor's information – her signals about the employees' investments as well as the exogenous uncertainty in the evaluation measures – are public information. As a consequence, the realizations of the performance evaluation of each employee, V_i and V_j, are also public information.

With regard to terminology for the information structure, we will use "information" to refer to conditioning information and/or prior information. We will use "belief" or "opinion" to refer to the probability distribution that an agent takes away from a certain situation, be it a prior probability in the absence of updating or a posterior, conditional probability.

2.7 Eff ciency

The normalization of the prof t realizations to zero and one, respectively, makes the expected prof t of a project equal to the probability that it matches the state. Choosing the project that the employee's signal indicates is prof table maximizes the expected prof t from investment. Assuming that an employee makes full use of his information by following this eff cient investment strategy, the expected prof t of his investment is equal to ε_i. Since the employee's effort, $\lambda_i = \varepsilon_i - \mu$, increases this expected prof t one for one, its marginal social benef t is unity.

A key feature of the model is that the amount of information that the employee has access to is an endogenous choice. An interior first-bes eff cient choice of effort, which we will denote by $\lambda^{**} \equiv \varepsilon^{**} - \mu$, equates the marginal social benef t to the marginal social cost of effort: $1 = c\lambda^{**} \Leftrightarrow \lambda^{**} = 1/c$. The highest effort level that warrants its cost, i.e., the highest effort level that dominates putting in no effort at all, will be denoted by $\bar{\lambda}$: hence $\bar{\lambda} - C(\bar{\lambda}) = 0 \Leftrightarrow \bar{\lambda} - 0.5c\bar{\lambda}^2 = 0 \Leftrightarrow \bar{\lambda} = 2(1/c)$. Since it is strictly smaller than $\bar{\lambda}$, λ^{**} strictly dominates no effort at all, making it the unique socially optimal level of effort for both employees. To ensure that full information is not feasible we assume that $\bar{\lambda} < 1 - \mu \Leftrightarrow c > 2(1/(1-\mu))$.

Finally, the value of the employees' information also depends on what the supervisor can do with the information that she receives in period 3. If the supervisor could overrule the employees' choices, then with suff ciently accurate supervisor information it would be diff cult to explain why the f rm would need the employees in the f rst place. To preclude this degenerate outcome, we assume that the employees' project choices are irreversible. Hence, the supervisor can use her information only to evaluate – not to change – the employees' decisions. An alternative approach would be to assume that the supervisor's information is suffic ently inaccurate. However, making investment irreversible instead has the advantage of imposing fewer restrictions on the information structure, thus allowing for a more thorough analysis.

Table 1 provides a list of our notation.

Table 1
List of Symbols

$z_i \in \{g, r\}$	Employee i's investment choice, either green or red.
$Z \in \{G, R\}$	The investment uncertainty, either Green or Red.
$\pi_i(z_i, Z)$	The net payoff from employee i's investment.
$\widehat{\pi}_i(\cdot)$	The supervisor's expectation of π_i.
$\mu \geq 1/2$	Public information: the prior probability that $Z = G$.
$e_i \in \{\gamma, \rho\}$	Employee i's signal about Z, either γ (indicating G) or ρ (indicating R).
$\varepsilon_i = \mu + \lambda_i$	The accuracy of employee i's signal.
$\lambda_i \in [0, 1 - \mu]$	Employee i's effort to collect information about Z.
$C(\lambda_i) = (1/2)c\lambda_i^2$	The employees' cost of effort, parameterized by $c > 2(1/(1 - \mu))$.
$MC(\lambda_i) = c\lambda_i$	The employees' marginal cost of effort.
$\lambda^{**} = 1/c$	The f rst-best efficien amount of employee effort.
$\overline{\lambda} = 2(1/c)$	The largest profitabl amount of employee effort.
$s_i \in \{h, l\}$	The supervisor's signal about the net payoff from employee i's investment project.
$\sigma \geq 1/2$	The accuracy of the supervisor's signals.
W_1, W_2	The employees' payoffs from receiving and being denied the promotion, respectively.
$V_i = \hat{\pi}_i + \eta_i$	The supervisor's evaluation measure for employee i.
η_i	Uncertainty about the evaluation of employee i unrelated to her investment.
$\xi = \eta_j - \eta_i$	Employee j's advantage over employee i due to factors other than their investments.
P_i	The probability that employee i gets promoted.
ϕ	The probability that ξ falls between $-1/2$ and $1/2$.

3 Analysis

When studying the outcomes of the model, we will focus on symmetric pure-strategy sequential equilibria refined by universal divinity. We restrict ourselves to pure strategies because the use of mixed strategies by the supervisor is incompatible with the rules of the tournament. Sequential equilibrium with universal divinity ensures that players at every information set make a rational choice given their beliefs and that those beliefs in turn are rational in the sense that they are consistent with the strategies that are used. Hence, at information sets along the equilibrium path of play, Bayes's rule specifies beliefs. But at information sets off the equilibrium path of play, Bayesian updating cannot define rational beliefs, so we use universal divinity as proposed by Banks and Sobel (1987) to ensure that only economically sensible beliefs are allowed.[8] As for symmetry, looking at behavior that is the same for both employees is sensible for a number of reasons. First, symmetric behavior is intuitively natural, since the employees are identical. Second, studying symmetric equilibria is standard in the tournament literature. Third, symmetric equilibria constitute a natural focal point in a symmetric model because they require no additional coordination across players. Finally, and perhaps most importantly, we are interested in the sustainability of efficient behavior by the employees, and since such behavior is symmetric, the equilibria that we study should be as well (Marino and Zábojník, 2004).

3.1 Equilibrium when the Supervisor Knows the Employees' Signal Realizations

We start out by analyzing the model when the employees' signal realizations are known to the supervisor. This will serve as a benchmark for the more likely situation in which the supervisor is ignorant about the outcome of the employees' information collection. The unique equilibrium, which is summarized in Proposition 1, replicates the standard conclusion in the tournament literature that a first-best efficient outcome can be implemented.

Working our way backwards from the end of the game, we first look at the supervisor's promotion decision. In doing so, we take as given the employees' symmetric effort level, $\lambda_i = \lambda_j = \lambda$, and the resulting signal accuracy, $\varepsilon_i = \varepsilon_j = \varepsilon$, as well as their choice of investment projects, z_i and z_j. The supervisor promotes the employee with the higher evaluation measure, $V_i = \hat{\pi}_i + \eta_i$, and her assessments of the profitability of the projects chosen are based on the employees' information-collection effort, the realizations of their signals, their choice of investment projects, and the realizations of the supervisor's own signals: $\hat{\pi}_i = \mathrm{E}[\pi_i(z_i, Z) \mid \lambda_i, \lambda_j, e_i, e_j, z_i, z_j, s_i, s_j]$.

---

[8] Universal divinity specifies that a deviation from equilibrium behavior that takes the game to an information set off the equilibrium path of play must be interpreted as having not been made by a type of informed player (in our case an employee with a certain signal realization) that would find it beneficial to do so in the face of a smaller set of strategies chosen by the uninformed player (in our case the supervisor).

Moving to the employees' choices of investment project, employee i will choose the project that maximizes his probability of getting promoted. Unable to affect the evaluation measure of employee j, this amounts to maximizing the supervisor's assessment of the profitability of his own project, $\hat{\pi}_i$ (this gives the highest expected value of his evaluation measure, $V_i = \hat{\pi}_i + \eta_i$, as the exogenous uncertainty is by definition beyond his control). We can think of employee i's choice of projects as having two effects on $\hat{\pi}_i$, before and after the supervisor's signals. If employee i makes the inefficient choice of picking the project that his signal suggests is the less profitable one, both of these effects will decrease $\hat{\pi}_i$. First, knowing the employees' signals, the supervisor will recognize that he chooses the project that is less profitable. Second, the less profitable project is less likely to generate a favorable signal for the supervisor. Hence, following their own signals when choosing project is rational for the employees.

Finally, consider the employees' choice of effort and the firm's choice of tournament prizes. The employees will exert effort, λ_i^*, until its expected marginal benefit to them, $\phi[W_1 - W_2]$, is equal to its marginal cost, $c\lambda_i$: $\phi[W_1 - W_2] = c\lambda_i^* \Leftrightarrow \lambda_i^* = (\phi/c)[W_1 - W_2]$. Because it is used in full, the firm's marginal benefit of employee information and effort is equal to unity. It therefore chooses a tournament design that makes the employees choose the efficient effort level to support their investment decisions, by setting the employees' net benefit from promotion equal to $W_1^* - W_2^* = (c\lambda_i^{**})/\phi = 1/\phi$. We think of this as the benchmark first-best efficient incentive intensity in the tournament.

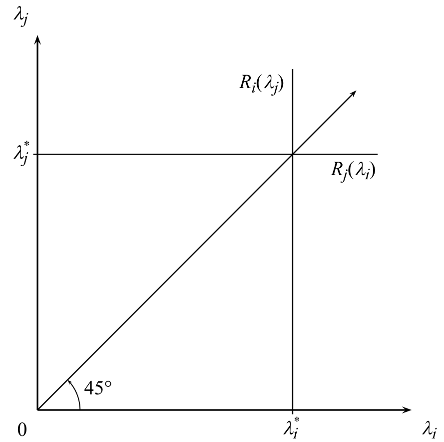
The symmetric equilibrium in effort levels is illustrated in Figure 2, where R_i and R_j denote the reaction functions of employee i and j, respectively. Notice that employee i's optimal choice of effort is independent of that of employee j. This absence of strategic interaction is due to the assumption that ξ is uniformly distributed.[9]

PROPOSITION 1 *When the supervisor knows the realizations of the employees' signals, the equilibrium is unique with the employees choosing the same effort level, using their private information fully, and each getting promoted with probability one-half. The firm chooses the tournament prizes that induce the employees to choose the efficient level of effort while affording them an expected net utility that is equal to zero:*

$$W_1^* = \frac{1}{2}\left(\frac{1}{c} + \frac{1}{\phi}\right) \quad \text{and} \quad W_2^* = \frac{1}{2}\left(\frac{1}{c} - \frac{1}{\phi}\right).$$

[9] With less stringent assumptions on the distribution of the difference in random shocks to performance, $\xi = \eta_j - \eta_i$, there could be several equilibria in effort levels. It is, however, standard in the tournament literature to assume a shock distribution that eliminates all equilibria save one that is symmetric. By contrast, asymmetric equilibria are typical and intuitively appealing in many contests where the probability of winning has a general logit form or a difference form (in a tournament it has a probit form) (Baik, 1998, and Hirshleifer, 1989). Examples include R&D races, military conflict and rent-seeking contests.

Figure 2
Equilibrium Choice of Effort when the Supervisor Knows
the Employees' Signal Realizations



3.2 Equilibrium when the Supervisor does Not Know the Employees' Signal Realizations

The more realistic information structure is that the employees' signal realizations are private information, so that the supervisor can discern what they know only through their choice of project. This allows the employees to manipulate the supervisor's beliefs about which signal they have received. Their attempts to do so may, in turn, distort their choice of investment project.

3.2.1 Promotion

Starting from the end of the game, the supervisor's decision is trivial: she promotes the employee with the higher performance evaluation measure, $V_i = \hat{\pi}_i + \eta_i$. The crucial difference from the symmetric-information case is that now the supervisor does not know the employees' signal realizations, only their choices of project: $\hat{\pi}_i = \mathrm{E}[\pi_i(z_i, Z) \mid \lambda_i, \lambda_j, z_i, z_j, s_i, s_j]$. However, the supervisor will use these choices to try to infer the underlying private information. How successful she is depends on the strategies that the employees use to choose their projects.

One possibility is that the employees' investment behavior is perfectly pooling. This happens if they ignore their private signals by always choosing the same project, either the green or the red. Since behavior does not depend on the signals, the supervisor is unable to infer anything about them: an employee who chooses the common project can have either a γ-signal or a ρ-signal.[10] As a consequence, the

[10] A choice of project that is never supposed to happen cannot be evaluated using Bayes's rule. Here we specify that the supervisor views this behavior in a way that is

employees' investment choice does not change the supervisor's prior belief about the investment uncertainty, i.e., she still believes that $\Pr(G \mid \lambda_i, \lambda_j, z_i, z_j) = \mu$.

With pure strategies, the only other possibility is that investment behavior is perfectly separating. This happens if the employees follow their private signal, choosing the green project if the signal is γ and the red project if the signal is ρ (it also happens if the employees choose the project that their signal indicates is less prof table, but these strategies are strictly dominated and can therefore be disregarded). Since it is determined by the signal realization, this behavior allows the supervisor to infer the underlying employee signals perfectly: an employee who chooses the green project has a γ-signal, and one who chooses the red project has a ρ-signal. This means that when the supervisor receives her own signals, her belief about the investment uncertainty is $\Pr(G \mid \lambda_i, \lambda_j, z_i, z_j, e_i, e_j)$. What this belief looks like depends on the choice of investment projects that the employees actually make. If they choose different projects, then the supervisor knows that they have different signals. Since these signals are equally accurate, they cancel each other out, leaving the supervisor with her prior belief, $\Pr(G \mid \lambda_i, \lambda_j, z_i, z_j, e_i, e_j) = \mu$. On the other hand, if both employees choose the same project, then they both indicate that the same state of the world is more likely, skewing the supervisor's belief in that direction. For example, if both employees choose the green project, Bayes's rule would give the supervisor a posterior when she receives her own signal that is equal to

$$\Pr(G \mid \lambda_i, \lambda_j, z_i, z_j, e_i, e_j) = \frac{\mu\varepsilon^2}{\mu\varepsilon^2 + (1-\mu)(1-\varepsilon)^2}.$$

The f nal revision of the supervisor's belief about the investment uncertainty, again using Bayes's rule, will be based on the realizations of her own two signals. The supervisor receives one signal for each of the two investment projects, indicating whether the profitabilit of each is high or low. For example, suppose that the employees use the perfectly separating strategy of following their own signal, both employees choose the red project, and both signals received by the supervisor are low. Then the supervisor can infer that both employees have ρ-signals, giving her the following belief about the investment uncertainty when she makes her promotion decision:

$$\Pr(G \mid \lambda_i, \lambda_j, z_i, z_j, e_i, e_j, s_i, s_j) = \frac{\mu(1-\varepsilon)^2\sigma^2}{\mu(1-\varepsilon)^2\sigma^2 + (1-\mu)\varepsilon^2(1-\sigma)^2}.$$

Three things about how the supervisor's signals affect her beliefs are worth pointing out. First, the investment uncertainty is the same for both projects, which means that the signal about employee i's project does not carry more information about that project than it does about employee j's project: if both employees have

consistent with universal divinity. In the context of our model, this amounts to the supervisor interpreting a choice that is not supposed to be made as the employee who makes it following his signal: if the green project is supposed to be chosen regardless of which signal is received, then the supervisor will believe that a red project can only have been chosen by an employee with a ρ-signal, and vice versa.

chosen the red project, then a high signal for employee i's project simply indicates that the Red state is more likely and therefore that the expected prof t for both projects is higher. It does not, however, give more good news about employee i's project than about that of employee j. Second, the interpretation of the signal in terms of which state is more likely depends on the choice of project: a high signal for employee i indicates that the state is Red if he has chosen the red project, but that the state is Green if he as chosen the green project. Finally, once again, the two signals are equally accurate, and therefore if they contradict one another they cancel out, leaving the supervisor with the same belief that she had before she received the signals.

3.2.2 Investment

Moving back through the game, consider next the employees' investment decisions. Our f rst step towards understanding what these decisions look like is to answer a less ambitious question: is eff cient investment choice an equilibrium outcome? We will do this by asking ourselves whether either type of employee, with a γ-signal or a ρ-signal, has an incentive to unilaterally deviate from the eff cient investment strategy of following his own signal.[11] Throughout this analysis, keep in mind that eff cient project choice by the employees is perfectly separating and therefore reveals their signals to the supervisor.

Suppose employee i has a ρ-signal. Just as when information is symmetric, his objective is to try to get promoted by maximizing the supervisor's assessment of the prof tability of his investment project, $\hat{\pi}_i$. However, at the time when he chooses his project, this is inf uenced by three things that employee i does not know: employee j's choice of project and the realizations of the supervisor's two signals. The uncertainty facing employee i if he chooses the red project is illustrated in Figure 3.[12] If employee j also chooses the red project, represented by the top branch labeled 1, then the supervisor knows that both employees have a ρ-signal, revising her prior belief in favor of the Red state. The fact that the two employees have the same project means that the supervisor must believe that they have the same profitabilit , regardless of which signals she receives.[13] Along this branch, for all realizations of the supervisor's signals each employee can therefore expect to be promoted with probability one-half.

In the bottom half of Figure 3, the employees choose different projects. Then the supervisor knows that the employees have signals that contradict one another and therefore cancel each other out, leaving her with only the prior belief after observing

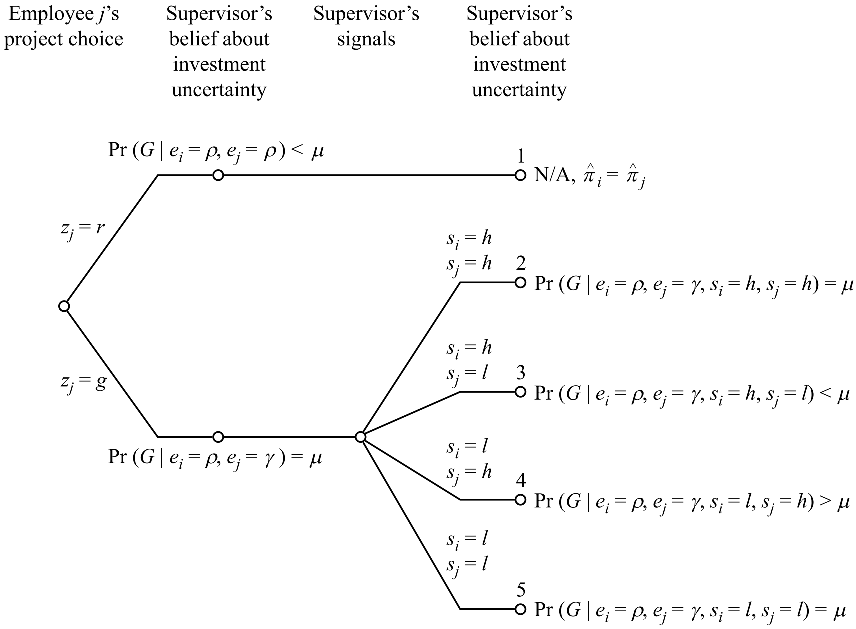
---

11 Since the game is symmetric, we need to check for only one employee.

12 The f gure if employee i instead were to choose the green project would look the same, except that the investment alternatives for employee j would be reversed.

13 If both players choose the red project, then the expected profi of both projects is equal to the probability that the state is Red. And since the supervisor's signals provide information only about which state is more likely, not about other features of the projects, if, say, $s_i = h$ and $s_j = l$, it must still be the case that the supervisor's assessment of the probability that the state is Red is the same for both projects.

Figure 3
Uncertainty Facing Employee i if he Chooses the Red Project



the employees' project choice. The supervisor can now receive four different signal realizations. In the top and bottom branches, labeled 2 and 5, respectively, the signals for both projects are the same (i.e., green is high and red is high, or both are low), in which case they indicate different states of the world and therefore cancel. This leaves the supervisor with only her prior belief to rely on. In branch 3, the red and green projects receive a high and a low signal, respectively, both of which indicate that the state is Red. Finally, branch 4 has the opposite signals, both of them indicating that the state is Green.

It is useful to f rst consider the special case of the supervisor having no a priori preference for one of the projects, i.e., $\mu = 1/2$. If both employees choose the same project – branch 1 in Figure 3 – there is a draw, and each employee receives the promotion with probability one-half. If they choose different projects but the supervisor's signal cancel out – branches 2 and 5 – then the supervisor believes that $\Pr(G) = \mu = 1/2$, so once again there is a draw and each employee is promoted with probability one-half. Finally, if different projects are chosen and the supervisor's signals both point toward the same state, then the project with the support of the signals – branch 3 for the red project, and branch 4 for the green project – has a strictly higher expected prof tability and therefore gives its employee the promotion with probability one.

Consider an employee with a ρ-signal. Is it rational for him to choose the eff cient red project, or does the green project offer him a strictly higher probability of winning the promotion? In the three situations where there is a tie – branches 1, 2, and 5 – both projects are equally good, each giving the promotion with probability one-half. In the remaining two contingencies, both projects win in one and lose in the other, depending on whether the supervisor's signals are for or against. But his ρ-signal makes the ρ-employee believe that the Red state more likely than the Green:

$$\Pr(G \mid \lambda_i, e_i = \rho) = \frac{\mu(1-\varepsilon)}{\mu(1-\varepsilon) + (1-\mu)\varepsilon} = 1 - \varepsilon < \frac{1}{2}\,.$$

This means that the ρ-employee thinks that the red project is more likely to be prof table and, as a result, also more likely to generate a high signal for the supervisor than the green project. Therefore, choosing the red project increases the chance of winning the promotion and decreases the risk of losing it. Hence, since there is no incentive to deviate from eff cient investment, with a symmetric prior it can always occur in equilibrium.[14]

But it is quite unlikely that the supervisor's prior experience leaves her completely indifferent between the two projects, i.e., that $\mu = 1/2$. Instead, one would expect that the prior probability indicates that one project – in this case and without loss of generality the green one – is more likely to be prof table than the other, i.e., that $\mu > 1/2$. This slight and generic perturbation has a dramatic effect on the outcome, because it handicaps the red project when it comes to getting employee i promoted if he follows his ρ-signal. If employee j also chooses the red project, it is still the case that there is a tie and each employee is promoted with probability one-half. But if the other employee chooses the green project, which is more likely when $\mu > 1/2$, then the red project now has a much smaller chance of winning the promotion. If the supervisor's signals cancel (branches 2 and 5), her prior belief now indicates that the Green state is more likely, so she promotes the employee with the green project. Moreover, even if her signals support the red project (branch 3), the supervisor's prior bias may be so strong that she still gives the promotion to the green project. Hence, while it wins in one outcome and reaches a draw in three when the prior is balanced, the red project loses in at least three – and possibly in four – outcomes when the prior is unbalanced. This reinforces the incentive for an employee with a γ-signal to invest eff ciently, but encourages an employee with a ρ-signal to deviate from the eff cient strategy and become a yes man by choosing the project that the prior suggests is more promising (i.e., the green one).

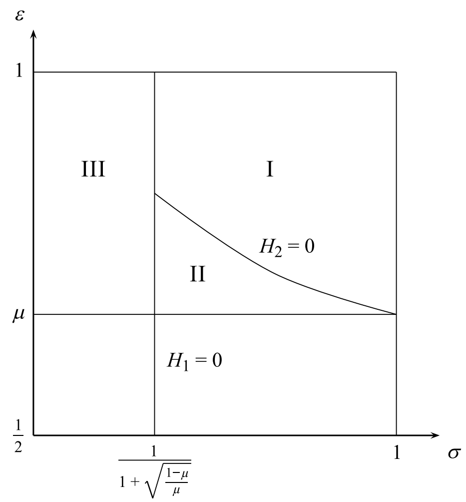
The ρ-type employee's trade-off that determines whether or not he will become a yes man is analyzed in Proposition 2 and illustrated in Figure 4. A f rst condition that must be satisf ed for the ρ-employee to choose the red project is that it has some chance of beating its green counterpart. With an unbalanced prior, the red project loses not only when the supervisor's signals favor green, but also when they cancel.

[14] Since a balanced prior makes the model symmetric, the argument for an employee with a γ-signal is exactly the same.

Therefore, for the ρ-employee to invest efficiently, it is necessary that the red project win when the supervisor's signals go in its favor (branch 3 in Figure 3). This, in turn, requires that the supervisor's posterior probability that the state is Red after receiving two signals indicating that this is the case exceed one-half:[15]

$$\Pr(R \mid e_i = \rho, e_j = \gamma, s_i = h, s_j = l) = \frac{(1-\mu)\varepsilon(1-\varepsilon)\sigma^2}{(1-\mu)\varepsilon(1-\varepsilon)\sigma^2 + \mu(1-\varepsilon)\varepsilon(1-\sigma)^2} > \frac{1}{2}.$$

Figure 4
Efficient and Inefficient Project Choice



Rearrangement of this condition yields inequality (1) in Proposition 2, which confirms the intuition that it is satisfied, i.e., the supervisor can be convinced that the red project is the more profitable one, if the accuracy of her signals, σ, is high enough relative to the strength of her initial conviction, μ. In Figure 4, this is the area of the (σ, ε) space (for a given μ) that lies to the right of the vertical line denoted by $H_1 = 0$.[16] To the left of this line, in region III where $\sigma \leq 1/[1 + \sqrt{1 - \mu/\mu}]$, efficient investment is not an equilibrium outcome: since the red project never wins over the green one, the ρ-employee will never choose it, even though he thinks it is more profitable.

But it is not sufficient that the supervisor receives enough new information to potentially overcome her initial bias towards the green project, i.e., that we are to the right of the line $H_1 = 0$ in Figure 4. For the ρ-type employee to refrain from

[15] Recall that the supervisor knows the employees' signals because efficient investment is perfectly separating.

[16] Notice that ε is not an exogenous parameter, but rather a choice variable. Its equilibrium level will be determined in section 3.2.3.

yes-man behavior, the red project must give him a higher probability of winning the promotion than the green project does (given that the red project indeed does win with two favorable supervisor signals):

$$\mathrm{E}\{P_i(r) \mid \rho\} - \mathrm{E}\{P_i(g) \mid \rho\} \geq 0\,.$$

This condition is equivalent to inequality (2) in Proposition 2, which says that, conditional on a ρ-signal, it must be more likely than not that the supervisor receives two signals indicating that the state is Red, so that she gives the promotion to the red project. This happens if either ε or σ is large enough. It is because an accurate signal for the ρ-employee makes him conf dent that the red project is prof table and an accurate signal for the supervisor makes it likely that she will learn this from her signals. In Figure 4, this is the area above the curve where $H_2 = 0$, denoted region I. In region II, below the curve where $H_2 = 0$, ε and σ are both so low that the ρ-employee has a better chance of winning the promotion with the green project than with the red one. Hence, region I is the only part of the parameter space where eff cient investment is an equilibrium outcome. In regions II and III, it is not rational for the employees to choose the efficien red project after they receive a ρ-signal.

An increase in the prior bias towards the green project, μ, shrinks region I in Figure 4 and makes yes-man behavior more of a problem by shifting $H_1 = 0$ to the right and $H_2 = 0$ up. However, these two effects are quite different in that they affect the posteriors of two different agents. The shift in H_1 occurs because it makes the supervisor more opinionated and convinced that the green project is the more prof table one. This requires a more accurate signal, i.e., a higher σ, to convince her otherwise. The shift in H_2 occurs because it makes the ρ-employee less convinced that the red project is more prof table and therefore more pessimistic about the chance that this project will generate favorable supervisor information.

PROPOSITION 2 *The equilibrium choice of project is effi ient if the supervisor's and employees' signals are both accurate enough compared to the skew in the prior, as captured by the following conditions:*

(1) $$\left(\frac{\sigma}{1-\sigma}\right)^2 > \left(\frac{\mu}{1-\mu}\right),$$

(2) $$\Pr(R \mid \rho)\sigma^2 + \Pr(G \mid \rho)(1-\sigma)^2 \geq \frac{1}{2}\,.$$

In particular, if the prior is balanced, i.e., if $\mu = 1/2$, then eff cient project choice is always an equilibrium outcome.

An increase in the prior bias towards the green project, μ, makes eff cient project choice less likely, while an increase in the accuracy of the supervisor's information, σ, makes effi ient project choice more likely.

When yes-man behavior precludes eff cient investment in regions II and III, the unique equilibrium choice of investment project is that the employees completely ignore their private information and always choose the green project.

PROPOSITION 3 *If inequality* (1) *or* (2) *is violated, then the unique equilibrium project choice is that the employees always choose the green project regardless of their own signal.*

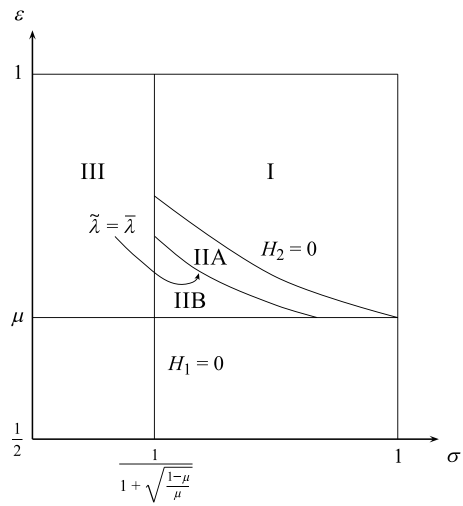
3.2.3 Effort Choice and Tournament Design

Next, consider the effort stage of the game. In a symmetric equilibrium the employees choose the same level of effort and each wins the tournament with probability one-half. The employees' choice of effort is controlled by the f rm through its choice of the benef t of winning the promotion, $W_1 - W_2$. However, in view of the risk of yes-man behavior, the f rm's value of employee effort depends on how the employees will end up using their information when they choose investment projects.

There are two main cases to consider, illustrated in Figure 5. Take the strength of prior beliefs, μ, and the accuracy of the supervisor's signals, σ, as given. Suppose f rst that it is impossible to convince the supervisor that the red project is superior, i.e., that μ is large enough relative to σ to make $\sigma \leq 1/[1 + \sqrt{1 - \mu/\mu}]$. This puts us in region III, where investment choice is ineffic ent regardless of the accuracy of the employees' signals and therefore also regardless of their effort. Hence, the firm s marginal benefi of effort is always equal to zero, and as a consequence it dismantles the tournament and extracts no effort at all.

The second possibility is that the supervisor can be convinced that the red project is better, i.e., that μ is small enough relative to σ so that $\sigma > 1/[1 + \sqrt{1 - \mu/\mu}]$. This puts us in region I or II, depending on the accuracy of the employees' information.

Figure 5
Overall Equilibrium Depending on the Cost of Effort, the Accuracy of the Public Information, and the Accuracy of the Supervisor's Signals



Suppose first that the cost of effort, c, is small enough to make the first-best efficient amount of employee effort, $\lambda^{**} = 1/c$, so large that it would put us in region I. The first-best efficient amount of employee information would now be used fully and therefore be perfectly revealed. As a consequence, the overall equilibrium would be the first-best efficient one described in Proposition 1.

By contrast, if the cost of effort is so high that the first-best efficient quantity of employee effort puts us in region II, then this amount of employee information would not be used at all. This means that its marginal benefit is equal to zero, which is inconsistent with the expenditure of costly effort to get it. Therefore, the first-best efficient outcome is not an equilibrium outcome. However, it is possible that a quantity of employee effort above and beyond the first-best efficient one could take us into region I, where the information would be used fully. Let $\widetilde{\lambda}$ be the threshold amount of effort that brings us to the boundary between regions II and I. Since λ^{**} puts us in region II, it must be the case that $\widetilde{\lambda} > \lambda^{**}$. Now $\widetilde{\lambda}$ increases in μ (it shifts the boundary $H_2 = 0$ up) and decreases in σ (it moves us closer to the boundary $H_2 = 0$). Even though $\widetilde{\lambda}$ is suboptimally high, as long as it gets some net benefit from extracting that level of effort, the firm would choose to do so. This is feasible as long as the amount of effort required is not so large that it wipes out the surplus completely, i.e., as long as $\widetilde{\lambda} \leq \overline{\lambda}$. This is the case in region IIA in Figure 5. Here the firm extracts $\widetilde{\lambda}$ units of effort and the employees use their private information to guide their investment decision.

When $\widetilde{\lambda} > \overline{\lambda}$ – which happens if μ is high and σ is low – having the employees put in enough effort to take us to region I is too costly, so it is better for the firm to extract no effort at all. This is the case in region IIB in Figure 5.

PROPOSITION 4 *The overall equilibrium depends on the exogenous parameters μ, σ, and c.*

(1) If $\sigma \leq 1/[1 + \sqrt{1 - \mu/\mu}]$, then the firm chooses not to use the tournament (setting $W_1^* = W_2^* = 0$), the employees supply no effort, and only public information is used in the investment decisions.

(2) If $\sigma > 1/[1 + \sqrt{1 - \mu/\mu}]$ and c is low enough so that when the employees supply the first-best efficient amount of effort, λ^{**}, the accuracy of their private information, $\mu + \lambda^{**}$, is high enough to make $H_2 \geq 0$, then the equilibrium is the first-best efficient one described in Proposition 1.

(3) If $\sigma > 1/[1 + \sqrt{1 - \mu/\mu}]$ and c is high enough so that when the employees supply the highest effort level that generates a surplus, $\overline{\lambda}$, the accuracy of their private information, $\mu + \overline{\lambda}$, is low enough to make $H_2 < 0$, then the firm chooses not to use the tournament (setting $W_1^* = W_2^* = 0$), the employees supply no effort, and only public information is used in the investment decisions.

(4) If $\sigma > 1/[1 + \sqrt{1 - \mu/\mu}]$ and c is in the intermediate range where $\mu + \lambda^{**}$ makes $H_2 < 0$ and $\mu + \overline{\lambda}$ makes $H_2 \geq 0$, then the firm sets the tournament incentive intensity above the first-best efficient level, $W_1^* - W_2^* > 1/\phi$, to make the employees provide the effort level $\widetilde{\lambda} \in (\lambda^{**}, \overline{\lambda})$ that makes $H_2 = 0$, and the employees choose the investment project that their own private signal indicates is more profitable.

To summarize, by turning the employees into yes men the tournament may render the collection of valuable information impossible. This happens if the accuracy of the supervisor's information relative to that of the public information, σ/μ, is suff ciently low (region III) or if the employees' cost of effort, c, is suff ciently high (region IIB). If the supervisor's information is suffic ently accurate relative to the public information – σ/μ is suff ciently high – and the employees' cost of effort, c, is suffic ently low (region I), then the yes-man problem does not affect either investment, effort, or tournament design. Finally, if the supervisor's information is suff ciently accurate relative to the public information – σ/μ is suff ciently high – and the employees' cost of effort, c, falls in an intermediate range (region IIA), then the yes-man problem increases the incentive intensity in the tournament and the employees' effort above the f rst-best eff cient level, but allows investment choice to be effic ent, given employee effort, and in fact more likely to be profi able than under f rst-best eff cient effort.

3.2.4 Eff ciency

Finally, Proposition 5 describes what the eff ciency loss looks like. When the yes-man problem altogether precludes information collection (regions IIB and III), the loss amounts to the expected value of the forgone information. When it forces excessive information collection in order to induce the employees to use it (region IIA), the loss is the negative welfare from the employee effort above and beyond the eff cient level.

PROPOSITION 5 *The per-employee effi iency loss in equilibrium, L^*, looks as follows.*

(1) If the equilibrium allows no information collection (cases 1 and 3 in Proposition 4), then the loss is the value of the information that is not collected and used: $L^* = \lambda^{**} - C(\lambda^{**}) = 1/2c$. The loss is strictly decreasing in the cost of effort, c.

(2) If the equilibrium involves excess effort (case 4 in Proposition 4), then the loss is the welfare loss from the excess effort: $L^* = -\{[\widetilde{\lambda} - \lambda^{**}] - [C(\widetilde{\lambda}) - C(\lambda^{**})]\}$. The loss is strictly increasing in the amount of public information, μ, and in the cost of effort, c. It is strictly decreasing in the accuracy of the supervisor's information, σ.

In general, more information for the supervisor weakens the yes-man incentive in the sense of moving us towards the part of the parameter space where eff cient investment can be sustained. This is also true when the ineff ciency consists of too much information being collected in region IIA: by moving the boundary where $H_2 = 0$ closer to the eff cient outcome, an increase in σ decreases the amount of excess effort – and loss of eff ciency – required. On the other hand, when the ineff ciency consists of no information collection, then the loss is independent of σ: a small increase in σ cannot take us out of region IIB or III, even though it takes us a little closer to their boundaries.

Public information, μ, does the opposite of σ. In general, a higher μ aggravates yes-man problems by shifting the boundary where $H_1 = 0$ to the right and the one where $H_2 = 0$ up. And when the inefficiency consists of excessive effort, an increase in μ aggravates the inefficiency by forcing an even higher effort level as it moves $H_2 = 0$ further away. Once again, if no effort is supplied, then the inefficiency is independent of μ.

The effect of an increase in the cost of effort on the inefficiency from yes-man behavior also depends on whether too little or too much effort is supplied in equilibrium. When the inefficiency consists of information not being used at all, a higher cost of effort leads to a lower efficiency loss. This is because a cost increase decreases the efficient amount of information that could have been collected, thus diminishing the forgone net gain from not having it. When the inefficiency consists of too much information being collected, a higher cost leads to a higher efficiency loss. This is because a cost increase diminishes the net gain from both the efficient amount of effort and the higher equilibrium amount of effort. However, this decrease is strictly smaller for the efficient amount of effort, because it is rationally chosen, and therefore adjusts to soften the blow of a higher cost.

4 Extensions

We pursue two extensions of our basic model. The first is to allow the two employees to choose investment projects with payoffs that are less than perfectly correlated. This will give some insight into how yes-man tendencies are affected by how similar the tasks of the employees are. The second is to compare yes-man incentives in a tournament with those when the employees instead have individual compensation contracts. The purpose of this exercise is to highlight which aspects of a tournament it is that play a key role in turning employees into yes men.

4.1 Learning about Different Sources of Uncertainty

In the baseline model, the two employees collect information about the same uncertainty. One way to interpret this is that the employees' tasks are identical. It is natural to ask whether yes-man incentives change if an employee competes for promotion against someone who is charged with a very different task. To address this issue, in this section we generalize the uncertainties that the employees learn about to allow for any arbitrary level of correlation between them.[17] All signal processes remain independent of one another, conditional on the state of the world. We now denote the states of the world governing the profitability of the employees' investment projects by $Z_i \in \{G, R\}$ and $Z_j \in \{G, R\}$, respectively. It is still the case that for both employees' investments, the prior, unconditional probability that the state is Green is μ. But the two uncertainties are no longer independent, and we parameterize the correlation between the payoffs from the employees' investment projects by $\kappa \in [0, 1]$.[18] Table 2 gives the joint probability distribution of Z_i and Z_j.

[17] This generalization is taken from Heidhues and Lagerlöf (2003).

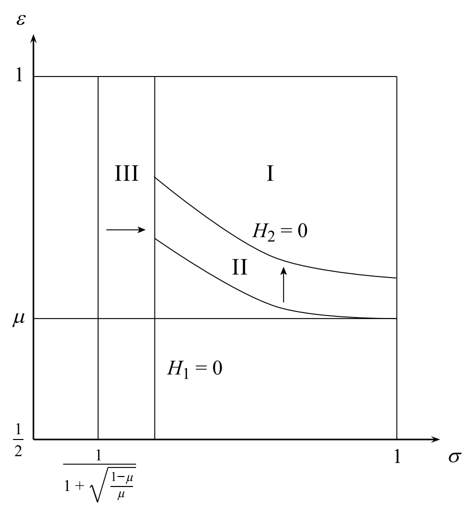
[18] Notice, however, that κ is not the correlation coefficient. Also notice that the main analysis covers the special case when $\kappa = 1$.

Table 2
Joint Probability Distribution of Z_i and Z_j

Z_i / Z_j	G	R	Σ
G	$\mu^2 + \kappa\mu(1-\mu)$	$(1-\kappa)\mu(1-\mu)$	μ
R	$(1-\kappa)\mu(1-\mu)$	$(1-\mu)^2 + \kappa\mu(1-\mu)$	$1-\mu$
Σ	μ	$1-\mu$	

A weaker correlation between the employees' sources of uncertainty aggravates the yes-man problem because it provides the supervisor with less information about the employees' investment decisions. This is illustrated in Figure 6, which is the analogue of Figure 4 for the case of less than perfectly correlated investment uncertainties. A lower κ diminishes the amount of information that the supervisor can infer about employee i's project prof tability from her information about employee j's decision. As a result, supervisor information of any accuracy, σ, now makes less of a dent in her initial conviction that green projects are better and therefore shifts the boundary $H_1 = 0$ to the right. Moreover, even if the supervisor could be convinced that the red project is the more prof table one, with a lower κ the ρ-employee should expect the supervisor to get less information that supports the red project. This shifts the boundary $H_2 = 0$ upward. Hence, the subset of the (ε, σ) space where eff cient choice of investment project is an equilibrium outcome, region I, becomes strictly smaller as the correlation between the investment uncertainties weakens.

Figure 6
Efficien and Ineff cient Project Choice with Less than Perfectly Correlated Investment Uncertainties



PROPOSITION 6 *As the correlation between the sources of uncertainty that the employees learn about grows weaker, yes-man behavior becomes a more serious problem in the sense that the region of the parameter space where effi ient investment is an equilibrium outcome shrinks.*

One important conclusion in the theory of incentives is that comparing the performance of employees with similar tasks is benefic al because it shields them from common risk (Holmström, 1982). But when employees manage private information on behalf of their employer, such an arrangement also has another advantage, namely that it gives the employer a more accurate estimate of what that information might be. This second benef t from comparing employees with similar tasks therefore has nothing to do with risk-sharing. Rather, it allows the supervisor to replace preconceived notions with a better-informed evaluation, something that in turn lets the employees' own knowledge inform their decisions.

4.2 A Comparison with Individual Contracts

We now return to the baseline model where the uncertainties of the investment opportunities are perfectly correlated.

As mentioned in the introduction, rank-order tournaments are not alone in turning employees into yes men; reputation concerns and individual contracts can do the same (Gentzkow and Shapiro, 2006, and Prendergast, 1993b, respectively). It is therefore natural to ask whether yes-man problems tend to be more or less severe with a tournament than with individual contracts. In this section, we compare these two incentive mechanisms with respect to their ability to generate efficien investment. Just like the tournament, individual contracts can implement a particular level of employee effort to support the investment decision, but we leave this part of the analysis aside.

A tournament models competition between employees for promotion, which gives it two features that are often highlighted as def ning. First, pure competition implies that only a relative performance measure should be used. Second, since the prizes in the competition are different jobs, the def nitions of which are often inf uenced by considerations other than incentive provision, their total value is f xed no matter what happens in the tournament. Individual incentive contracts typically do not have these two properties, in economic theory or in real life. But in our analysis they will, in the interest of focusing on less obvious differences between the two incentive mechanisms. Hence, we will assume that our individual contracts also distribute a constant total compensation of $W_1 + W_2$ and use the same, purely relative performance measure, namely the supervisor's assessment of the difference in prof tability of the two projects, $\widehat{\pi}_i - \widehat{\pi}_j$.[19]

Even with these restrictions, individual compensation contracts turn out to be better than a tournament at discouraging yes-man behavior. This is because a tour-

[19] Again, because of risk neutrality, the random shocks in the performance measure do not affect the investment decisions.

nament must always distribute the same two f xed prizes among the contestants, which means that not only is the total compensation f xed, but the difference between winning and losing, $W_1 - W_2$, is as well. And this effectively translates into a wasteful use of the information contained in the performance measure: in a tournament the reward for winning depends only on the sign of the performance measure, not on its magnitude. In other words, the reward for winning is forced to be the same whether it is by a wide margin or by just a hair. By contrast, individual compensation contacts, which have prizes that are calibrated with an exclusive focus on incentive provision, can allow the benef t of winning to depend on the margin of victory. This enables such contracts to decrease the reward for having a higher performance measure when this is due merely to turning your coat to the wind by catering to prior opinions.

To see that this is the case, we want to compare the incentive to deviate from efficien investment through yes-man behavior in a tournament with that under individual incentive contracts of the type specif ed above. Therefore, we start from the assumption that the employees invest eff ciently by following their signals. The performance measure given to employee i's investment, $\widehat{\pi}_i - \widehat{\pi}_j$, can now take on seven values, depending on the employees' choice of projects and the realizations of the supervisor's signals. If both employees choose the same project, then $\widehat{\pi}_i - \widehat{\pi}_j = 0$, regardless of what the supervisor's signals turn out to be. The expected prize or wage in this case is denoted by W_0 (see Figure 7). If the employees choose different projects, then they must have received different signals, which the supervisor can infer because effic ent investment is perfectly separating. She therefore knows that the two contradictory employee signals cancel and her own signals provide the only new information. If employee i chooses the green (red) project and the supervisor receives signals s_i and s_j, then $\widehat{\pi}_i - \widehat{\pi}_j = \Pr(G \mid s_i, s_j) - \Pr(R \mid s_i, s_j)$ $(\widehat{\pi}_i - \widehat{\pi}_j = \Pr(R \mid s_i, s_j) - \Pr(G \mid s_i, s_j))$. This gives us six realizations of $\widehat{\pi}_i - \widehat{\pi}_j$, symmetrically distributed around zero. To get a briefer notation for the dependence of $\widehat{\pi}_i - \widehat{\pi}_j$ on both employee i's project choice and the supervisor's signals, we indicate the choice of project simply by the state in which it is prof table, and the realizations of the supervisor's signals by which state they indicate is more likely, using the same symbols as for the employees' signals. For example, if employee i chooses the red project (and employee j therefore the green) and $s_i = h$ and $s_j = l$, so that both signals indicate that $Z = R$, then $\widehat{\pi}_i - \widehat{\pi}_j = \Pr(R \mid \rho\rho) - \Pr(G \mid \rho\rho)$.

The question we want to address is whether individual contracts can implement effic ent investment when the tournament fails to do so. Consider fir t region II of the parameter space in Figure 4, where $H_1 > 0$ and $H_2 < 0$, i.e., the supervisor can be convinced that the red project is more prof table, but this is so unlikely that the ρ-employee chooses the green project instead. Figure 7 illustrates how the two incentive mechanisms compensate employee i for different realizations of the performance measure, $\widehat{\pi}_i - \widehat{\pi}_j$.[20] The solid line is the wage structure of the

[20] The individual incentive contract that is derived here is only one example of an inf nite number of individual incentive contracts that can implement eff cient investment.

tournament. It relies exclusively on relative performance evaluation and uses it only as an ordinal measure: the wage depends only on the sign of the evaluation measure, not on its absolute magnitude. Notice that the red project wins the ρ-agent the promotion (a prize of W_{++}) only if both of the supervisor's signals indicate that this is the best choice. Both of the other winning realizations of $\widehat{\pi}_i - \widehat{\pi}_j$ (with prizes of W_+ and W_{+++}) belong to the green project. The tournament therefore administers the appropriate rewards and punishments relative to the supervisor's signal. When the signals indicate that the state is Green (Red), then the green and the red project receive W_{+++} and W_{---} (W_{--} and W_{++}), respectively. It is W_+ and W_- that play a key role in generating yes-man behavior: when the supervisor's signals contradict one another and cancel, the skewed prior dictates that the green project be rewarded with W_+ and the red one be punished with W_-. An optimal individual contract, illustrated by the dashed line in Figure 7, can remove this perverse incentive and implement eff cient investment. This is done by decreasing W_+ and increasing W_-, and the required adjustment is small enough to allow the contract to remain monotonic, i.e., $W_- \leq W_0 \leq W_+$.[21]

Ineff cient investment also occurs in region III in Figure 4, where $H_1 \leq 0$, so that not even the strongest information from her own signals can convince the supervisor that the red project is more prof table. Figure 8 illustrates wage structures in this case, once again with the tournament being the solid line and the individual contract being the dashed line. Now, the tournament incentives are even worse. Not only is the green project rewarded (W_{++} compared to W_{--} for the red project) when the supervisor receives no new information because her two signals contradict one another. It also receives a higher compensation (W_+ rather than W_-) when the supervisor's signals both indicate that the state is not Green, but rather Red. Since the red project never wins the promotion outright (all three positive evaluation realizations are generated by the green project), the ρ-employee turns into a yes man. But an optimal individual contract, which is allowed to change the compensation depending on the margin by which the red project loses, once again can overcome this problem. In this case, that requires that the red project be given the higher prize even when it has a lower expected prof tability, but by the smallest margin. Hence, in order to discourage yes-man behavior, the individual incentive contract must be nonmonotonic.[22] And just as was the case in region II, the individual contract effectively turns the supervisor's signals net of the prior into the performance measure by removing the green project's advantage, f rst, in the absence of new information by setting $W_{++} = W_{--} = W_0$, and, second, in the face of contradictory new information by increasing W_- and decreasing W_+.

[21] One contract that would achieve this that is appealing because of its simplicity is one with $W_- = W_0 = W_+$. This effectively weeds out the prior from the performance measure, leaving only the supervisor's signal realizations, which give an unbiased prediction of profit bility.

[22] However, the optimal W_- can be strictly smaller than W_1 and the optimal W_+ can be strictly larger than W_2.

Figure 7
The Wage Structure of the Tournament and an Eff cient Individual Contract in Region II, where $H_1 > 0$ and $H_2 < 0$

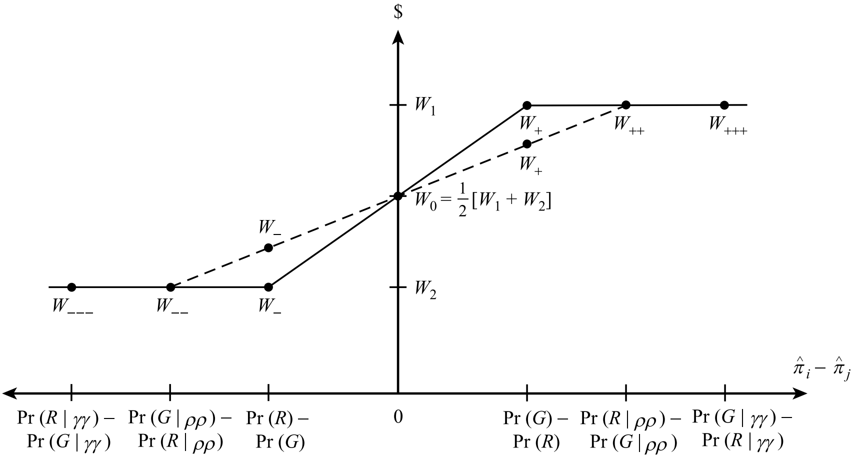
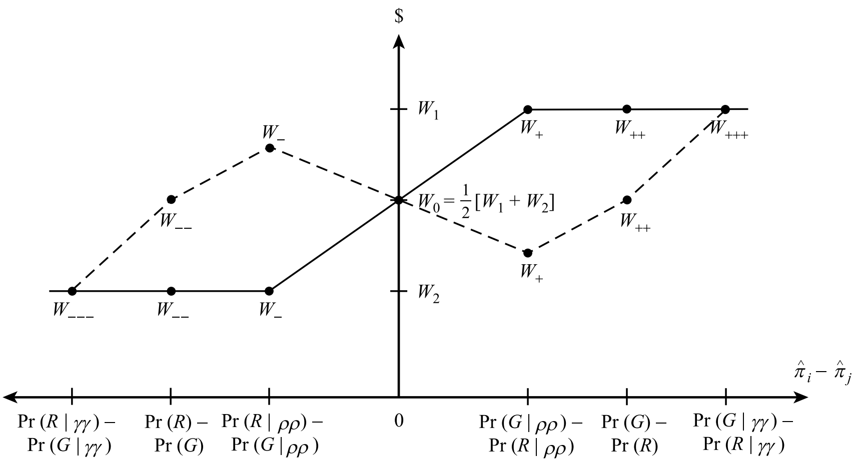


Figure 8
The Wage Structure of the Tournament and an Eff cient Individual Contract in Region III, where $H_1 \leq 0$



PROPOSITION 7 *Individual contracts with the same performance measure and the same fi ed wage bill can correct the yes-man problem in the tournament. In region II in Figure 4, eff cient individual contracts can remain monotonic, but in region III effi ient investment can be achieved only if individual contracts are nonmonotonic.*

The conclusion that individual contracts do a better job of discouraging yes-man behavior once again raises the often-asked question as to what offsetting advantages tournaments may have that could explain their use. One aspect of tournaments that has been highlighted as a major advantage because it eliminates moral-hazard problems on the principal's side is that it distributes a fixed overall wage bill. However, this is a feature that contracts that eliminate yes-man problems are able to replicate. But even though they pay out a constant total wage, the individual contracts have a wage structure that is more fine-tuned. This requires more accurate performance evaluation, and it may well be that this is where a tournament has a decisive advantage: it allows the supervisor to get away with determining only which employee does a better job, not how much better a job he does.

5 *Discussion*

Our analysis illustrates how a rank-order tournament compromises the ability of employees to act with integrity and, in fact, can turn them into yes men who make decisions that conform to preconceived – but relatively ill-informed – opinions. In other words, the competitive pressure that provides employees with the incentive to collect potentially valuable information may be self-defeating because it also prevents them from using the information they collect, thus rendering it worthless.

The same tension between incentive provision and information management that we study is also the focus of Prendergast (1993b), who demonstrates that the tendency to conform to the opinions of those who administer rewards is nothing unique to a tournament, but appears when incentives are provided with an individual contract as well. Apart from the incentive mechanisms studied, the main difference between Prendergast's analysis and ours lies in the perception of the information that the supervisor has access to. In Prendergast's model, the supervisor collects information at only one point in time.[23] By contrast, our model makes a sharp distinction between what the employees know and what they do not know about the supervisor's information because of when it arrives. Information that the supervisor receives before the employees' decisions makes the tournament unfair in that it favors one project over the other. Unfortunately, the employees can react to such early information by choosing away from alternatives with the deck stacked against them and instead choose the supervisor's favorite alternative. Information that the supervisor receives after the employees have committed to an alternative, on the other hand, discourages yes-man behavior because the employees can expect it to reenforce the tendency for the supervisor to ultimately see the world the way they do.

The source of the yes-man incentive in tournaments is the supervisor's prior opinion. Therefore, to the extent that her objective is to make employees behave in a way that maximizes the firm's profit, the supervisor should ignore her prior

[23] As a consequence, Prendergast's model lacks any countervailing cost to yes-man behavior of the type that the supervisor's signals provide in our model.

information: with a balanced prior, the employees have no reason to privilege one project over the other unless their own information about profitability tells them that they should. However, we think of the supervisor's prior beliefs and opinions as capturing not a fleeting impression that can be erased at will, but rather past information that has shaped the very core of the supervisor's view of the world in a way that can be reversed only by new information.

But even if it may be difficult for the individual supervisor to prevent her preconceived notions from influencing promotion decisions, the firm as an organization may be able to take measures to do so.[24] We have argued that individual contracts diminish the influence of the supervisor's prior on performance evaluation. But there may be other ways to do this as well. Notice that it is the employees' perception of the supervisor's initial bias that matters. Therefore, institutional arrangements that make it more difficult for employees to discern which prior they should cater to would benefit the management of private information inside the firm. One way to accomplish this might be to let a group make the promotion decision: compared to gauging the biases of a single person, it is likely to be more difficult for employees to predict which prior opinions will carry the day in a group decision. Similarly, engaging a less well-known outside evaluator may serve the same purpose of confusing the employees about which direction they should turn their coats. This provides a new rationale – in addition to discouraging various inefficient influence activities – for taking the distribution of rewards to employees out of the hands of their immediate supervisors.

The model we develop is extremely simple, but many of the simplifying assumptions that we make do not alter the qualitative economic results. For example, this is the case with the restriction to only two investment projects, the normalization of their payoffs to zero and one, the conditional independence of the signals, the rudimentary strategic interaction in the effort stage of the game, and the fact that the employees use their information in a real decision rather than simply report it to their supervisor. Furthermore, it is our conjecture that the restriction of the evaluation shock to a uniform distribution is innocuous as well.

By contrast, other assumptions are necessary for our results. A crucial one is that the expected profit is used to evaluate the employees' investment decisions. This seems reasonable if the task that the employees handle is complex, because then expected profit serves as an overall measure of the effects of what the employees do on the welfare of the firm's shareholders. But the yes-man incentive highlighted in this paper certainly constitutes a drawback to rewarding employees according to the shareholders' less than perfectly informed preferences over their choices.

---

24 A balanced prior also plays an important role in the legal setting (Demougin and Fluet, 2006, and Posner, 1999). Since it is considered important that only the evidence presented in a case influences the ruling, there are procedural rules, such as the exclusion of certain types of evidence, that push courts towards acting as if they have a balanced prior. Moreover, one potential benefit of judgment by jury is that with several people involved, their priors may cancel each other out.

Another key assumption is that we do not allow for the possibility that employees are heterogeneous in their innate ability to collect information. If this were the case, the firm might want to promote abler employees. The employees would then have an incentive to signal their own ability, to the extent that they know it. One way this could be done is by expressing an opinion that in fact contradicts the conventional wisdom. Hence, the desire to signal innate learning ability may introduce a countervailing incentive for employees to become not "yes men," but rather "no men" (Kim and Ruy, 2003, and Prendergast and Stole, 1996).

Finally, we restrict what the employees can do in a crucial way in that they cannot choose how to allocate their effort across the different projects: an increase in effort increases the accuracy of the signals about both projects by the same amount. An interesting question is whether the competitive pressure of a promotion tournament not only makes employees neglect information that they have, but also distorts their allocation of information-collection effort across different projects. We hope to address this issue in future research.

Appendix

A.1 Proof of Proposition 1

Consider first the employees' last decision of choosing the investment project. Recall that each employee makes this choice with the objective of maximizing the expected profit from his project, given the supervisor's information. The supervisor's expectation is unbiased, and the employee's choice does not reveal or conceal any of his own information, so the employee's expectation of the supervisor's assessment of the expected profit is simply equal to the employee's own expectation of the profit. Thus, the employee chooses the project that his own signal indicates is the most profitable one.

Next, consider employee i's effort choice. His marginal benefit of effort is equal to

$$\begin{aligned}
&\frac{\partial}{\partial \lambda_i}\left\{\Pr(\xi < \mathrm{E}[\hat{\pi}_i(\varepsilon_i,\varepsilon_j,e_i,e_j,z_i,z_j,s_i,s_j)] - \mathrm{E}[\hat{\pi}_j(\varepsilon_i,\varepsilon_j,e_i,e_j,z_i,z_j,s_i,s_j)])[W_1 - W_2]\right\}\\
&\qquad = \frac{\partial \Pr(\xi < x)}{\partial x}\left\{\frac{\partial \mathrm{E}[\hat{\pi}_i]}{\partial \varepsilon_i}\frac{\partial \varepsilon_i}{\partial \lambda_i} - \frac{\partial \mathrm{E}[\hat{\pi}_j]}{\partial \varepsilon_i}\frac{\partial \varepsilon_i}{\partial \lambda_i}\right\}[W_1 - W_2]\\
&\qquad = \phi\left\{\frac{\partial \mathrm{E}[\hat{\pi}_i]}{\partial \varepsilon_i}\frac{\partial \varepsilon_i}{\partial \lambda_i} - \frac{\partial \mathrm{E}[\hat{\pi}_j]}{\partial \varepsilon_i}\frac{\partial \varepsilon_i}{\partial \lambda_i}\right\}[W_1 - W_2]\,.
\end{aligned}$$

Again, at the time when he chooses effort, the employee's expectation of the supervisor's expectation of project profits is simply equal to his own expectation. When information is used efficiently, this expectation of the supervisor's expectation is equal to the accuracy of the employee's signal:

$$\mathrm{E}[\hat{\pi}_i] = \mathrm{E}[\pi_i(\varepsilon_i, z_i)] = \varepsilon_i, \qquad \mathrm{E}[\hat{\pi}_j] = \mathrm{E}[\pi_j(\varepsilon_j, z_j)] = \varepsilon_j\,.$$

The fact that $\varepsilon_i = \mu + \lambda_i$ makes employee i's marginal benefit of effort equal to $\phi[W_1 - W_2]$. His rational choice of effort, λ_i^*, is the one that equates its marginal benef t and marginal cost: $\phi[W_1 - W_2] = c\lambda_i^*$, so that $\lambda_i^* = (\phi/c)[W_1 - W_2]$. Since both employees face the same benef t and cost of effort, the effort stage has a unique equilibrium that is symmetric and given by the above optimality condition. Because the employees choose the same effort level, $\lambda_i^* = \lambda_j^* = \lambda^* = (\phi/c)[W_1 - W_2]$, each of them can expect to win the promotion with probability one-half.

Finally, consider the f rm's prof t-maximizing choice of tournament prizes, W_1^* and W_2^*. The prof t-maximizing tournament should implement eff cient effort levels, so $W_1^* - W_2^* = c\lambda_i^{**}/\phi = 1/\phi$. Moreover, prof t maximization implies that the employees' participation constraints must bind, so $[W_1^* + W_2^*]/2 = C(\lambda^{**}) = 1/2c$. These two conditions imply that

$$W_1^* = \frac{1}{2}\left(\frac{1}{c} + \frac{1}{\phi}\right) \quad \text{and} \quad W_2^* = \frac{1}{2}\left(\frac{1}{c} - \frac{1}{\phi}\right).$$

Q.E.D.

A.2 Proof of Proposition 2

In a symmetric equilibrium, the employees have chosen the same effort level, ε. Considering an unbalanced prior, $\mu > 0$, we f rst derive condition (1):

$$H_1 \equiv \Pr(R \mid e_i = \rho, e_j = \gamma, s_i = h, s_j = l) - \frac{1}{2} > 0,$$
$$\frac{(1-\mu)\varepsilon(1-\varepsilon)\sigma^2}{(1-\mu)\varepsilon(1-\varepsilon)\sigma^2 + \mu(1-\varepsilon)\varepsilon(1-\sigma)^2} > \frac{1}{2},$$
$$\left(\frac{\sigma}{1-\sigma}\right)^2 > \frac{\mu}{1-\mu}.$$

H_1 is strictly decreasing in μ and strictly positive if $\mu = 1/2$. It is obvious that if condition (1) is violated so that $H_1 < 0$, then eff cient investment cannot occur in equilibrium, because the probability of winning the promotion with the red project against a green project is equal to zero. The same is true if $H_1 = 0$, but to economize on space, the proof concerning this nongeneric outcome is omitted.

If the red project wins with two favorable signals, i.e., if $H_1 > 0$, then eff cient project choice is an equilibrium outcome as long as an employee with a ρ-signal receives a positive expected net benefi from choosing the red project rather than the green one, denoted by H_2:

$$H_2 \equiv \mathrm{E}\{P_i(r) \mid \rho\} - \mathrm{E}\{P_i(g) \mid \rho\} \geq 0,$$

$$\Pr(z_j = r \mid \rho)\frac{1}{2} + \Pr(z_j = g, s_i = h, s_j = l \mid \rho)$$
$$\geq \Pr(z_j = g \mid \rho)\frac{1}{2} + [\Pr(z_j = r \mid \rho) - \Pr(z_j = r, s_i = l, s_j = h \mid \rho)],$$

$$\begin{aligned}
&\Pr(R \mid \rho)\left\{\Pr(z_j = r \mid R)\frac{1}{2} + \Pr(z_j = g, s_i = h, s_j = l \mid R)\right\} \\
&\quad + \Pr(G \mid \rho)\left\{\Pr(z_j = r \mid G)\frac{1}{2} + \Pr(z_j = g, s_i = h, s_j = l \mid G)\right\} \\
\geq\ &\Pr(R \mid \rho)\left\{\Pr(z_j = g \mid R)\frac{1}{2} + [\Pr(z_j = r \mid R) - \Pr(z_j = r, s_i = l, s_j = h \mid R)]\right\} \\
&\quad + \Pr(G \mid \rho)\left\{\Pr(z_j = g \mid G)\frac{1}{2} + [\Pr(z_j = r \mid G) - \Pr(z_j = r, s_i = l, s_j = h \mid G)]\right\},
\end{aligned}$$

$$\begin{aligned}
&(1-\varepsilon)\frac{1}{2} + \Pr(R \mid \rho)(2\varepsilon - 1)\frac{1}{2} + \{\Pr(R \mid \rho)(1-\varepsilon)\sigma^2 + \Pr(G \mid \rho)\varepsilon(1-\sigma)^2\} \\
\geq\ &(1-\varepsilon)\frac{1}{2} - \Pr(G \mid \rho)(2\varepsilon - 1)\frac{1}{2} + \varepsilon - \{\Pr(R \mid \rho)\varepsilon\sigma^2 + \Pr(G \mid \rho)(1-\varepsilon)(1-\sigma)^2\},
\end{aligned}$$

$$\Pr(R \mid \rho)\sigma^2 + \Pr(G \mid \rho)(1-\sigma)^2 \geq \frac{1}{2}.$$

The left-hand side is strictly increasing in both σ and ε, so it follows from the implicit-function theorem that the boundary in σ–ε space implicitly defined by $H_2 = 0$ has a negative slope. Moreover, condition (2) holds for any value of σ if ε is large enough, as well as for any value of ε if σ is large enough. Finally, it is easy to confirm that H_2 is strictly decreasing in μ. *Q.E.D.*

A.3 Proof of Proposition 3

Outside region I, both employee types choosing the green project is an equilibrium in the investment stage of the game. The supervisor's out-of-equilibrium beliefs after observing a red project must be specified to complete the proposed equilibrium investment. To rule out that such beliefs are economically unreasonable, we required that they satisfy universal divinity. The only belief that survives this criterion is that the supervisor interprets a red project as having been for sure chosen by a ρ-employee. This is because $\Pr(R \mid \gamma) < \Pr(R \mid \rho)$ and therefore the probability of generating supervisor signals that indicate that the red project is profitable is strictly smaller with a γ-signal than with a ρ-signal. Since promotion is based on the expected profit of the chosen projects, consider the probability that the supervisor promotes the employee that chooses the red project after observing signals indicating that it is profitable. The threshold probability of promotion with the red project that makes this choice rational is strictly higher with a γ-signal than with a ρ-signal. With a strictly smaller set of supervisor mixed strategies making it rational, the supervisor must rule out the possibility that the deviation was made by a γ-employee. And since conditions (1) and/or (2) assume that the supervisor recognizes a red project as being chosen by the ρ-employee, this belief makes the green project a sequentially rational choice for both types of employees.

The above equilibrium investment is unique. The only remaining possible pure-strategy equilibrium in the investment stage is a perfectly pooling one where both employee types choose the red project. However, an analogous argument to the one above rules this out as violating universal divinity. Since a green project is more likely to be prof table with a γ-signal than with the ρ-signal, a strictly smaller set of probabilities of promotion with the green project makes this choice attractive to the ρ-employee. Hence, the only out-of-equilibrium belief for the supervisor that satisf es universal divinity is that a green project is for sure chosen by a γ-employee. This, in turn, makes it optimal for the γ-employee to deviate and choose the green project instead. *Q.E.D.*

A.4 Proof of Proposition 5

The loss is the difference in expected net payoff from the employee's investment decision in the f rst-best eff cient outcome and in the equilibrium outcome. Start with cases 1 and 3 in Proposition 4 when no effort is provided and only public information is used to guide investment. Then the equilibrium expected welfare loss is equal to

$$L^* = [\varepsilon(\lambda^{**}) - C(\lambda^{**})] - [\mu - 0] = [\mu + \lambda^{**} - C(\lambda^{**})] - \mu = \lambda^{**} - \left(\frac{1}{2}\right)c(\lambda^{**})^2$$
$$= \lambda^{**}\left[1 - \left(\frac{1}{2}\right)c\lambda^{**}\right] = \left(\frac{1}{c}\right)\left[1 - \left(\frac{1}{2}\right)c\left(\frac{1}{c}\right)\right] = \frac{1}{2c}.$$

This expression is obviously strictly decreasing in c.

Next, consider case 4 in Proposition 4. The loss is now equal to

$$L^* = \left[\varepsilon\left(\lambda^{**}\right) - C\left(\lambda^{**}\right)\right] - \left[\varepsilon\left(\widetilde{\lambda}\right) - C\left(\widetilde{\lambda}\right)\right] = \left[\mu + \lambda^{**} - C\left(\lambda^{**}\right)\right] - \left[\mu + \widetilde{\lambda} - C\left(\widetilde{\lambda}\right)\right]$$
$$= \left[\lambda^{**} - C\left(\lambda^{**}\right)\right] - \left[\widetilde{\lambda} - C\left(\widetilde{\lambda}\right)\right] = -\left\{\left[\widetilde{\lambda} - \lambda^{**}\right] - \left[C\left(\widetilde{\lambda}\right) - C\left(\lambda^{**}\right)\right]\right\}.$$

Turning to the comparative statics of this loss function, consider f rst the effect of the public information, μ. We have that $\lambda^{**} - C(\lambda^{**}) = 1/2c$ is independent of μ, whereas $\widetilde{\lambda}$ is strictly increasing in μ because an increase in μ shifts the boundary between regions I and II upwards. Since a larger $\widetilde{\lambda}$ represents a larger deviation from the fir t-best eff cient effort level λ^{**}, the net surplus from $\widetilde{\lambda}$ must decrease. Hence, L^* is strictly increasing in μ.

The loss is strictly decreasing in the accuracy of the supervisor's private information, σ. Again, the net surplus from f rst-best eff cient effort is independent of σ, whereas the net surplus from $\widetilde{\lambda}$ is strictly increasing in σ because σ decreases $\widetilde{\lambda}$.

Finally, the loss is strictly increasing in the cost of effort, c. Notice that $\widetilde{\lambda}$ is independent of c, whereas λ^{**} is strictly decreasing in c. Write the loss as

$$L^* = [\lambda^{**} - C(\lambda^{**})] - \left[\widetilde{\lambda} - C\left(\widetilde{\lambda}\right)\right] = \frac{1}{2c} - \left[\widetilde{\lambda} - \left(\frac{1}{2}\right)c\left(\widetilde{\lambda}\right)^2\right].$$

Differentiating this expression with respect to c gives us

$$\frac{\partial L^*}{\partial c} = -\left(\frac{1}{2}\right)\left(\frac{1}{c}\right)^2 + \left(\frac{1}{2}\right)\left(\widetilde{\lambda}\right)^2 = \left(\frac{1}{2}\right)\left[\left(\widetilde{\lambda}\right)^2 - (\lambda^{**})^2\right].$$

Since $\widetilde{\lambda} > \lambda^{**}$, the bracketed expression is strictly positive, making L^* strictly increasing in c. *Q.E.D.*

A.5 Proof of Proposition 6

It is still the case that in a symmetric equilibrium, the employees have chosen the same effort level, denoted by ε. Consider employee i's incentive to choose the red project if his own signal indicates that it is prof table. Just as with perfectly correlated uncertainty, a necessary condition for the ρ-type employee to choose the red project is that it earns him the promotion if both the supervisor's signals indicate that red is a good choice and green is a bad choice. Again, denote the supervisor's net preference for the red project in this outcome by H_1. It is the difference between, on the one hand, the supervisor's conditional probability that the red project is prof table and the green one is not and, on the other, her conditional probability that the opposite is true:

$$\begin{aligned} H_1 \equiv\ & \Pr(Z_i = R, Z_j = R \mid e_i = \rho, e_j = \gamma, s_i = h, s_j = l) \\ & - \Pr(Z_i = G, Z_j = G \mid e_i = \rho, e_j = \gamma, s_i = h, s_j = l) \\ =\ & \frac{[(1-\mu)^2 + \kappa\mu(1-\mu)]\varepsilon(1-\varepsilon)\sigma^2 - [\mu^2 + \kappa\mu(1-\mu)](1-\varepsilon)\varepsilon(1-\sigma)^2}{(1-\mu)\varepsilon + \mu(1-\varepsilon)} \\ =\ & \frac{\varepsilon(1-\varepsilon)\{[(1-\mu)+\kappa\mu](1-\mu)\sigma^2 - [\mu + \kappa(1-\mu)]\mu(1-\sigma)^2\}}{(1-\mu)\varepsilon + \mu(1-\varepsilon)}\,. \end{aligned}$$

The expression in braces in the numerator determines the sign of H_1. It converges to the corresponding expression derived in Proposition 1 when κ approaches unity. Furthermore, it is independent of ε, strictly increasing in σ and κ, and strictly decreasing in μ. It therefore follows from the implicit-function theorem that an increase in κ shifts the boundary in the (σ, ε) parameter space where $H_1 = 0$ to the left, expanding region I where eff cient investment can occur.

Next, consider the ρ-employee's expected net benefi , again denoted by H_2, from choosing the red project rather than the green one given that $H_1 > 0$. With less than perfect correlation, if the employees choose the same project, then the promotion is split evenly if they receive the same signal and otherwise is given outright to the one with the high signal. If the employees choose different projects, then the promotion is given to the employee with the green project unless he himself receives a low signal and his competitor receives a high signal. Furthermore, notice that

$$\Pr(Z \mid e_i = \rho) = \frac{\Pr(e_i = \rho \mid Z)\Pr(Z)}{\Pr(e_i = \rho)}$$

and that the symmetry of the employees' signal process implies the following four equalities:

$$\begin{aligned} \Pr(s_i = l, s_j = h \mid z_i = g, z_j = r, Z) &= \Pr(s_i = h, s_j = l \mid z_i = r, z_j = g, Z), \\ \Pr(s_i = l, s_j = h \mid z_i = r, z_j = r, Z) &= \Pr(s_i = h, s_j = l \mid z_i = g, z_j = g, Z), \\ \Pr(s_i = h, s_j = h \mid z_i = r, z_j = r, Z) &= \Pr(s_i = l, s_j = l \mid z_i = g, z_j = g, Z), \\ \Pr(s_i = l, s_j = l \mid z_i = r, z_j = r, Z) &= \Pr(s_i = h, s_j = h \mid z_i = g, z_j = g, Z)\,. \end{aligned}$$

This implies that, after some algebraic manipulation, the expression for H_2 simplifie to

$$
\begin{aligned}
H_2 &\equiv \mathrm{E}\{\pi_i(r) \mid \rho\} - \mathrm{E}\{\pi_i(g) \mid \rho\}\\
&= \frac{1}{\Pr(e_i=\rho)} \sum_Z \Pr(e_i=\rho \mid Z)\Pr(Z)\Bigg\{ \Pr(s_i=l, s_j=h \mid z_i=g, z_j=r, Z)\\
&\quad - \Pr(s_i=l, s_j=h \mid z_i=r, z_j=r, Z) - \frac{1}{2}\Pr(s_i=h, s_j=h \mid z_i=r, z_j=r, Z)\\
&\quad - \frac{1}{2}\Pr(s_i=l, s_j=l \mid z_i=r, z_j=r, Z)\Bigg\}\\
&= \left[\frac{1}{\varepsilon(1-\mu)+(1-\varepsilon)\mu}\right]\Bigg\{\big\{\varepsilon(1-\mu)[(1-\mu)+\kappa\mu]\sigma^2\\
&\quad + (1-\varepsilon)\mu[\mu+\kappa(1-\mu)](1-\sigma)^2\big\} - \frac{1}{2}\big\{\varepsilon(1-\mu)[(1-\mu)+\kappa\mu]\\
&\quad + (1-\varepsilon)\mu[\mu+\kappa(1-\mu)] + 2(1-\kappa)\mu(1-\mu)\big[\varepsilon(1-\sigma)^2+(1-\varepsilon)\sigma^2\big]\big\}\Bigg\}.
\end{aligned}
$$

H_2 is continuous in κ, and it approaches the corresponding expression derived in Proposition 2 as κ approaches unity. Moreover, it is straightforward to show that at the boundary where its value is equal to zero, H_2 is strictly increasing in ε, σ, and κ, and strictly decreasing in μ. It therefore follows from the implicit-function theorem that an increase in the correlation between the employees' sources of uncertainty shifts the boundary where $H_2 = 0$ down and to the left, expanding region I where effic ent investment can occur. *Q.E.D.*

A.6 Proof of Proposition 7

Consider fir t the case of $H_1 > 0$ and $H_2 < 0$. The compensation and the signal realizations for this case can be found in Figure 7. Compared to the tournament, an eff cient individual contract will adjust W_- and W_+. Denote the symmetric adjustment by a: $W_- \equiv W_2 + a$ and $W_+ \equiv W_1 - a$. Notice that $a = (1/2)[W_1 - W_2] \Rightarrow W_- = W_+ = (1/2)[W_1 + W_2] = W_0$.

Under the individual contract, the net benef t from choosing the green project, conditional on the employee's own signal, looks as follows:

$$
\begin{aligned}
&\mathrm{E}\{W_i(g) \mid e_i\} - \mathrm{E}\{W_i(r) \mid e_i\}\\
&\quad = \Big\{ \Pr(e_j=\gamma \mid e_i)W_0 + \Pr(e_j=\rho, s_i=h, s_j=l \mid e_i)W_{+++}\\
&\qquad + \Pr(e_j=\rho, s_i=s_j \mid e_i)W_+ + \Pr(e_j=\rho, s_i=l, s_j=h \mid e_i)W_{--}\Big\}\\
&\qquad - \Big\{ \Pr(e_j=\rho \mid e_i)W_0 + \Pr(e_j=\gamma, s_i=h, s_j=l \mid e_i)W_{++}\\
&\qquad + \Pr(e_j=\gamma, s_i=s_j \mid e_i)W_- + \Pr(e_j=\gamma, s_i=l, s_j=h \mid e_i)W_{---}\Big\}\\
&\quad = \Big\{ \Pr(e_j=\gamma \mid e_i)\frac{1}{2}[W_1+W_2] + \Pr(e_j=\rho, s_i=h, s_j=l \mid e_i)W_1
\end{aligned}
$$

$$+ \Pr(e_j = \rho, s_i = s_j \mid e_i)[W_1 - a] + \Pr(e_j = \rho, s_i = l, s_j = h \mid e_i) W_2 \Big\}$$

$$- \Big\{ \Pr(e_j = \rho \mid e_i) \frac{1}{2}[W_1 + W_2] + \Pr(e_j = \gamma, s_i = h, s_j = l \mid e_i) W_1$$

$$+ \Pr(e_j = \gamma, s_i = s_j \mid e_i)[W_2 + a] + \Pr(e_j = \gamma, s_i = l, s_j = h \mid e_i) W_2 \Big\} .$$

Some algebraic manipulation gives

$$\begin{aligned} &\mathrm{E}\{W_i(g) \mid e_i\} - \mathrm{E}\{W_i(r) \mid e_i\} \\ &\quad = [W_1 - W_2] \Big\{ \frac{1}{2} - \big[\Pr(e_j = \rho, s_i = l, s_j = h \mid e_i) \\ &\qquad + \Pr(e_j = \gamma, s_i = h, s_j = l \mid e_i) \big] \Big\} - a \Pr(s_i = s_j \mid e_i) . \end{aligned}$$

This expression is continuous and strictly decreasing in a. Moreover, since $H_2 < 0$, it is strictly positive for both employee types when $a = 0$. Some algebra conf rms that when $a = (1/2)[W_1 - W_2]$, the sign of the net benef t from choosing green over red depends on the employee's signal, being strictly positive if $e_i = \gamma$ and strictly negative if $e_i = \rho$:

$$a = \frac{1}{2}[W_1 - W_2] \;\Rightarrow$$

$$\mathrm{E}\{W_i(g) \mid e_i\} - \mathrm{E}\{W_i(r) \mid e_i\} = [W_1 - W_2]\left(\sigma - \frac{1}{2}\right)[\Pr(G \mid e_i) - \Pr(R \mid e_i)] .$$

It therefore follows that there exists some cutoff value of a, $\hat{a} \in (0, (1/2)[W_1 - W_2])$, that makes the net benef t of choosing the green project with a ρ-signal vanish and that defi es a range of individual contracts that can implement eff cient investment: $a^* \in [\hat{a}, (1/2)[W_1 - W_2]]$.

Consider next the case of $H_1 < 0$. The compensation and the signal realizations for this case can be found in Figure 8. Compared to the tournament, an efficien individual contract will now adjust W_{--} and W_{++} as well as W_- and W_+. Denote the symmetric adjustment to W_{--} and W_{++} by a and the one to W_- and W_+ by b: $W_{--} \equiv W_2 + a$, $W_{++} \equiv W_1 - a$, $W_- = W_2 + b$, and $W_+ = W_1 - b$. Notice that $a = (1/2)[W_1 - W_2] \Rightarrow W_{--} = W_{++} = (1/2)[W_1 + W_2] = W_0$. Moreover, $b = (1/2)[W_1 - W_2] \Rightarrow W_- = W_+ = (1/2)[W_1 + W_2] = W_0$ and $b = [W_1 - W_2] \Rightarrow W_- = W_1$ and $W_+ = W_2$.

Under the individual contract, the net benef t from choosing the green project, conditional on the employee's own signal, looks as follows:

$$\begin{aligned} &\mathrm{E}\{W_i(g) \mid e_i\} - \mathrm{E}\{W_i(r) \mid e_i\} \\ &\quad = \Big\{ \Pr(e_j = \gamma \mid e_i) W_0 + \Pr(e_j = \rho, s_i = h, s_j = l \mid e_i) W_{+++} \\ &\qquad + \Pr(e_j = \rho, s_i = s_j \mid e_i) W_{++} + \Pr(e_j = \rho, s_i = l, s_j = h \mid e_i) W_+ \Big\} \\ &\qquad - \Big\{ \Pr(e_j = \rho \mid e_i) W_0 + \Pr(e_j = \gamma, s_i = h, s_j = l \mid e_i) W_- \end{aligned}$$

$$
\begin{aligned}
&\quad + \Pr(e_j = \gamma, s_i = s_j \mid e_i) W_{--} + \Pr(e_j = \gamma, s_i = l, s_j = h \mid e_i) W_{---} \Big\} \\
&= \Big\{ \Pr(e_j = \gamma \mid e_i) \frac{1}{2}[W_1 + W_2] + \Pr(e_j = \rho, s_i = h, s_j = l \mid e_i) W_1 \\
&\quad + \Pr(e_j = \rho, s_i = s_j \mid e_i)[W_1 - a] + \Pr(e_j = \rho, s_i = l, s_j = h \mid e_i)[W_1 - b] \Big\} \\
&\quad - \Big\{ \Pr(e_j = \rho \mid e_i) \frac{1}{2}[W_1 + W_2] + \Pr(e_j = \gamma, s_i = h, s_j = l \mid e_i)[W_2 + b] \\
&\quad + \Pr(e_j = \gamma, s_i = s_j \mid e_i)[W_2 + a] + \Pr(e_j = \gamma, s_i = l, s_j = h \mid e_i) W_2 \Big\}.
\end{aligned}
$$

Some algebraic manipulation gives

$$
\begin{aligned}
&\mathrm{E}\{W_i(g) \mid e_i\} - \mathrm{E}\{W_i(r) \mid e_i\} \\
&\quad = \frac{1}{2}[W_1 - W_2] - a \Pr(s_i = s_j \mid e_i) \\
&\qquad - b[\Pr(e_j = \rho, s_i = l, s_j = h \mid e_i) + \Pr(e_j = \gamma, s_i = h, s_j = l \mid e_i)].
\end{aligned}
$$

This expression is continuous and strictly decreasing in both a and b. Moreover, since $H_1 < 0$, it is strictly positive for both employee types when $a = b = 0$. Some algebra conf rms that when $a = b = (1/2)[W_1 - W_2]$, the net benef t from choosing green over red remains strictly positive for both types of employees:

$$
a = b = \frac{1}{2}[W_1 - W_2] \;\Rightarrow
$$

$$
\begin{aligned}
\mathrm{E}\{W_i(g) \mid e_i\} - \mathrm{E}\{W_i(r) \mid e_i\} &= \frac{1}{2}[W_1 - W_2]\big\{1 - \big[\Pr(s_i = s_i \mid e_i) \\
&\quad + \Pr(e_j = \rho, s_i = l, s_j = h \mid e_i) + \Pr(e_j = \gamma, s_i = h, s_j = l \mid e_i)\big]\big\} > 0\,.
\end{aligned}
$$

However, when $a = (1/2)[W_1 - W_2]$ and $b = [W_1 - W_2]$, the sign of the net benef t from choosing green over red depends on the employee's signal, being strictly positive if $e_i = \gamma$ and strictly negative if $e_i = \rho$:

$$
a = \frac{1}{2}[W_1 - W_2], b = [W_1 - W_2] \;\Rightarrow
$$

$$
\mathrm{E}\{W_i(g) \mid e_i\} - \mathrm{E}\{W_i(r) \mid e_i\} = [W_1 - W_2]\left(\sigma - \frac{1}{2}\right)[\Pr(G \mid e_i) - \Pr(R \mid e_i)]\,.
$$

It therefore follows that there exists some cutoff value of b, $\hat{b} \in ((1/2)[W_1 - W_2], [W_1 - W_2])$, that, in conjunction with $a = (1/2)[W_1 - W_2]$, makes the net benefi of choosing the green project with a ρ-signal vanish and that def nes a range of individual contracts that can implement eff cient investment: $a^* = (1/2)[W_1 - W_2]$ and $b^* \in [\hat{b}, [W_1 - W_2]]$.

The same type of adjustment to the tournament can implement eff cient investment when $H_1 = 0$, but we omit this nongeneric case to economize on space. *Q.E.D.*

References

Baik, K. H. (1998), "Difference-Form Contest Success Functions and Effort Levels in Contests," *European Journal of Political Economy*, 14(4), 685–701.

Baker, G. P., M. C. Jensen, and K. J. Murphy (1988), "Compensation and Incentives: Practice vs. Theory," *The Journal of Finance*, Papers and Proceedings, 43(3), 593–616.

Banks, J. S., and J. Sobel (1987), "Equilibrium Selection in Signaling Games," *Econometrica*, 55(3), 647–661.

Bernhardt, D. (1995), "Strategic Promotion and Compensation," *The Review of Economic Studies*, 62(2), 315–339.

Bhattacharya, S. (1983), "Tournaments and Incentives: Heterogeneity and Essentiality," Case Study No. 695, Graduate School of Business, Stanford University.

Brandenburger, A., and B. Polak (1996), "When Managers Cover their Posteriors: Making the Decisions the Market Wants to See," *The RAND Journal of Economics*, 27(3), 523–541.

Carmichael, L. (1983), "Firm-Specifi Human Capital and Promotion Ladders," *The Bell Journal of Economics*, 14(1), 251–258.

Cummins, J. G., and I. Nyman (2005), "The Dark Side of Competitive Pressure," *The RAND Journal of Economics*, 36(2), 361–377.

Demougin, D., and C. Fluet (2006), "Preponderance of Evidence," *European Economic Review*, 50(4), 963–976.

Drago, R., and G. T. Garvey (1998), "Incentives for Helping on the Job: Theory and Evidence," *Journal of Labor Economics*, 16(1), 1–25.

Dye, R. A. (1984), "The Trouble with Tournaments," *Economic Inquiry*, 22(1), 147–149.

Fairburn, J. A., and J. M. Malcomson (2001), "Performance, Promotion, and the Peter Principle," *The Review of Economic Studies*, 68(1), 45–66.

Gentzkow, M., and J. M. Shapiro (2006), "Media Bias and Reputation," *Journal of Political Economy*, 114(2), 280–316.

Gibbons, R. (1997), "Incentives and Careers in Organizations," in: D. M. Kreps and K. F. Wallis (eds.), *Advances in Economics and Econometrics: Theory and Applications, Seventh World Congress, Volume 2*, Cambridge University Press, New York, pp. 1–37.

— and M. Waldman (1999), "Careers in Organizations: Theory and Evidence," in: O. C. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics, Volume 3B*, North-Holland, Amsterdam, pp. 2373–2437.

Heidhues, P., and J. Lagerlöf (2003), "Hiding Information in Electoral Competition," *Games and Economic Behavior*, 42(1), 48–74.

Hirshleifer, J. (1989), "Conflic and Rent-Seeking Success Functions: Ratio vs. Difference Models of Relative Success," *Public Choice*, 63(2), 101–112.

Holmström, B. (1982), "Moral Hazard in Teams," *The Bell Journal of Economics*, 13(2), 324–340.

Kim, J.-Y., and K. Ruy (2003), "Yes-Men and No-Men: Does Defianc Signal Talent?" *Journal of Institutional and Theoretical Economics (JITE)*, 159(3), 468–490.

Koch, A. K., and E. Peyrache (2005), "Tournaments, Individualized Contracts, and Career Concerns," Discussion Paper No. 1841, IZA, Bonn.

Lazear, E. P. (1989), "Pay Equality and Industrial Politics," *Journal of Political Economy*, 97(3), 561–580.

— and S. Rosen (1981), "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, 89(5), 841–864.

Malcomson, J. M. (1984), "Work Incentives, Hierarchy, and Internal Labor Markets," *Journal of Political Economy*, 92(3), 486–507.

Marino, A. M., and J. Zábojník (2004), "Internal Competition for Corporate Resources and Incentives in Teams," *The RAND Journal of Economics*, 35(4), 710–727.

McLaughlin, K. J. (1988), "Aspects of Tournament Models: A Survey," in: R. G. Ehrenberg (ed.), *Research in Labor Economics, Volume 9*, JAI Press, Greenwich (CT), pp. 225–256.

Nalebuff, B. J., and J. E. Stiglitz (1983), "Prizes and Incentives: Towards a General Theory of Compensation and Competition," *The Bell Journal of Economics*, 14(1), 21–43.
Posner, R. A. (1999), "An Economic Approach to the Law of Evidence," *Stanford Law Review*, 51(6), 1477–1546.
Prendergast, C. (1993a), "The Role of Promotion in Inducing Specifi Human Capital Acquisition," *The Quarterly Journal of Economics*, 108(2), 523–534.
— (1993b), "A Theory of 'Yes Men'," *The American Economic Review*, 83(4), 757–770.
— (1999), "The Provision of Incentives in Firms," *Journal of Economic Literature*, 37(1), 7–63.
— and L. Stole (1996), "Impetuous Youngsters and Jaded Old-Timers: Acquiring a Reputation for Learning," *Journal of Political Economy*, 104(6), 1105–1134.
Ricart i Costa, J. E. (1988), "Managerial Task Assignment and Promotions," *Econometrica*, 56(2), 449–466.
Waldman, M. (1984), "Job Assignments, Signalling, and Eff ciency," *The RAND Journal of Economics*, 15(2), 255–267.
Zábojník, J. (2002), "Centralized and Decentralized Decision Making in Organizations," *Journal of Labor Economics*, 20(1), 1–22.
— and D. Bernhardt (2001), "Corporate Tournaments, Human Capital Acquisition, and the Firm Size–Wage Relation," *The Review of Economic Studies*, 68(3), 693–716.

Jason G. Cummins
Brevan Howard Asset Management
1776 I Street, NW
Suite 250
Washington, DC 20006
U.S.A.
jason.cummins@brevanhoward.com

Ingmar Nyman
Department of Economics
Hunter College
695 Park Avenue
New York, NY 10065
U.S.A.
ingmar.nyman@hunter.cuny.edu

Dynamic Games of R&D Competition in a Differentiated Duopoly

by

Changying Li and Jianhu Zhang*

Received May 20, 2012; in revised form May 15, 2013; accepted June 4, 2013

Based on a Hotelling-type model, this paper analyzes a differential game where two f rms engage in quality-enhancing research and development (R&D). The analysis is formulated in terms of open-loop and feedback solutions. We f nd that the open-loop stationary levels of R&D and quality are socially eff cient. Moreover, compared to open-loop strategies, feedback strategies lead to higher stationary levels of prices and prof ts, but lower levels of R&D, quality, consumer surplus, and social welfare. In addition, compared to the social optimum, both open-loop and feedback strategies yield a closer stationary distance between the two fir s. (JEL: C73, L13, D43, D92)

1 Introduction

It is now widely recognized that, for consumer products, market values decrease very quickly over time. In order to boost sales and market values, f rms must spend effort in improving their product qualities. For example, mobile-phone manufacturers have been trying to introduce new features into their new-generation phones. Signif cant features include increased capacity and faster speed, as well as convenient services such as multimedia, videoconferencing, web browsing, e-mail, fax, and navigational maps.[1] Computer processing technology is also advancing rapidly. Processor speeds are growing faster all the time, and their sophistication is increasing with every

* Changying Li (corresponding author) and Jianhu Zhang: School of Economics, Shandong University, P. R. China. We thank Gerd Muehlheusser and three anonymous referees for their valuable comments. We would like to thank Hong Hwang, Wen-Jung Liang, Yan-Shu Lin, and seminar participants at National Taiwan University for their helpful comments. Changying Li gratefully acknowledges f nancial support from the Shandong Science Foundation (ZR2012GM017) and Shandong University (IFYT1217). Jianhu Zhang acknowledges f nancial support from CPSF (2011M501100, 2012T50599) and IIFSU (2011GN017, IFYT12069). We, of course, retain responsibility for any remaining errors.

[1] http://ieeemacau.eee.umac.mo/ieee_student/history of mobile phone.htm.

Journal of Institutional and Theoretical Economics 169, 660–679 – ISSN 0932-4569
DOI: 10.1628/093245613X671184 – © 2013 Mohr Siebeck

enhancement. Another example is the hard-drive industry. The price and physical size of hard drives are decreasing year in and year out, while their capacity and speed are increasing dramatically with every innovation.

Based on a Hotelling-like framework,[2] this paper develops a differentiated duopoly model where two firms invest in improving their product quality, which deteriorates at a certain constant proportional rate. Changes in product quality thus depend on both firm's R&D effort and the depreciation rate. Each firm chooses paths of price, R&D, and location and thus induces product-quality and market-share paths so as to maximize its total discounted profit. Our analysis is conducted within the context of a differential game in which both open-loop and feedback solutions are formulated.

We find that the open-loop stationary levels of R&D and quality are socially efficient. By contrast, the feedback models lead to higher levels of profits, but lower levels of R&D, consumer surplus, and social welfare. Moreover, under the two information structures, a firm with higher initial quality will conduct more R&D and choose a location closer to the midpoint, in order to keep its advantage during the process of market adjustment. However, quality levels are equal and locations are symmetric at the stationary points. Indeed, there is some evidence that supports our theoretical predictions. For example, in the manufacturing of processors, Intel had a quality advantage over its competitors at the beginning of the 1970s; however, other firms, such as Advanced Micro Devices (AMD), were quickly catching up.

By using a similar Hotelling-plus-quality model, Brekke et al. (2010) also find that investment and quality are lower in the feedback solution than in the open-loop solution. However, they leave out price setting and location choice by firms, since prices are regulated and locations of firms are fixed at the opposite ends of the linear segment. In the current model, however, we allow firms to choose their prices, locations, and R&D investments.

In addition, the logic is different even though the two models yield similar results. In their model, quality investments are strategic complements. More specifically, an increase in a firm's R&D will induce a higher level of R&D by its competitor, which in turn will hurt the first firm. This business-stealing effect of quality improvement discourages firms' incentive to invest. In our model, however, quality investments are intertemporal strategic substitutes. An increase in R&D effort by a firm will cause an investment-reducing response by its opponent. From the perspective of an investor, the instantaneous gain in market share by increasing the supply of its quality is to be weighed against the future loss of price reduction – a strategic reaction – by its rival firm. Since prices are strategic complements, the strategic interaction during the evolution process will drive the supply of quality to a lower level in the steady state, creating a collusive effect.

2 In most markets, products are horizontally differentiated. For example, computers have different platforms (microprocessor architecture), clothes have different fashions, and so on. The Hotelling approach allows us to capture this phenomenon.

A differential game approach is widely used to study the dynamic aspect of advertising (goodwill)[3] and capital accumulation.[4] The nature of the R&D process in this paper bears similarities with models that deal with advertising games. However, most of those papers assume that prices are independent of state variables (e.g., Bass et al., 2005, Nair and Narasimhan, 2006). The current model, in contrast, assumes that prices depend on firms’ R&D investments. Moreover, the Hotelling framework allows us to capture the interactions between location choice and R&D investment.

The way in which product quality deteriorates over time is reminiscent of the literature on durable goods (e.g., Bulow, 1982, 1986, and Waldman, 1993, 1996). In order to reduce the intertemporal consistency problem associated with selling, durable goods’ producers may have incentives to eliminate the market for second-hand goods (i.e., planned obsolescence). This paper, however, assumes that the current demand for a good depends only on its current level of quality and price, and thus is independent of its past sales. One justification for this assumption is the arrival of new consumers (i.e., population change). Another justification could be the change in consumer tastes. The focus of our analysis is, therefore, both on firms’ incentives to conduct R&D and on their location choice, rather than on the intertemporal consistency problem.

Our paper is related to the literature that deals with R&D competition. Early contributions focus on either the relationship between market structure and firms’ R&D incentives[5] or strategic R&D between competitors.[6] None of these models, however, considers the cumulative effect of R&D (i.e., a firm’s product quality depends on its past and current R&D investments), which is one focus of the present paper.[7] Another novel feature of our model is that it incorporates a Hotelling framework, and thus generates some results that differ from those of previous models.

Harter (1993) adds a R&D process to a product-variety model and looks at Hotelling’s principle of minimum differentiation. Lambertini (2002) studies the entry process in a spatial market over an infinite time horizon. However, the setups used in those papers are basically static rather than dynamic. The static nature of these models rules out the dynamic aspect of quality accumulation, and thus neglects the dynamic interaction between firm ’ location choice and R&D strategy. Two recent papers by Narajabad and Watson (2008 and 2011) investigate the relationship between innovation and the dynamics of location in Hotelling models. In their models, however, firms are restricted to be located at either end of the Hotelling

---

[3] See Friedman (1983), Fershtman (1984), Erickson (1995), Piga (1998, 2000), and Bertuzzi and Lambertini (2010).

[4] See, for instance, Fershtman and Muller (1984) and Reynolds (1987).

[5] See, among others, Dasgupta and Stiglitz (1980), Lee and Wilde (1980), and Reinganum (1985).

[6] See, for example, Brander and Spencer (1983), Katz (1986), D’ Aspremont and Jacquemin (1988), Dixit (1988), and Kamien, Muller, and Zang (1992).

[7] Mukhopadhyay and Kouvelis (1997) use a differential game model to study the optimal design quality and pricing decisions over the life of the product. However, their results rest on numerical experimentations over a variety of model parameters.

line, and rivalry in quality upgrades takes the form of a quality ladder, in which the size of the step is an exogenous parameter.

Under the assumption that product development is coupled with technical risk, Li and Zhang (2012) build a spatial model where two f rms compete for developing a new product. An increase in one firm s R&D enhances its probability of succeeding. However, the model used in that paper is also static rather than dynamic, and it has nothing to do with quality accumulation. By contrast, R&D effort in this paper is used to improve product quality. The dynamic formulation allows us to model the quality accumulation and firms catch-up, which is absent in Li and Zhang (2012).

Our paper is also related to the literature on quality regulation. A major concern of this literature is the issue of regulation associated with quality verif ability and optimal contract (see, for example, Lewis and Sappington, 1991, 1992, and Auray, Mariotti, and Moizeau, 2011). Our analysis, however, departs from modeling information uncertainties, and focuses on R&D competition between Bertrand competitors, rather than on incentive contracts between the regulator and the agents.

The organization of the paper is as follows. The upcoming section presents the basic model. In section 3, we analyze the dynamic game of R&D under both open-loop and feedback models. In section 4, we consider the case of the socially effic ent outcome. Section 5 contains a summary of our analysis. Proofs are relegated to the Appendix.

2 *The Basic Model*

Consider a Hotelling-type market. The market is unit-length and exists over $t \in [0, +\infty)$. Two f rms, indexed by 1 and 2, produce the same physical good, and maximize their discounted prof ts. Throughout the time horizon, the two f rms have the same discounted rate $\rho \in (0, 1)$. Each f rm has a unit production cost $c \geq 0$. Suppose that f rm 1 locates at point $y_1(t)$ and that f rm 2 locates at point $1 - y_2(t)$.

At date t, f rm i engages in product innovation with an investment, $\alpha I_i^2(t)/2$, to improve the quality of its product, $\alpha > 0$. The parameter α reflec s f rms' R&D effic ency. $I_i(t)$ is f rm i's R&D effort at date t. Quality evolves over time according to the following equation:[8]

$$(1) \qquad \frac{dX_i(t)}{dt} = X_i'(t) = I_i(t) - \delta X_i(t), \quad X_i(0) = X_{i0}, \quad i = 1, 2,$$

where $\delta \in (0, 1)$ is a constant proportional depreciation rate, $X_i(t)$ is f rm i's quality at date t, and X_{i0} is f rm i's initial quality at the beginning of the time period.

Equation (1) implies that quality depreciates over time, and thus in order to improve the quality, sustained effort is required. Equation (1) also implies that a f rm's effort has a long-lasting impact. As a result, the current quality level depends on its past levels and thus, in turn, depends on the f rm's continual efforts.

[8] Similar dynamic relations have been used in the literature on capital accumulation and quality evolution (e.g., Fershtman and Muller, 1984, Auray, Mariotti, and Moizeau, 2011).

Consumers are uniformly distributed over the interval $[0, 1]$. The population is normalized to one without loss of generality. Each consumer buys at most one unit of the good. The utility of a consumer located at $x \in [0, 1]$ who will buy from f rm i is

$$S + X_i(t) - p_i(t) - d\left|x - \left[(1 + (-1)^i)/2 - (-1)^i y_i(t)\right]\right|, \quad i = 1, 2,$$

where $S > 0$ is the reservation price obtained by any consumer who buys from either of the two firm . Assume that S is large enough so that all consumers buy. Obviously, the higher the quality induced by a f rm's R&D effort, the higher the consumer's willingness to pay.[9] $p_i(t)$ is f rm i's product price at date t, and d is a unit-length transportation cost.

Following Bertuzzi and Lambertini (2010, p. 157), we assume that changing location is costly. Such costs can be interpreted as f rms' relocating cost along the linear market, or as R&D cost associated with product differentiation. In particular, fir i bears the following location cost:

$$\gamma_i[y_i(t)] = \beta[\eta - y_i(t)]^2, \quad i = 1, 2,$$

where $y_1(0) = \eta$ and $y_2(0) = 1 - \eta$. As noted by Bertuzzi and Lambertini (2010, p. 157), "points η and $1 - \eta$ can be interpreted as representing given product specif-cations that are public domain and therefore costless to supply, or the geographical locations of firm along the city, that owners may change at a cost."

In order to ensure the existence of equilibrium and to ensure that the second-order conditions are satisfie , we assume that

$$\beta > \max\left\{\frac{d(\delta + 2\rho)}{6\rho}, \frac{\Lambda + \sqrt{\Lambda^2 - 24d^2(1 - 4\eta)}}{12(1 - 4\eta)}\right\}, \quad \Lambda = d(7 - 4\eta) + 2\left|X_{20} - X_{10}\right|.$$

Following Bertuzzi and Lambertini (2010), we further assume $\eta < 1/4$, in order for the game to yield a pure-strategy price equilibrium at $t = 0$.[10]

In order to guarantee the stability of the equilibrium and to eliminate the possibility of a monopoly, it is assumed that $\alpha > (3A)/(2\delta(\delta + \rho))$ and $(X_{10}, X_{20}) \in \Omega$, where $A = ((8\beta - d)\beta)/(d(d - 6\beta)^2))$ and

$$\Omega = \left\{X(t) \equiv (X_1(t), X_2(t)) \,\|\, X_1(t) - X_2(t)\right| \le 3d - d^2/(2\beta)\right\}. \tag{2}$$

Equation (2) implies that the difference between the two f rms' initial qualities is not too large; otherwise, the low-quality f rm will be driven out of the market. The purpose of this paper is to investigate issues related to dynamic R&D competition. Therefore, it is useful to rule out the case of a monopoly.

---

9 An increase in quality acts very much like an increase in the reservation price, and therefore the investment to increase quality here is very close to an advertising campaign (Lambertini, 2005). Thanks to a referee for suggesting this point.

10 As pointed by Bertuzzi and Lambertini (2010, p. 157, footnote 12), "if $\eta \in [1/4, 3/4]$ and f rms chose not to modify their locations at the initial instant, then clearly the undercutting argument would destroy the price equilibrium at the very outset of the game."

At the date t, the marginal consumer, who is indifferent between purchasing from either firm is located at $\hat{x}(t)$ as given by

$$\hat{x}(t) = \big(p_2(t) - p_1(t) + d + dy_1(t) - dy_2(t) + X_1(t) - X_2(t)\big)/(2d)\,.$$

Aggregate demand for firm 1 and 2 is, respectively, as follows:

$$D_1(t) = \min\left\{\max\left\{\hat{x}(t), 0\right\}, 1\right\} \quad \text{and} \quad D_2(t) = 1 - D_1(t)\,. \tag{3}$$

The instantaneous prof t function of f rm i is

$$\pi_i(t) = [p_i(t) - c]D_i(t) - \beta[\eta - y_i(t)]^2 - \alpha \cdot I_i^2(t)/2\,.$$

Hence, the problem of f rm i is

$$\max_{\{p_i(t), y_i(t), I_i(t)\}} G^i = \int_0^\infty \pi_i(t) \cdot e^{-t \cdot \rho} dt\,, \tag{4}$$

$$\begin{aligned} \text{subject to} \quad & X_i'(t) = I_i(t) - \delta X_i(t), \\ & X_j'(t) = I_j(t) - \delta X_j(t), \\ & X_i(0) = X_{i0}, \quad X_j(0) = X_{j0}, \quad i, j = 1, 2\,, \end{aligned}$$

where $p_i(t)$, $y_i(t)$, and $I_i(t)$ are decision variables, and $X_i(t)$ is the state variable.

3 Model Analysis

According to the information structure, two strategies will be investigated. One is the open-loop strategy and the other is the feedback strategy. Under the open-loop rule, each fir chooses its investment and location strategy as well as price schedule at the initial date, and commits to them forever.[11] Therefore, open-loop strategies vary with time only and thus do not allow for strategic possibilities between f rms through the evolution of qualities over time. In contrast, the feedback rule does allow firm to choose their strategies by taking into account their strategic interactions, i.e., feedback strategies are functions not only of time, but also of f rms' qualities over time.

3.1 Open-Loop Solution

In this subsection, we will examine a situation where each fir solves its prof t-maximizing problem (as def ned by equation (4)), taking the action path of its rival as given. The Hamiltonian function of f rm $i \in \{1, 2\}$ is

$$H_i(t) = D_i(t)[p_i(t) - c] - \beta[\eta - y_i(t)]^2 - \alpha \cdot I_i^2(t)/2 + \lambda_i(t)[I_i(t) - \delta X_i(t)]\,, \tag{5}$$

where $\lambda_i(t)$ is the costate variable associated with $X_i(t)$. The outcome of the open-loop model can be summarized as follows:

[11] One possible explanation for such a precommitment is that a rule of the game requires precommitment, which is not observable after the start of the game (Reynolds, 1987; Wirl, 2010). Another possible explanation is that such a precommitment is included in a binding contract signed with a third party (e.g., the government) (Hanig, 1987).

PROPOSITION 1 *There is a unique open-loop Nash equilibrium:*

$$I_i^{\mathrm{OL}}(t) = (-1)^i \cdot \Psi(\delta - \mu^{\mathrm{OL}}) \cdot e^{-t\cdot\mu^{\mathrm{OL}}} + I^{\mathrm{OL}}, \quad i = 1,2\,,$$

$$p_i^{\mathrm{OL}}(t) = c + d + (-1)^i \cdot (4\beta\Psi)/(6\beta - d) \cdot e^{-t\cdot\mu^{\mathrm{OL}}}, \quad i = 1,2\,, \tag{6}$$

$$y_i^{\mathrm{OL}}(t) = d/(4\beta) + \eta + (-1)^i \cdot \Psi/(6\beta - d) \cdot e^{-t\cdot\mu^{\mathrm{OL}}}, \quad i = 1,2\,, \tag{7}$$

where

$$\delta > \mu^{\mathrm{OL}} = \left(\sqrt{(2\delta+\rho)^2 - 8\beta/(d\alpha(6\beta - d))} - \rho\right)/2 > 0, \quad \Psi = (X_{20} - X_{10})/2\,.$$

Firm i's quality path is

$$X_i^{\mathrm{OL}}(t) = \Phi \cdot e^{-t\cdot\delta} + (-1)^i \cdot \Psi \cdot e^{-t\cdot\mu^{\mathrm{OL}}} + X^{\mathrm{OL}}, \quad i = 1,2\,,$$

where $\Phi = (X_{20} + X_{10} - 2X^{\mathrm{OL}})/2$. *Firm 1's market share is*

$$D_1^{\mathrm{OL}}(t) = 1/2 - (2\beta\Psi)/(6d\beta - d^2) \cdot e^{-t\cdot\mu^{\mathrm{OL}}}\,.$$

I^{OL} *and* X^{OL} *are the steady-state equilibrium R&D and quality, i.e.,*

$$I^{\mathrm{OL}} = 1/(2\alpha(\delta+\rho)), \quad X^{\mathrm{OL}} = I^{\mathrm{OL}}/\delta\,. \tag{8}$$

Proposition 1 implies that open-loop strategies act as a convergence force during the process of market adjustment. A f rm with a higher initial quality will start with a higher R&D investment, to increase its market share to a certain level, and then lower its R&D effort gradually over time; in contrast, the rival f rm will begin with a lower R&D expenditure, to allow for a decline in its market share, and then increase its R&D effort over time. In other words, the quality advantage of a f rm with superior technology will remain for a long time, but it diminishes gradually over time. The dynamics of price adjustment exhibit a similar pattern to the dynamic adjustment of quality.

In addition, a f rm with a higher initial quality will f rst choose a location close to the midpoint and then move back to the endpoint; in contrast, the rival f rm will begin with a location close to the endpoint and then move toward the midpoint. In the steady state, the two f rms have symmetric locations. During the process of market adjustment, the distance between the two firm keeps constant.

3.2 Feedback Solution

In this subsection, we will analyze the feedback equilibrium, in which f rms will not precommit to investment paths at the beginning of the time period; instead, they will behave strategically, and therefore, their prices, locations, and R&D decisions are contingent on state variables over time. The feedback solutions satisfy the following

Hamilton–Jacobi–Bellman equation:

$$\rho V_i = \max_{\{p_i(t), y_i(t), I_i(t)\}} \Big\{ D_i(t)[p_i(t) - c] - \beta[\eta - y_i(t)]^2 - \alpha \cdot I_i^2(t)/2 \tag{9}$$
$$+ \frac{\partial V_i}{\partial X_i(t)}(I_i(t) - \delta X_i(t)) + \frac{\partial V_i}{\partial X_j(t)}(I_j(t) - \delta X_j(t)) \Big\},$$

$$i, j = 1, 2, \quad i \neq j\,.$$

PROPOSITION 2 *There is a unique feedback Nash equilibrium. The pair of strategies is* $\{I_i^{\mathrm{FB}}(t), y_i^{\mathrm{FB}}(t), p_i^{\mathrm{FB}}(t)\}$, *where*

$$I_i^{\mathrm{FB}}(t) = I^{\mathrm{FB}} + X_i^{\mathrm{FB}}(t) \cdot (\delta - \mu^{\mathrm{FB}})/3 - X_j^{\mathrm{FB}}(t) \cdot (\delta - \mu^{\mathrm{FB}})/3, \qquad i, j = 1, 2, \quad i \neq j\,, \tag{10}$$

$$y_i^{\mathrm{FB}}(t) = d/(4\beta) + \eta + (X_i^{\mathrm{FB}}(t) - X_j^{\mathrm{FB}}(t))/(12\beta - 2d), \quad i = 1, 2, \quad i \neq j\,, \tag{11}$$

$$p_i^{\mathrm{FB}}(t) = c + d + 2\beta(X_i^{\mathrm{FB}}(t) - X_j^{\mathrm{FB}}(t))/(6\beta - d), \quad i, j = 1, 2, \quad i \neq j\,; \tag{12}$$

$$\delta > \mu^{\mathrm{FB}} = \Big(\sqrt{(2\delta + \rho)^2 - (6\beta(8\beta - d))/(d\alpha(6\beta - d)^2)} - \rho\Big)/2 > 0\,.$$

The induced paths of quality and market share are $X_i^{\mathrm{FB}}(t)$ *and* $D_1^{\mathrm{FB}}(t)$, *where*

$$X_i^{\mathrm{FB}}(t) = \mathrm{T}e^{-t\cdot\delta} + (-1)^i \cdot \Psi e^{-t(\delta + 2\mu^{\mathrm{FB}})/3} + X^{\mathrm{FB}}\,,$$
$$\mathrm{T} = (X_{20} + X_{10} - 2X^{\mathrm{FB}})/2, \quad i = 1, 2\,,$$

$$D_1^{\mathrm{FB}}(t) = 1/2 - e^{-t(\delta + 2\mu^{\mathrm{FB}})/3} \cdot (2\beta\Psi)/(6d\beta - d^2)\,.$$

The steady-state levels of investment and quality are, respectively,

$$I^{\mathrm{FB}} = 3(8\beta - d)/(4\alpha(6\beta - d)(2\delta + \mu^{\mathrm{FB}} + 3\rho)) \quad \textit{and} \quad X^{\mathrm{FB}} = I^{\mathrm{FB}}/\delta\,.$$

Note that $\delta - \mu^{\mathrm{FB}} > 0$; thus (10) implies that a firm's R&D investment at date t depends on the level of its own quality and of its rival's quality at that time. Specifically, firm i's R&D expenditure is positively related to its own quality level and negatively related to firm j's quality level, which suggests that firms' R&D investments are *intertemporal strategic substitutes* (Jun and Vives, 2004).[12] In a static model, strategic substitutability implies that $(\partial^2\pi_i(X))/(\partial X_i \partial X_j) < 0$. In the current dynamic model, equation (A13) in Appendix 2 implies that $(\partial I_i(X))/(\partial X_j) = (1/\alpha) \cdot (\partial^2 V_i(X))/(\partial X_i \partial X_j)$. Therefore, equation (10) suggests that the static strategic substitutability of the model translates into intertemporal strategic substitutability in the dynamic game (with $(\partial^2 V_i(X))/(\partial X_i \partial X_j) < 0$).

Equation (11) shows that, with an increase in R&D, a firm will move toward the midpoint, in order to obtain a greater market share. This will force the rival firm to move away, in order to mute the effect of competition in the market. Equation (12) suggests that an increase in a firm's current R&D will decrease the price of its competitor.

[12] This dynamic property is called feedback substitutability in Figuieres (2002).

Under the feedback model, a firm with an initial quality advantage will enjoy this benefit over its rival firm during the process of market adjustment. However, this advantage will gradually diminish and finally disappear in the steady state. Therefore, the feedback strategies act as a convergence force, in terms of quality evolution. This pattern is similar to that generated by the open-loop strategies.

Direct comparison between the open-loop and the feedback equilibria leads to the following statement:

PROPOSITION 3 *(1) Compared to the open-loop strategies, the feedback strategies generate higher levels of stationary profits but lower levels of R&D, quality, consumer surplus, and social welfare. (2) For any initial levels of qualities,* $(X_{10}, X_{20}) \in \Omega$, $p_1^N(t) \geq p_2^N(t)$ *if* $X_{10} \geq X_{20}$, *and* $p_1^N(t) < p_2^N(t)$ *otherwise,* $N = \{\mathrm{OL}, \mathrm{FB}\}$. *However, the steady-state prices are equal, i.e.,* $\lim_{t\to+\infty} p_i^{\mathrm{OL}}(t) = \lim_{t\to+\infty} p_i^{\mathrm{FB}}(t) = c + d$. *(3) The stationary locations are symmetric, i.e.,* $\lim_{t\to+\infty} y_i^{\mathrm{OL}}(t) = \lim_{t\to+\infty} y_i^{\mathrm{FB}}(t) = d/(4\beta) + \eta$.

The first result can be explained as follows. With a feedback model, a firm's current investment will (1) preempt some amount of later investment by its rival firm[13] (2) force its opponent to move away, and (3) reduce its opponent's price. The first two effects encourage the firm to increase its investment, but the third effect forces the firm to reduce its investment, since prices are strategic complements. In the present model, the third effect dominates, which induces firms to cut their R&D, in order to soften market competition. This creates a *collusive effect* and produces less R&D in the feedback than in the open-loop equilibrium. The collusive effect makes firms better off but makes consumers and the society worse off.

The second result of Proposition 3 is not surprising. The price of the high-quality good is higher during the process of quality evolution, irrespective of the types of strategies. However, as the quality approaches its stationary point, the quality gap disappears, and thus the prices become equal in the steady state.

The result regarding symmetric locations can be explained as follows. It is well known that, in a spatial competition model, a firm has to balance the *market share effect* and the *competition effect* when choosing its location. In the current model, both of the effects are influenced by product qualities. In the steady state, the feedback strategy yields the same quality levels as the open-loop strategy, and thus leads to symmetric equilibrium locations.

It is interesting to compare the equilibria of the dynamic game with the equilibrium of the static one. The question is whether the dynamic setup adds anything to results derived for the multistage game (that is, firms first invest in quality, then choose locations, and finally set prices).[14] In order to guarantee that the second-order conditions are satisfied and to eliminate the possibility of a monopoly, it is assumed that

$$\alpha > \max\{3A/(2\delta(\delta + \rho)), ((18\beta - d)\beta)/(2d(d - 9\beta)^2)\}.$$

[13] This effect is also noted by Reynolds (1987).

[14] We thank the editor and a referee for suggesting this point.

The f rms' prof t function becomes

$$\pi_i = [p_i - c] \cdot D_i - \beta[\eta - y_i]^2 - \alpha \cdot X_i^2/2, \quad i = 1, 2\,.$$

Standard computations lead to the equilibrium results:

$$p^* = c + d = \lim_{t\to\infty} p_i^{\mathrm{OL}}(t), \quad X^* = (18\beta - d)/(6\alpha(9\beta - d)), \quad \text{and}$$

$$y^* = d/(6\beta) + \eta < \lim_{t\to\infty} y_i^{\mathrm{OL}}(t) = \lim_{t\to\infty} y_i^{\mathrm{FB}}(t) = d/(4\beta) + \eta\,.$$

It is easy to show that $X^* \geq$ or $\leq X^{\mathrm{FB}}$ and $X^* < X^{\mathrm{OL}}$.

COROLLARY 1 *Compared with the outcome of a static game, the equilibrium distance between the two firm is smaller in a dynamic setting. Moreover, the open-loop solution produces a higher R&D effort, while the feedback solution generates either a higher or a lower R&D effort.*

Compared with the outcome of the dynamic game, the static game leads to a less competitive outcome, where firm move away from each other, in order to invoke weaker competition in the price-setting stage.[15] *Conversely, the dynamic game produces a competitive outcome where f rms move closer, since f rms choose their investments, locations, and prices simultaneously, not sequentially.*

In the static competition model, R&D investment is a one-shot game, where f rms make their decisions at the outset of the game. However, in the dynamic setting, R&D investment takes place over a time horizon, and therefore, the R&D cost becomes lower. This *cost-eff ciency* effect explains the outcome that the open-loop solution produces a higher R&D effort relative to the static game. The feedback solution, however, might induce either a higher or a lower R&D effort, depending on the tradeoff between the cost-effic ency effect and the aforementioned collusive effect, caused by the dynamic strategic responses between the two players.

4 *Social Optimum*

In this section, we look for the socially effic ent case where social welfare (the sum of firm ' profi s and consumer surplus) achieves a maximal level. Based on the model in section 2, the instantaneous welfare function can be derived:

$$\begin{aligned} w = &\int_0^{D_1} [S + X_1 - p_1 - d|x - y_1|]dx + (p_1 - c)D_1 - \beta(\eta - y_1)^2 - \alpha I_1^2/2 \\ &+ \int_{D_1}^{1} [S + X_2 - p_2 - d|1 - y_2 - x|]dx + (p_2 - c)(1 - D_1) - \beta(\eta - y_2)^2 - \alpha I_2^2/2 \\ = &\, S - c - d\cdot(1/2 - D_1 + D_1^2) + D_1(X_1 - X_2) + X_2 - \alpha(I_1^2 + I_2^2)/2 \\ &+ d[y_1^2 + (y_2 - 1)y_2 + D_1(y_2 - y_1)] - \beta(\eta - y_1)^2 - \beta(\eta - y_2)^2\,. \end{aligned}$$

[15] If firm make their decisions in a one-stage game (that is, players choose their investments, locations, and prices simultaneously, not sequentially), the static game generates the same equilibrium locations, but different R&D levels, compared to the dynamic game.

The problem of a social planner can be written as

$$\max_{\{D_1(t), I_1(t), I_2(t), y_1(t), y_2(t)\}} W = \int_0^{\infty} w \cdot e^{-t\cdot\rho} dt\,,$$

$$\text{subject to} \quad \begin{aligned} &X_i'(t) = I_i(t) - \delta X_i(t),\\ &X_j'(t) = I_j(t) - \delta X_j(t),\\ &X_i(0) = X_{i0}, \quad X_j(0) = X_{j0}, \quad i, j = 1, 2\,, \end{aligned}$$

where $D_1(t)$, $I_1(t)$, $I_2(t)$, $y_1(t)$, and $y_2(t)$ are decision variables, and $X_1(t)$ and $X_2(t)$ are state variables.

Solving the welfare-maximization problem leads to the following statement:

PROPOSITION 4 *The socially optimal pair of strategies is* $\{D_1^S(t), I_i^S(t), y_i^S(t)\}$, *where*

$$\begin{aligned} &D_1^S(t) = 1/2 - e^{-t\cdot\mu^S} \cdot (2\Psi(d+\beta))/(d^2 + 2d\beta),\\ &I_i^S(t) = (-1)^i \Psi(\delta - \mu^S) \cdot e^{-t\cdot\mu^S} + I^S, \quad i = 1, 2, \quad \textit{and}\\ &y_i^S(t) = (d + 4\beta\eta)/(4d + 4\beta) + (-1)^i \cdot \Psi/(2\beta + d) \cdot e^{-t\cdot\mu^S}, \quad i = 1, 2;\\ &\delta > \mu^S = \left(\sqrt{(2\delta+\rho)^2 - 8(d+\beta)/(d^2\alpha + 2d\alpha\beta)} - \rho\right)/2 > 0\,. \end{aligned}$$

The quality path is

$$X_i^S(t) = \Phi \cdot e^{-t\cdot\delta} + (-1)^i \Psi \cdot e^{-t\cdot\mu^S} + X^S, \quad i = 1, 2\,.$$

The steady-state equilibrium levels of R&D and quality are, respectively,

$$I^S = 1/(2\alpha(\delta+\rho)) \quad \textit{and} \quad X^S = I^S/\delta\,. \tag{13}$$

The pattern of quality evolution for the social optimum is similar to that for the open-loop strategies, i.e., from a welfare perspective, it is more eff cient if a f rm with an initial quality advantage starts with a greater R&D effort, to increase its market share to a desirable level, and then reduces its R&D effort gradually over time; and the rival f rm moves in the opposite fashion. In the steady-state equilibrium, the two f rms have the same levels of R&D investments and qualities.

Direct comparison between equations (8) and (13) leads to the following statement:

COROLLARY 2 *The open-loop strategies generate socially effi ient stationary levels of R&D and quality.*

Direct comparison between Propositions 3 and 4 yields the following statement:

COROLLARY 3 *Compared to the socially effi ient location, the stationary distance between the two firm is closer under both open-loop and feedback strategies.*

PROOF

$$\lim_{t\to+\infty} y_i^N(t) - \lim_{t\to+\infty} y_i^S(t) = \frac{d}{4\beta} + \eta - \frac{d+4\beta\eta}{4(d+\beta)} = \frac{d(d+4\beta\eta)}{4\beta(d+\beta)} > 0, \quad N = \{\text{OL}, \text{FB}\}\,.$$

Q.E.D.

5 Conclusions

This paper introduces R&D competition by two Bertrand competitors into a differential game. The model is built on the work of Hotelling (1929), which enables us to study the effect of product differentiation on the equilibrium results. The paper shows that: First, the stationary levels of R&D and quality for the open-loop strategies are socially optimal. Second, the stationary feedback levels of R&D, quality, consumer surplus, and social welfare are always lower than their open-loop counterparts: the opposite then is true with regard to the prof ts. Third, regardless of the type of strategies, R&D investment remains higher for a f rm with an initial quality advantage during the market evolution; however, the levels of R&D investments and product qualities of the two f rms will ultimately converge to stationary points. Fourth, compared to the socially optimal location, the stationary distance between the two f rms is closer under the two models. Finally, for the two information structures, a f rm with higher initial quality will choose a location close to the midpoint, in order to keep its advantage during the process of market adjustment. However, the two firm ’ location will ultimately converge to a symmetric stationary state.

Appendix

A.1 Proof of Proposition 1

Consider the f rst-order conditions (FOCs) w.r.t. $p_i(t)$, $y_i(t)$, and $I_i(t)$, calculated using equation (5):

$$\text{(A1)} \qquad \frac{\partial H_i(t)}{\partial p_i(t)} = \frac{c + d + dy_i(t) - dy_j(t) - 2p_i(t) + p_j(t) + X_i(t) - X_j(t)}{2d} = 0\,,$$

$$\text{(A2)} \qquad \frac{\partial H_i(t)}{\partial y_i(t)} = \frac{p_i(t) - c + 4\beta[\eta - y_i(t)]}{2} = 0, \quad \text{and}$$

$$\text{(A3)} \qquad \frac{\partial H_i(t)}{\partial I_i(t)} = \lambda_i(t) - \alpha I_i(t) = 0\,.$$

The adjoint equations for the optimum are

$$\text{(A4)} \qquad \lambda_i'(t) = \rho\lambda_i(t) - \frac{\partial H_i(t)}{\partial X_i(t)} = (\rho+\delta)\lambda_i(t) + \frac{c - p_i(t)}{2d}, \quad i = 1,2\,.$$

The transversality conditions are

$$\text{(A5)} \qquad \lim_{t\to\infty} \lambda_i(t)\cdot X_i(t)\cdot e^{-t\cdot\rho} = 0, \quad i = 1,2\,.$$

Solving equations (A1)–(A3) yields

$$\text{(A6)} \qquad p_i(t) = c + d + \frac{2\beta}{6\beta - d}[X_i(t) - X_j(t)], \quad i = 1,2, \quad i \neq j\,,$$

$$\text{(A7)} \qquad y_i(t) = \frac{d}{4\beta} + \eta + \frac{1}{2(6\beta - d)}[X_i(t) - X_j(t)], \quad i = 1,2, \quad i \neq j\,,$$

$$\text{(A8)} \qquad \lambda_i(t) = \alpha I_i(t), \quad i = 1,2\,.$$

Substituting equations (A6)–(A8) into equation (A4) leads to

$$\lambda_i'(t) = \frac{d^2 - 2\beta[3d + X_i(t) - X_j(t)]}{2d(6\beta - d)} + \alpha(\delta + \rho)I_i(t)\,.$$

Differentiating equation (A8) gives us

$$I_i'(t) = \frac{d^2 - 2\beta[3d + X_i(t) - X_j(t)]}{2d\alpha(6\beta - d)} + (\delta + \rho)I_i(t)\,. \tag{A9}$$

Combining equations (1) and (A9), we have

$$\begin{cases} I_i'(t) = \dfrac{d^2 - 2\beta[3d + X_i(t) - X_j(t)]}{2d\alpha(6\beta - d)} + (\delta + \rho)I_i(t) \\ X_i'(t) = I_i(t) - \delta X_i(t) \end{cases}, \quad i = 1, 2, \quad i \neq j\,. \tag{A10}$$

From $I_i'(t) = X_i'(t) = 0$, we obtain the steady-state equilibrium R&D and quality

$$I^{\mathrm{OL}} = \frac{1}{2\alpha(\delta + \rho)} \quad \text{and} \quad X^{\mathrm{OL}} = \frac{I^{\mathrm{OL}}}{\delta}\,.$$

Solving equation (A10) and taking into account equation (A5) leads to

$$\begin{bmatrix} I_1^{\mathrm{OL}}(t) \\ I_2^{\mathrm{OL}}(t) \\ X_1^{\mathrm{OL}}(t) \\ X_2^{\mathrm{OL}}(t) \end{bmatrix} = \Phi e^{-t\cdot\delta} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} + \Psi e^{-t\cdot\mu^{\mathrm{OL}}} \begin{bmatrix} -(\delta - \mu^{\mathrm{OL}}) \\ \delta - \mu^{\mathrm{OL}} \\ -1 \\ 1 \end{bmatrix} + \begin{bmatrix} I^{\mathrm{OL}} \\ I^{\mathrm{OL}} \\ X^{\mathrm{OL}} \\ X^{\mathrm{OL}} \end{bmatrix}, \tag{A11}$$

where

$$\delta > \mu^{\mathrm{OL}} = \frac{1}{2}\sqrt{(2\delta + \rho)^2 - \frac{8\beta}{d\alpha(6\beta - d)}} - \frac{\rho}{2} > 0\,.$$

Observe that the initial conditions imply

$$\Phi = \frac{1}{2}(X_{20} + X_{10} - 2X^{\mathrm{OL}}) \quad \text{and} \quad \Psi = \frac{1}{2}(X_{20} - X_{10})\,.$$

Combining equations (A6), (A7), and (A11), we have

$$\begin{aligned} p_i^{\mathrm{OL}}(t) &= c + d + (-1)^i \frac{4\beta\Psi}{6\beta - d} e^{-t\cdot\mu^{\mathrm{OL}}}, \quad i = 1, 2, \\ y_i^{\mathrm{OL}}(t) &= \frac{d}{4\beta} + \eta + (-1)^i \frac{\Psi}{6\beta - d} e^{-t\cdot\mu^{\mathrm{OL}}}, \quad i = 1, 2\,. \end{aligned} \tag{A12}$$

Now

$$\beta > \frac{\Lambda + \sqrt{\Lambda^2 - 24d^2(1 - 4\eta)}}{12(1 - 4\eta)}$$

implies that

$$y_1^{\mathrm{OL}}(t) < \frac{1}{4} \quad \text{and} \quad 1 - y_2^{\mathrm{OL}}(t) > \frac{3}{4}$$

at all times $t \in [0, \infty)$. If so, then the prices in equation (A12) are indeed equilibrium prices.

From equations (3), (A6), (A7), and (A11), we have

$$D_1^{\mathrm{OL}}(t) = \frac{1}{2} - \frac{2\beta\Psi}{6d\beta - d^2} e^{-t\cdot\mu^{\mathrm{OL}}}.$$

The steady-state equilibrium values are

$$D_1^{\mathrm{OL}} = \frac{1}{2}, \quad p_1^{\mathrm{OL}} = p_2^{\mathrm{OL}} = c + d, \quad y_1^{\mathrm{OL}} = y_2^{\mathrm{OL}} = \frac{d}{4\beta} + \eta,$$

$$I_1^{\mathrm{OL}} = I_2^{\mathrm{OL}} = I^{\mathrm{OL}} = \frac{1}{2\alpha(\delta+\rho)}, \quad \text{and} \quad X_1^{\mathrm{OL}} = X_2^{\mathrm{OL}} = X^{\mathrm{OL}} = \frac{1}{2\alpha\delta(\delta+\rho)}.$$

Firms' prof ts, consumer surplus, and social welfare are, respectively,

$$\pi_i^{\mathrm{OL}} = (p_i^{\mathrm{OL}} - c)D_i^{\mathrm{OL}} - \frac{\alpha}{2}(I_i^{\mathrm{OL}})^2 = \frac{1}{2}\left(d - \frac{1}{4\alpha(\delta+\rho)^2}\right), \quad i = 1, 2,$$

$$\begin{aligned} CS^{\mathrm{OL}} &= \int_0^{D_1^{\mathrm{OL}}} \left[S + X_1^{\mathrm{OL}} - p_1^{\mathrm{OL}} - d\left|x - y^{\mathrm{OL}}\right|\right]dx \\ &\quad + \int_{D_1^{\mathrm{OL}}}^1 \left[S + X_2^{\mathrm{OL}} - p_2^{\mathrm{OL}} - d\left|1 - y^{\mathrm{OL}} - x\right|\right]dx \\ &= S - c - \frac{5d}{4} + \frac{1}{2\alpha\delta(\delta+\rho)} - \frac{d[d^2 + 2d\beta(4\eta - 1) + 2\beta^2(5 - 4\eta + 8\eta^2)]}{8\beta^2}, \\ W^{\mathrm{OL}} &= \pi_1^{\mathrm{OL}} + \pi_2^{\mathrm{OL}} + CS^{\mathrm{OL}} \\ &= S - c - \frac{d}{4} + \frac{\delta + 2\rho}{4\alpha\delta(\delta+\rho)^2} - \frac{d[d^2 + 2d\beta(4\eta - 1) + 2\beta^2(1 - 4\eta + 8\eta^2)]}{8\beta^2}. \end{aligned}$$

Q.E.D.

A.2 Proof of Proposition 2

Taking the f rst-order conditions of equation (9) yields

$$I_i(t) = \frac{1}{\alpha}\frac{\partial V_i(X(t))}{\partial X_i(t)}, \quad i = 1, 2, \tag{A13}$$

$$\frac{c + d + dy_i(t) - dy_j(t) - 2p_i(t) + p_j(t) + X_i(t) - X_j(t)}{2d} = 0, \quad i, j = 1, 2, \quad i \neq j, \tag{A14}$$

$$\frac{p_i(t) - c + 4\beta[\eta - y_i(t)]}{2} = 0, \quad i = 1, 2. \tag{A15}$$

From equations (A14) and (A15), we have

$$p_i(t) = c + d + \frac{2\beta}{6\beta - d}\left[X_i^{\mathrm{OL}}(t) - X_j^{\mathrm{OL}}(t)\right], \quad i = 1, 2, \quad i \neq j, \quad \text{and} \tag{A16}$$

$$y_i(t) = \frac{d}{4\beta} + \eta + \frac{1}{2(6\beta - d)}\left[X_i^{\mathrm{OL}}(t) - X_j^{\mathrm{OL}}(t)\right], \quad i = 1, 2, \quad i \neq j. \tag{A17}$$

Substituting equations (A13), (A16), and (A17) into equation (9) leads to

$$\text{(A18)}\quad \begin{aligned}\rho V_i(X(t)) = {} & \frac{(8\beta-d)[d^2-2(3d+X_i(t)-X_j(t))\beta]^2}{16d\beta(d-6\beta)^2}+\frac{1}{2\alpha}\left(\frac{\partial V_i(X(t))}{\partial X_i(t)}\right)^2 \\ & -\delta X_i(t)\frac{\partial V_i(X(t))}{\partial X_i(t)}+\frac{1}{\alpha}\frac{\partial V_i(X(t))}{\partial X_j(t)}\left(\frac{\partial V_j(X(t))}{\partial X_j(t)}-\alpha\delta X_j(t)\right).\end{aligned}$$

It is diff cult to derive the feedback equilibrium without obtaining a closed-form solution for this system. We conjecture a quadratic value function for f rm i, i.e.,

$$\text{(A19)}\quad \begin{aligned}V_i(X(t)) = {} & m_1+\alpha m_2 X_i(t)+m_3X_j(t)+\frac{\alpha m_4}{2}X_i^2(t)+\alpha m_5 X_i(t)X_j(t) \\ & +\frac{m_6}{2}X_j^2(t)\,,\end{aligned}$$

where m_k, $k = 1, ..., 6$, are yet undetermined parameters.

From equation (A13), we obtain

$$\text{(A20)}\quad I_i(t)=m_2+m_4X_i(t)+m_5X_j(t),\quad i=1,2,\quad i\neq j\,.$$

Substituting equations (A19) and (A20) into equation (A18) leads to the following six equation systems in the value function parameters:

$$\text{(A21)}\quad -A-2\alpha[m_4^2+2m_5^2-m_4(2\delta+\rho)]=0\,,$$

$$\text{(A22)}\quad A/2-m_5[m_6-\alpha(2m_4-2\delta-\rho)]=0\,,$$

$$\text{(A23)}\quad 2\alpha m_5^2+A-2m_6(2m_4-2\delta-\rho)=0\,,$$

$$8\beta-d-4m_3m_5(d-6\beta)-4\alpha m_2(d-6\beta)(m_4+m_5-\delta-\rho)=0\,,$$

$$d-8\beta+4m_2(m_6+m_5\alpha)(d+6\beta)-4m_3(d-6\beta)(m_4-\delta-\rho)=0\,,\quad \text{and}$$

$$-8d-8m_2(2m_3+m_2\alpha)+d^2/\beta+16\rho m_1=0\,.$$

Solving equations (A21)–(A23) leads to the parameter values of m_4, m_5, m_6. In order to ensure that the differential system

$$\text{(A24)}\quad \begin{cases}X_1'(t)=I_1(t)-\delta X_1(t)=m_2+(m_4-\delta)X_1(t)+m_5X_2(t),\\ X_2'(t)=I_2(t)-\delta X_2(t)=m_2+(m_4-\delta)X_2(t)+m_5X_1(t)\end{cases}$$

is stable, the Routh–Hurwicz condition must be satisfied This requirement generates the following two inequalities:

$$m_4-\delta<0\quad\text{and}\quad(m_4-\delta)^2-m_5^2>0\,.$$

Therefore we have

$$\text{(A25)}\quad \begin{aligned}& m_2=\frac{3(8\beta-d)}{4\alpha(6\beta-d)(2\delta+\mu^{\mathrm{FB}}+3\rho)},\quad m_4=\frac{\delta-\mu^{\mathrm{FB}}}{3},\quad\text{and}\\ & m_5=-\frac{\delta-\mu^{\mathrm{FB}}}{3}\,.\end{aligned}$$

The steady-state levels of investment and quality can be derived from equations (A24) and (A25):

$$I^{\mathrm{FB}}=\frac{3(8\beta-d)}{4\alpha(6\beta-d)\delta(2\delta+\mu^{\mathrm{FB}}+3\rho)}\quad\text{and}\quad X^{\mathrm{FB}}=\frac{I^{\mathrm{FB}}}{\delta}\,.$$

Equation (A24) and the initial conditions generate the induced quality path

$$\begin{bmatrix} X_1^{\text{FB}}(t) \\ X_2^{\text{FB}}(t) \end{bmatrix} = \mathrm{T} \cdot e^{-t\cdot\delta} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \Psi \cdot e^{-\frac{t}{3}(\delta+2\mu^{\text{FB}})} \begin{bmatrix} -1 \\ 1 \end{bmatrix} \begin{bmatrix} X^{\text{FB}} \\ X^{\text{FB}} \end{bmatrix},$$

where $\mathrm{T} = (X_{20} + X_{10} - 2X^{\text{FB}})/2$. Equations (A13), (A16), (A17), and (A25) lead to the pair of strategies $\{I_i^{\text{FB}}(t), y_i^{\text{FB}}(t), p_i^{\text{FB}}(t)\}$ where

$$\text{(A26)} \quad I_i^{\text{FB}}(t) = I^{\text{FB}} + \frac{\delta - \mu^{\text{FB}}}{3} X_i^{\text{FB}}(t) - \frac{\delta - \mu^{\text{FB}}}{3} X_j^{\text{FB}}(t), \quad i, j = 1, 2, \quad i \neq j,$$

$$\text{(A27)} \quad y_i^{\text{FB}}(t) = \frac{d}{4\beta} + \eta + \frac{1}{2(6\beta - d)} \left[X_i^{\text{FB}}(t) - X_j^{\text{FB}}(t)\right], \quad i = 1, 2, \quad i \neq j,$$

$$\text{(A28)} \quad p_i^{\text{FB}}(t) = c + d + \frac{2\beta}{6\beta - d} \left[X_i^{\text{FB}}(t) - X_j^{\text{FB}}(t)\right], \quad i, j = 1, 2, \quad i \neq j.$$

Combining equations (3), (A26)–(A28) yields the path of f rm 1's market share:

$$D_1^{\text{FB}}(t) = \frac{1}{2} - \frac{2\beta\Psi}{d(6\beta - d)} e^{-\frac{t}{3}(\delta+2\mu^{\text{FB}})}.$$

The steady-state equilibrium values can be summarized as

$$D_1^{\text{FB}} = \frac{1}{2}, \quad p_1^{\text{FB}} = p_2^{\text{FB}} = c + d, \quad I_1^{\text{FB}} = I_2^{\text{FB}} = I^{\text{FB}} = \frac{3(8\beta - d)}{4\alpha(6\beta - d)(2\delta + \mu^{\text{FB}} + 3\rho)},$$

and

$$X_1^{\text{FB}} = X_2^{\text{FB}} = X^{\text{FB}} = \frac{I^{\text{FB}}}{\delta}, \quad y_1^{\text{FB}} = y_2^{\text{FB}} = y^{\text{FB}} = \frac{d}{4\beta} + \eta.$$

The steady-state equilibrium prof ts, consumer surplus, and social welfare are, respectively,

$$\pi_1^{\text{FB}} = \pi_2^{\text{FB}} = \frac{d}{2} - \frac{9\Theta^2}{32\alpha(2\delta + \mu^{\text{FB}} + 3\rho)^2},$$

$$CS^{\text{FB}} = S - c - \frac{3\Theta}{4\alpha\delta(2\delta + \mu^{\text{FB}} + 3\rho)} - \frac{d[d^2 + 2d\beta(4\eta - 1) + 2\beta^2(5 - 4\eta + 8\eta^2)]}{8\beta^2},$$

$$W^{\text{FB}} = S - c + \frac{3\Theta[(8 - 3\Theta)\delta + 4(\mu^{\text{FB}} + 3\rho)]}{16\alpha\delta(2\delta + \mu^{\text{FB}} + 3\rho)^2} - \frac{d[d^2 + 2d\beta(4\eta - 1) + 2\beta^2(1 - 4\eta + 8\eta^2)]}{8\beta^2},$$

where

$$\Theta = \frac{8\beta - d}{6\beta - d} \in \left(\frac{5}{3} - \frac{\delta}{3(\delta + \rho)}, \frac{4}{3}\right).$$

Q.E.D.

A.3 Proof of Proposition 3

Based on the results in Propositions 1 and 2, we obtain the following results:

$$I^{\text{OL}} - I^{\text{FB}} = \frac{\mathrm{E}}{4\alpha(\delta + \rho)(2\delta + \mu^{\text{FB}} + 3\rho)} > 0,$$

where $\mathrm{E} = \delta(4 - 3\Theta) + 2\mu^{\mathrm{FB}} + 3\rho(2 - \Theta) > 2\mu^{\mathrm{FB}} + \rho > 0$;

$$X^{\mathrm{OL}} - X^{\mathrm{FB}} = \frac{I^{\mathrm{OL}} - I^{\mathrm{FB}}}{\delta} > 0\,,$$

$$\pi_i^{\mathrm{FB}} - \pi_i^{\mathrm{OL}} = \frac{\mathrm{E} \cdot [(4 + 3\Theta)\delta + 2\mu^{\mathrm{FB}} + 3\rho(2 + \Theta)]}{32\alpha(\delta + \rho)^2(2\delta + \mu^{\mathrm{FB}} + 3\rho)^2} > 0\,,$$

$$CS^{\mathrm{FB}} - CS^{\mathrm{OL}} = -\frac{\mathrm{E}}{4\alpha\delta(\delta + \rho)(2\delta + \mu^{\mathrm{FB}} + 3\rho)} < 0\,,$$

$$\begin{aligned} W^{\mathrm{FB}} - W^{\mathrm{OL}} &= \frac{\mathrm{E} \cdot [(3\Theta - 4)\delta^2 - 4\rho(\mu^{\mathrm{FB}} + 3\rho) + \delta\rho(3\Theta - 14) - 2\delta\mu^{\mathrm{FB}}]}{16\alpha\delta(\delta + \rho)^2(2\delta + \mu^{\mathrm{FB}} + 3\rho)^2} \\ &< -\frac{\mathrm{E} \cdot [4\rho(\mu^{\mathrm{FB}} + 3\rho) + \delta(2\mu^{\mathrm{FB}} + 9\rho)]}{16\alpha\delta(\delta + \rho)^2(2\delta + \mu^{\mathrm{FB}} + 3\rho)^2} < 0\,, \end{aligned}$$

(A29) $$p_i^{\mathrm{FB}}(t) = c + d + (-1)^i \frac{4\beta\Psi}{6\beta - d} e^{-t\cdot(\delta + 2\mu^{\mathrm{FB}})/3}, \quad i = 1, 2\,,$$

(A30) $$y_i^{\mathrm{FB}}(t) = \frac{d}{4\beta} + \eta + (-1)^i \frac{\Psi}{6\beta - d} e^{-t\cdot(\delta + 2\mu^{\mathrm{FB}})/3}, \quad i = 1, 2\,,$$

$$p_1^{\mathrm{OL}}(t) - p_2^{\mathrm{OL}}(t) = (X_{10} - X_{20}) \frac{4\beta}{6\beta - d} e^{-t\cdot\mu^{\mathrm{OL}}}, \quad \text{and}$$

$$p_1^{\mathrm{FB}}(t) - p_2^{\mathrm{FB}}(t) = (X_{10} - X_{20}) \frac{4\beta}{6\beta - d} e^{-t\cdot(\delta + 2\mu^{\mathrm{FB}})/3}\,;$$

$$\mathrm{sign}\{p_1^{\mathrm{OL}}(t) - p_2^{\mathrm{OL}}(t)\} = \mathrm{sign}\{p_1^{\mathrm{FB}}(t) - p_2^{\mathrm{FB}}(t)\} = \mathrm{sign}\{X_{10} - X_{20}\}\,.$$

Taking the limits of equations (6), (7), (A29), and (A30) leads to

$$\lim_{t\to+\infty} p_i^{\mathrm{OL}}(t) = \lim_{t\to+\infty} p_i^{\mathrm{FB}}(t) = c + d, \quad \lim_{t\to+\infty} y_i^{\mathrm{OL}}(t) = \lim_{t\to+\infty} y_i^{\mathrm{FB}}(t) = \frac{d}{4\beta} + \eta\,.$$

Q.E.D.

A.4 Proof of Proposition 4

The Hamiltonian for the social planner has the form

$$H(t) = w(t) + \lambda_{1s}(t)[I_1(t) - \delta X_1(t)] + \lambda_{2s}(t)[I_2(t) - \delta X_2(t)]\,,$$

where $\lambda_{is}(t)$ are the costate variables associated with $X_i(t), i = 1, 2$. The corresponding f rst-order conditions are, respectively,

(A31) $$\frac{\partial H(t)}{\partial D_1(t)} = d[1 - 2D_1(t)] + X_1(t) - X_2(t) = 0\,,$$

(A32) $$\frac{\partial H(t)}{\partial y_1(t)} = d[D_1(t) - 2y_1(t)] + 2\beta[\eta - y_1(t)] = 0\,,$$

(A33) $$\frac{\partial H(t)}{\partial y_2(t)} = d[1 - D_1(t) - 2y_2(t)] + 2\beta[\eta - y_2(t)] = 0, \quad \text{and}$$

(A34) $$\frac{\partial H(t)}{\partial I_i(t)} = \lambda_{is}(t) - \alpha I_i(t) = 0, \quad i = 1, 2\,.$$

The adjoint equations for the optimum are

(A35) $$\lambda'_{1s}(t) = \rho\lambda_{1s}(t) - \frac{\partial H(t)}{\partial X_1(t)} = (\rho+\delta)\lambda_{1s}(t) - D_1(t) \quad \text{and}$$

(A36) $$\lambda'_{2s}(t) = \rho\lambda_{2s}(t) - \frac{\partial H(t)}{\partial X_2(t)} = (\rho+\delta)\lambda_{2s}(t) + D_1(t) - 1\,.$$

The transversality conditions are

(A37) $$\lim_{t\to\infty} \lambda_{is}(t)\cdot X_i(t)\cdot e^{-t\cdot\rho} = 0, \quad i = 1,2\,.$$

Equations (A31)–(A34) generate

(A38) $$D_1(t) = \frac{1}{2} + \frac{d+\beta}{d(d+2\beta)}[X_1(t) - X_2(t)]\,,$$

(A39) $$y_i(t) = \frac{d+4\beta\eta}{4(d+\beta)} + \frac{1}{2d+4\beta}[X_i(t) - X_j(t)], \quad i,j = 1,2, \quad i \neq j, \quad \text{and}$$

(A40) $$\lambda_{is}(t) = \alpha I_i(t), \quad i = 1,2\,.$$

Substituting equations (A38), (A40) into equations (A35) and (A36) leads to

(A41)
$$I'_i(t) = -\frac{1}{2\alpha} + (\rho+\delta)I_i(t) - \frac{d+\beta}{d\alpha(d+2\beta)}[X_i(t) - X_j(t)], \quad i,j = 1,2, \quad i \neq j\,.$$

Combining equations (1) and (A41) yields the following differential system:

(A42)
$$\begin{cases} I'_i(t) = -\dfrac{1}{2\alpha} + (\rho+\delta)I_i(t) - \dfrac{d+\beta}{d\alpha(d+2\beta)}[X_i(t) - X_j(t)] \\ X'_i(t) = I_i(t) - \delta X_i(t) \end{cases}, \quad i,j = 1,2, \quad i \neq j\,.$$

The steady-state equilibrium values can be derived from $I'_i(t) = X'_i(t) = 0$:

$$I^S = \frac{1}{2\alpha(\delta+\rho)} \quad \text{and} \quad X^S = \frac{I^S}{\delta}\,.$$

The initial conditions and equation (A37) generate the solutions to equation (A42):

(A43) $$\begin{bmatrix} I_1^S(t) \\ I_2^S(t) \\ X_1^S(t) \\ X_2^S(t) \end{bmatrix} = \Phi e^{-t\cdot\delta}\begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} + \Psi e^{-t\cdot\mu^S}\begin{bmatrix} -(\delta-\mu^S) \\ \delta-\mu^S \\ -1 \\ 1 \end{bmatrix} + \begin{bmatrix} I^S \\ I^S \\ X^S \\ X^S \end{bmatrix},$$

where

$$\delta > \mu^S = \frac{1}{2}\sqrt{(2\delta+\rho)^2 - \frac{8(d+\beta)}{d^2\alpha + 2d\alpha\beta}} - \frac{\rho}{2} > 0\,.$$

Combining equations (A38), (A39), and (A43) yields

$$D_1^S(t) = \frac{1}{2} - \frac{2(d+\beta)\Psi}{d(d+2\beta)}e^{-t\cdot\mu^S}, \quad y_i^S(t) = \frac{d+4\beta\eta}{4(d+\beta)} + (-1)^i\frac{\Psi}{2\beta+d}e^{-t\cdot\mu^S}\,.$$

Q.E.D.

References

Auray, S., T. Mariotti, and F. Moizeau (2011), "Dynamic Regulation of Quality," *The RAND Journal of Economics*, 42(2), 246–265.

Bass, F. M., A. Krishnamoorthy, A. Prasad, and S. P. Sethi (2005), "Generic and Brand Advertising Strategies in a Dynamic Duopoly," *Marketing Science*, 24(4), 556–568.

Bertuzzi, G., and L. Lambertini (2010), "Existence of Equilibrium in a Differential Game of Spatial Competition with Advertising," *Regional Science & Urban Economics*, 40(2–3), 155–160.

Brander, J. A., and B. J. Spencer (1983), "Strategic Commitment with R&D: The Symmetric Case," *The Bell Journal of Economics*, 14(1), 225–235.

Brekke, K. R., R. Cellini, L. Siciliani, and O. R. Straume (2010), "Competition and Quality in Health Care Markets: A Differential-Game Approach," *Journal of Health Economics*, 29(4), 508–523.

Bulow, J. (1986), "An Economic Theory of Planned Obsolescence," *The Quarterly Journal of Economics*, 101(4), 729–750.

Bulow, J. I. (1982), "Durable-Goods Monopolists," *Journal of Political Economy*, 90(2), 314–332.

D'Aspremont, C., and A. Jacquemin (1988), "Cooperative and Noncooperative R&D in Duopoly with Spillovers," *The American Economic Review*, 78(5), 1133–1137.

Dasgupta, P., and J. Stiglitz (1980), "Industrial Structure and the Nature of Innovative Activity," *The Economic Journal*, 90(358), 266–293.

Dixit, A. (1988), "A General Model of R&D Competition and Policy," *The RAND Journal of Economics*, 19(3), 317–326.

Erickson, G. M. (1995), "Differential Game Models of Advertising Competition," *European Journal of Operational Research*, 83(3), 431–438.

Fershtman, C. (1984), "Goodwill and Market Shares in Oligopoly," *Economica*, 51(203), 271–281.

— and E. Muller (1984), "Capital Accumulation Games of Infinit Duration," *Journal of Economic Theory*, 33(2), 322–339.

Figuieres, C. (2002), "Complementarity, Substitutability and Strategic Accumulation of Capital," *International Game Theory Review*, 4(4), 371–390.

Friedman, J. W. (1983), "Advertising and Oligopolistic Equilibrium," *The Bell Journal of Economics*, 14(2), 464–473.

Hanig, M. (1987), "Differential Gaming Models of Oligopoly," Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge (MA).

Harter, J. F. R. (1993), "Differentiated Products with R&D," *The Journal of Industrial Economics*, 41(1), 19–28.

Hotelling, H. (1929), "Stability in Competition," *The Economic Journal*, 39(153), 41–57.

Jun, B., and X. Vives (2004), "Strategic Incentives in Dynamic Duopoly," *Journal of Economic Theory*, 116(2), 249–281.

Kamien, M. I., E. Muller, and I. Zang (1992), "Research Joint Ventures and R&D Cartels," *The American Economic Review*, 82(5), 1293–1306.

Katz, M. L. (1986), "An Analysis of Cooperative Research and Development," *The RAND Journal of Economics*, 17(4), 527–543.

Lambertini, L. (2002), "Equilibrium Locations in a Spatial Model with Sequential Entry in Real Time," *Regional Science & Urban Economics*, 32(1), 47–58.

— (2005), "Advertising in a Dynamic Spatial Monopoly," *European Journal of Operational Research*, 166(2), 547–556.

Lee, T., and L. L. Wilde (1980), "Market Structure and Innovation: A Reformulation," *The Quarterly Journal of Economics*, 94(2), 429–436.

Lewis, T. R., and D. E. M. Sappington (1991), "Incentives for Monitoring Quality," *The RAND Journal of Economics*, 22(3), 370–384.

— and — (1992), "Incentives for Conservation and Quality-Improvement by Public Utilities," *The American Economic Review*, 82(5), 1321–1340.

Li, C., and J. Zhang (2012), "R&D Competition in a Spatial Model with Technical Risk," *Papers in Regional Science*, DOI: 10.1111/j.1435-5957.2012.00433.x.

Mukhopadhyay, S. K., and P. Kouvelis (1997), "A Differential Game Theoretic Model for Duopolistic Competition on Design Quality," *Operations Research*, 45(6), 886–893.

Nair, A., and R. Narasimhan (2006), "Dynamics of Competing with Quality- and Advertising-Based Goodwill," *European Journal of Operational Research*, 175(1), 462–474.

Narajabad, B., and R. Watson (2008), "Innovation and Horizontal Differentiation in a Continuous-Time Hotelling Model," Mimeo, Rice University Department of Economics, Houston (TX).

— and — (2011), "The Dynamics of Innovation and Horizontal Differentiation," *Journal of Economic Dynamics & Control*, 35(6), 825–842.

Piga, C. A. G. (1998), "A Dynamic Model of Advertising and Product Differentiation," *Review of Industrial Organization*, 13(5), 509–522.

— (2000), "Competition in a Duopoly with Sticky Price and Advertising," *International Journal of Industrial Organization*, 18(4), 595–614.

Reinganum, J. F. (1985), "Innovation and Industry Evolution," *The Quarterly Journal of Economics*, 100(1), 81–99.

Reynolds, S. S. (1987), "Capacity Investment, Preemption and Commitment in an Infinit Horizon Model," *International Economic Review*, 28(1), 69–88.

Waldman, M. (1993), "A New Perspective on Planned Obsolescence," *The Quarterly Journal of Economics*, 108(1), 273–283.

— (1996), "Planned Obsolescence and the R&D Decision," *The RAND Journal of Economics*, 27(3), 583–595.

Wirl, F. (2010), "Dynamic Demand and Noncompetitive Intertemporal Output Adjustments," *International Journal of Industrial Organization*, 28(3), 220–229.

Changying Li
School of Economics
Shandong University
27 Shanda South Road, Jinan
250100 Shandong
P. R. China
changyingli@sdu.edu.cn

Jianhu Zhang
School of Economics
Shandong University
27 Shanda South Road, Jinan
250100 Shandong
P. R. China
jianhu2528785@163.com

Personal Bankruptcy Law, Fresh Starts, and Judicial Practice

by

Régis Blazy, Bertrand Chopard, Eric Langlais, and Ydriss Ziane*

Received February 21, 2011; in revised form September 6, 2012;
accepted June 5, 2013

We explore the rationale behind and the mechanisms employed by French judges while discharging personal debts in exchange for liquidation of debtors' assets. Our empirical results highlight the determinants of judicial selection between debtors whose debts are wiped out and those who have to reimburse them. We fin that French judges tend to disqualify debtors with multiple creditors from debt discharge, and are sensitive to variables representative of economic activity in the courts' locality. These empirical results help us to understand better how much French personal bankruptcy law is rather pro-creditor than pro-debtor. (JEL: G33, K29)

1 Introduction

It is observed that, worldwide, personal bankruptcy laws consist of at least two formal procedures for resolving fi ancial distress: *debt restructuring* via a reimbursement plan, and *liquidation* with some debt discharge. The f rst procedure, debt restructuring, begins by starting a renegotiation process between the debtors and the main creditors. This process is supervised by a judge or an administrative authority. During this procedure, an automatic stay order comes into effect to prevent the

* Régis Blazy: EM Strasbourg Business School and IEP Strasbourg; Bertrand Chopard (corresponding author): Université Paris Ouest Nanterre la Défense and EconomiX, Nanterre; Eric Langlais: Université Paris Ouest Nanterre la Défense and EconomiX, Nanterre; and Ydriss Ziane: Sorbonne Graduate Business School, Paris. We would like to thank the anonymous referees and Yvan Lengwiler for their helpful comments. We equally benefite from the participants of the 2011 European Association of Law and Economics conference in Hamburg, the 2011 Infinit conference in Dublin, the 2010 French Association of Economics conference in Paris, the 2011 workshop in Law and Economics in Besançon, and the 2011 LIEN seminar at Université Paris Ouest Nanterre la Défense. Finally, we express our gratitude to the French National Research Agency for its financ al support (under the ANR research program DETTES).

Journal of Institutional and Theoretical Economics 169, 680–702 – ISSN 0932-4569
DOI: 10.1628/093245613X13806312325814 –

creditors from pursuing the debtors. If the renegotiation succeeds, debtors usually have to reimburse their debt (partially or wholly) out of their future flow of incomes within a predetermined period. Generally, the law decrees that part of future earnings is exempt from debt reimbursement, depending on family size, location, etc. The second bankruptcy procedure, liquidation, involves the discharge of personal debts carried out under the supervision of a judge. The main objective of this procedure is the cancellation of debts through repayment from the proceeds of nonexempt assets' liquidation. As many bankrupts do not own many assets, such liquidation can be viewed as a *fresh start*. However, for bankrupts to benefit from this fresh start, a judge has to first gauge whether or not they have a chance of repaying their debts in the future. If they do, the judge may order a new schedule of repayment of creditors, or ask the debtors and creditors to renegotiate a debt reschedule plan.

Facing a substantial and steady rise of these personal bankruptcy filings, some authors have then questioned the reasons for it, mainly in the U.S. (see appendix A.1 for the evolution of the annual number of filings recorded in France for the period 1993–2011). Several factors have been cited for the steady rise in bankruptcy filings: the development of credit debt and revolving debt with high interest (Dick and Lehnert, 2010); the misuse of credit or the debtors' inability to manage their budget (Lusardi and Tufano, 2008); the increase in major expenses, such as domestic rental and personal costs (especially following divorce, separation, or loss of job – see Gan and Sabarwal, 2005); and also the design of personal bankruptcy law (Fay, Hurst, and White, 2002). A second strand of empirical research seeks to explain whether U.S. personal bankruptcy law may affect the supply of credit. For instance, Gropp, Scholz, and White (1997) demonstrate that American debtors who live in those states that have higher exemption levels on assets and future incomes (i.e., the most pro-debtor states) necessarily have more difficulty accessing credit than others and pay a higher price for it.[1]

In comparison, there are not many country-specific empirical studies dealing with the personal bankruptcy rules applying in Europe, especially with the factors that determine the decision-making process of a judge discharging personal debts. Only White (2007) has initiated a cross-country comparison of the bankruptcy rules prevailing in the U.S., Canada, England, France, and Germany.[2] White (2007) concludes that France is clearly pro-creditor, in the sense that bankruptcy laws tend to protect the creditors' interests rather than the debtor' ones, for three reasons. First, the average duration of repayment plans is quite long, approximately eight years. Second, the levels of exemption on assets/future incomes are very low, meaning that bankrupts will have to adjust their net income to a minimal poverty-level standard of

[1] See Filer II and Fisher (2005), Filer and Fisher (2007), and Musto (2004) for a more comprehensive study of credit rationing for debtors who recently benefited from a debt reschedule payment plan.

[2] Her work relies on seven variables: the amount of debt discharged, the level of asset exemption, the level of income exemption, the fraction of income above the exemption that debtors must use to repay their debts, the length of the repayment obligation, the bankruptcy costs, and the bankruptcy punishment.

living during the debt reschedule plan. Third, individuals generally have to sell their assets (if they have any) before f ling for bankruptcy if they want to benef t from the procedure. Besides the lack of empirical studies in Europe, a major limitation of previous empirical works stems from the risk of misrepresenting bankruptcy laws, in particular the gap between the rules and the practices of bankruptcy courts. Indeed, some scholars have recognized for a long time the importance of "local legal culture" in the implementation of the personal bankruptcy rules (Braucher, 1993; Sullivan, Warren, and Westbrook, 1994). In the U.S. case, they have shown that the same bankruptcy laws are administered very differently within a single state. To sum up, they give strong evidence of district-level variations in (*i*) the bankruptcy f ling rates, (*ii*) the debtors' choice between Chapter 7 (the fresh start) and Chapter 13 (the reorganization process), (*iii*) the propensity to reaffir debts via Chapter 7, and (*iv*) the rate of successful completion of reimbursement plans. The same would probably be true on the cross-country level. So, even if one were to know the laws applying in each European country, that knowledge would remain incomplete unless one also knew how the law really worked in the courts.

Based on an empirical study of 1084 French bankruptcy f lings in 2004–2005, a f rst attempt is here made to take into account the working mechanisms of courts under the French legal system, especially how judges discharge personal debts in exchange for liquidation of debtors' nonexempt assets. Intuitively, for the judge, the choice to discharge the personal debts will be a function of the debtor's capacity to reimburse the debts. Controlling for this effect, we use regression analysis to test the hypothesis that judges tend also to disqualify debtors with multiple creditors from debt discharge. In other words, we test for the f rst time whether judges may also be inf uenced by their sense that debtors are responsible for overindebtedness. Our focus is also on the behavioral dynamics of judges with respect to the implementation of the laws, so that one can appreciate the relation between our paper and the tenets of the behavioral law-and-economics literature. In that perspective, Rachlinski, Guthrie, and Wistrich (2007) have explored the possibility of emotional inf uence in judges' decision-making with respect to U.S. personal bankruptcy law.[3] Here, we would like to test a second question: is the judges' choice whether or not to discharge debts sensitive to the macroeconomic context in the court's locality? In particular, our intuition is that judges will be more likely to enforce debt discharge in regions where there is a clear shortage of employment, or where debtors face imminent increased risk of an adverse event (job loss) in the future.

The rest of this paper is organized as follows. In section 2 we report the basic features of French bankruptcy law; section 3 presents our data and regression results; section 4 concludes.

[3] Rachlinski, Guthrie, and Wistrich (2007) show that debtors' apologies (for excessive accumulation of debts) have little effect on judges' choices of whether or not to discharge debts, even after controlling for judges' characteristics (gender, judicial experience, and political affili tion). Instead, they give some evidence that Republican judges are more likely than Democratic judges to make decisions in favor of creditors.

2 French Personal Bankruptcy Law: An Overview

Before benefiting from a debt discharge (or other remedy), financially distressed French debtors have first to file for a bankruptcy procedure under the supervision of an administrative authority, the *commission de surendettement* (CSUR). This initial bankruptcy filing is automatically associated with the debtor's record in a national file of bankrupt debtors. In addition, all creditors' pursuits (for example, asset seizure) are suspended during this bargaining process. But the CSUR may either accept or reject the debtor's bankruptcy filing. The CSUR authorizes debtors to continue the bankruptcy process only if they (1) appear to have significant difficulties in repaying their debts from their current incomes and assets and (2) are bona fide. More precisely, after having examined the debtor's situation, the CSUR (the judge does not intervene at this time) has the right to choose between two different bankruptcy procedures: the *plan de redressement* and the *procédure de rétablissement personnel* (PRP hereafter). In order to choose between these two outcomes, the CSUR calculates a standard level of charges for each debtor, based on a common scale for all CSURs, that takes into account family size, living expenses, and medical and school bills. When the difference between the monthly resources and the sum of these charges is positive, the CSURs attempt to favor a debt renegotiation between the filer and his or her creditors (i.e., a *plan de redressement*; see description below). In contrast, when the reimbursement capacity is negative, the CSUR will generally propose that the debtor benefit from full debt discharge after liquidation of his or her assets (i.e., a PRP; see description below), but the final outcome (a fresh start or not) has to be approved by a judge.

By using the annual univariate statistics computed by the Banque de France on the work of CSURs, we report that in the year 2004, (*i*) 73.6% of bankruptcy filers had a reimbursement capacity between 0 and €450 per month, and (*ii*) 31.8% of filers had a strictly negative reimbursement capacity. We also notice that in the year 2004, in a set of 188,176 bankruptcy filings, 153,185 debtors (or 83.1%) were allowed to file for a bankruptcy procedure (either *plan de redressement* or PRP). The rest (34,991 debtors) did not benefit from any delays in reimbursing their debt, reduction of interest, or debt release. Further, 131,151 debtors (or 85.6% of the debtors in the bankruptcy process for 2004) attempted to renegotiate a debt reschedule plan with their creditors (via the *plan de redressement*). The others (14.6%) filed directly for the PRP in order to benefit (or not) from a full debt discharge in exchange for liquidation of their nonexempt assets. So, the great majority of bankruptcy filers have to renegotiate their debts, and full debt discharge with asset liquidation is reserved for the most financially distressed debtors (because the CSUR automatically proposes the PRP when the debtor's reimbursement capacity is negative). But we cannot investigate more precisely which factors determined whether the CSUR orders a *plan de redressement* or a PRP, because we do not have any access to the (confidential) information gathered by the Banque de France.

Now, we illustrate the two bankruptcy procedures in more detail.

2.1 The plan de redressement

This procedure aims at elaborating a restructuring debt schedule through debt renegotiation with the debtor and his or her main creditors. Because, as we have seen, the CSUR initially calculates the debtor's reimbursement capacity by using a fi ed formula (the difference between the debtor's income and a standard level of charges), debts will generally be reimbursed according to the debtor's reimbursement capacity over a fi ed period (less than ten years[4]) and personal savings. Notice that the CSUR can even order the sale of the debtor's real estate assets in order to reimburse the debt. But such a measure can only be regarded as a last resort, and is rarely used. During the renegotiation process, the creditors may agree to accept some debt reductions or delays in repayment. For instance, for the year 2004, about half the fi ers obtained a full clearing of interest charges, the rest obtaining smaller interest reductions (source: Banque de France, 2008). Now, if renegotiation fails (as it does for about 1 in 3 f lers), there are three possible outcomes. First, the CSUR may ask a judge to enforce a schedule of repayment (on the strict condition that debts will be totally reimbursed during a period of ten years at most). Second, the CSUR may rule that the debtors do not have to reimburse their debts for a period of at most two years. Third, the CSUR may propose that the debtor f le for a PRP.

2.2 The procédure de rétablissement personnel (PRP)

Upon commencement of the fir t step, the CSUR estimates that there is no (or very little) chance that debts can be reimbursed from future income and assets, or that the debt renegotiation has failed. As a next step, a judge (with the debtor's authorization) has to decide whether debts (except for specif c debts such as secured loans, f nes, or child-support fees) will be discharged or not. In exchange, all the debtor's nonexempt assets are liquidated (the exempt assets being mainly the debtor's vehicle and other goods essential to life). Then, the liquidation proceeds are divided among creditors according to a strict priority rule. But prior to ordering a debt discharge (the main topic of this paper), judges have to verify that debtors are unable to repay their debts with their future income and current assets, and are bona f de. The f rst of these criteria means that the debtors with a higher capacity of repayment should be less likely to obtain a debt discharge through a PRP. More precisely, the judge has to compare, for each debtor, the difference between the debtor's current income and expenses (a proxy for the debtor's reimbursement capacity) with his or her total amount of debts. The second criterion aims at excluding individuals who strategically use the procedure in order to benef t from full debt discharge when they have accumulated too many debts in the past. In our database of bankruptcy debtors, we have found that only 4.2% of f lers were identif ed as mala f de. What explains that low percentage? First of all, debtors are presumed bona f de until the creditors establish the reverse. In other words, the judge does not run any investigations in order to establish whether the debtor is bona fid or not. That means also that one or

[4] This maximum was recently changed to eight years.

more creditors have to prove to the judge that the debtor is a mala f de one, which can be fairly diff cult to establish, as there does not exist an off cial and agreed-upon def nition of a mala f de borrower. Second, it is also plausible that creditors may have not incentives to monitor or control for debtors' bona f des, especially for lower values of debts or expected recoveries. After exploring our sample of mala fi e cases, we know only that these debtors were identif ed as mala f de ones because they had falsif ed private information in order to become voluntarily overindebted, or wrote checks without funds.

3 Empirical Findings

In this section, we test how judges handle the debt discharge in exchange for the liquidation of nonexempt assets within the PRP. More precisely, we compare bona f de debtors who obtain a full debt discharge in exchange for their nonexempt assets liquidation, and bona f de debtors who have to reimburse their debts. We do not consider in the rest of this paper the male f de debtors (who do not benef t from any debt discharge), for two main reasons. As noted above, there are very few debtors identif ed as mala f de ones. Further, and more crucially, it is quite diff cult to understand how our explanatory variables might help to explain statistically what is bona or mala f de. However, we propose in appendix A.2 a robustness analysis including the mala f de debtors, such that the judge's decision is now among three outcomes: debt discharge (for bona f de debtors), no debt discharge (for bona f de debtors), and no debt discharge due to debtor's mala f des. Now, we f rst describe our data on PRPs.

3.1 Data Description

Even though this subject has gained worldwide attention due to the steady rise of overindebtedness in European countries, only limited information is available on the legal treatment of f nancial distress. In order to improve the quality of the work done by the courts, or at least gain a better understanding of the judicial practices, the French Ministry of Justice ordered a large data collection on PRPs for the period 2004–2005. Information was gathered manually from documents in 158 French courts, including bankruptcy declarations, court decisions and motivations, lists of claims, and characteristics of bankrupts. Using this information, we assembled an original database of 4098 judgments of PRPs delivered in the period 2004–2005 in 20 regions (out of a total of 22 regions), representing nearly 11% of the entire population of debtors filin for a PRP during this period. For each bankruptcy case, we gathered data in relation to the debtor's f nancial situation at the triggering time: total amount of claims due; total income (including wages, unemployment benef ts, family income support, housing benef ts, rent allowances, sickness benef ts, and old-age pensions); total amount of expenses (debt service, dependents, taxes, rent, additional expenses calculated by the judge on the basis of family size); asset list (exempt or not); and the list of claims. Finally, we reported the duration of the judicial proceedings (i.e., the time between the date of PRP filing and the date of judgment).

As usual, there exists some prescreening (and also preselection here) of trial participants in our data set. More precisely, we identify two potential sources of selection bias (that we cannot control in our regression analysis due to the lack of data). First, we have to keep in mind in our empirical estimations that our sample is composed of debtors who have either failed to renegotiate their debts with their creditors under the supervision of a CSUR, or debtors who have been directly assigned to the judge because a CSUR has previously decided that there was no chance to successfully renegotiate a debt reschedule plan.[5] Second, it is also plausible that some debtors had recently filed for a bankruptcy procedure in the past, especially the *plan de redressement*. As a consequence, in our sample, some bankrupts may have filed for bankruptcy more than once, because either they failed to reimburse their debts according to their repayment plan, or a CSUR has rejected their bankruptcy filing in the first step of the bankruptcy process. To sum up, our sample thus represents the French debtors with the least ability to repay their debts. However, we are sure that these debtors did not benefit in the past from a full discharge of their personal debts in exchange for liquidation of their assets.

After controlling for missing data on some of the variables described here, the sample size falls into 1120 observations on a set of 20 French regions. More precisely, we have eliminated the observations that have missing values concerning debt structure, mainly the number of creditors (with the distinction between financial creditors and ordinary creditors). To evaluate this selection bias, we verified that the characteristics of our sample do not significantly differ from those of the initial database of 4098 judgments, in several ways. First, the amounts of resources and current expenses do not greatly differ: the mean amounts of resources and expenses are respectively €904 and €1004 in our final sample, and respectively €954 and €1089 in the initial data set of judgments (see Table 1 below). Second, the amounts of debts are quite similar in the two samples (€20241 in the initial database, and €21095 in the sample studied in this paper). Third, the sole difference is in the likelihood of debt discharge: Courts' decisions studied in the paper show a slightly lower frequency of debt discharge (25.73%) than observed in the initial database (34.3%), which is very close to the ratio for the entire population of filers (33.2%; source: Banque de France, 2008).

The next step in constructing the sample was to exclude 36 debtors who were identified as mala fide ones.[6] As a consequence, we analyzed data only for bona fide filers (or, at least, those who had not been identified by the judge). Further, we regrouped creditors into two sets: financial claimants (banks and firms specialized

[5] In other words, our analysis ignores, due to the lack of data on CSURs, the debtors who were steered to repayment plans (instead of PRPs) by a CSUR.

[6] Notice that, as indicated before, in the initial database of 4098 filers, only 4.2% of filers were identified as mala fide by judges. The corresponding figure is 3.2% in the sample studied in this paper.

Table 1a
Summary Statistics on the Sample ($N = 1084$)

Variables	Mean	Standard deviation
Sum of debts (euros)	21095.03	37452.3
Resources (euros)	904.07	361.15
Current expenses (euros)	1004.39	358.43
Duration of judicial proceedings (days)	34.61	16.98
Number of creditors	7.62	4.78
Number of financia creditors	2.70	2.16
Number of ordinary creditors	4.92	4.43

in consumer credit) and other claimants (for rent, taxes, energy or communication bills, private debts, commercial debts, unpaid alimony, tuition fees, f nes). Then, we introduced a dummy variable in order to distinguish between debtors who owned real estate at the triggering time and those who did not. Within our sample of 1084 debtors, 15 were owners (occupier or not). So, a very large majority of bankrupts had no assets to liquidate in order to benef t from debt release in exchange. Overall, this f gure may indicate that debtors with some real estate assets refuse to f le for this bankruptcy procedure. It is also plausible that such debtors have restricted access to this procedure, meaning that some CSURs impose a debt reschedule plan on these debtors, or in extreme cases prevent them from renegotiating their debts. However, this does not mean that the individuals had no assets at all. Of the 1069 debtors without real estate, 917 clearly had no assets, but 152 owned a car and/or furniture. More precisely, 106 bankrupts owned a car and no furniture, 11 owned both a car and furniture, and 35 owned furniture but no car. But these assets (cars and furniture) are generally exempt from liquidation because people need cars to get to work, and furniture has a very low liquidation value.

When analyzing our fi al sample (see Table 1a), we f rst notice that the mean amount of debtors' monthly resources is €904.07, the mean amount of debts is €21095.03, and the mean amount of monthly expenses is €1004.39. As a result, debtors have a negative mean capacity to reimburse their debts (as given by the difference between resources and expenses). More precisely, 771 debtors have a strictly negative capacity to reimburse, and 313 debtors have a positive one. Further, debtors are indebted to 7.62 creditors on average, with a minimum of one creditor and a maximum of 35 (median value equals 7). The mean number of fi ancial (other) creditors is 2.70 (4.92). Finally, the mean duration of the judicial proceedings is 34.61 days, with a maximum duration of 147 days. Now, if we compare debtors whose debts are discharged by the judge with other bankrupts, we f nd that 25.73% of debtors in our sample who f led for PRPs did not benef t from a debt discharge.

In Table 1b, we make a distinction between the group of debtors that benef ts from a debt discharge and the group that does not. For each group, we compute both the mean and the standard deviation for the following set of variables: size

Table 1b
Variable Means and Standard Deviations for the Bankruptcy Sample according to the Judgment: Debt Discharge ($N = 805$) versus No Debt Discharge ($N = 279$)

	No debt discharge	Debt discharge		No debt discharge	Debt discharge
Resources (euros)			*Current Expenses (euros)*		
Mean	1029.47	860.61	Mean	995.27	1007.57
t	6.1068		t	−0.4619	
p-value	< 0.0001		p-value	0.6444	
Standard deviation	419.62	327.79	Standard deviation	398.12	343.82
F	1.6387		F	1.3408	
p-value	< 0.0001		p-value	0.0011	
Sum of Debts (euros)			*Number of Creditors*		
Mean	22702.09	20538.04	Mean	8.20	7.43
t	0.9762		t	2.3268	
p-value	0.3293		p-value	0.0201	
Standard deviation	28535.52	40083.23	Standard deviation	4.93	4.71
F	0.5068		F	1.0962	
p-value	< 0.0001		p-value	0.1695	
Number of Financial Creditors			*Number of Ordinary Creditors*		
Mean	2.9820	2.6186	Mean	5.222	4.8149
t	2.1630		t	1.3231	
p-value	0.0311		p-value	0.1861	
Standard deviation	2.5377	2.0347	Standard deviation	4.5585	4.3863
F	1.5563		F	1.0804	
p-value	< 0.0001		p-value	0.2100	
Duration of Judicial Proceedings (days)					
Mean	37.89	33.47			
t	3.3777				
p-value	< 0.0001				
Standard deviation	19.82	15.72			
F	1.5876				
p-value	0.0004				

of the debt, resources, current expenses, number of creditors, number of f nancial creditors, number of ordinary creditors, and duration of the judicial proceedings. We also report the results of t-tests to evaluate the difference in means between the two groups of debtors. The equality of variances is verif ed with the F-test. Finally, the p-level reported for both tests represents the probability of error involved in accepting hypothesis about the existence of a difference (in mean or variance). In this connection, we stress several differences between the two sets of debtors. First, the bankrupts whose debt discharge is refused by the judge present the highest values for resources, number of creditors, number of f nancial creditors, and duration of the judicial proceedings. Second, there is no significan difference between the means (in terms of debt size, current expenses, and number of ordinary creditors) of debtors who benef t from a debt release and those who do not. We explore these differences in the following section and take a closer look at the way the judges deal with personal bankruptcy.

3.2 Do Judges Punish Default?

In this section, we examine the possibility that judges *punish* debtors for overindebtedness – that is, they might penalize some debtors by denying them debt discharge if they have too many debts. We also note that, from an economic point of view, debt discharge may be interpreted as a punishment for creditors having granted too much credit.

To study this effect, we run a logit regression analysis where the dependent variable equals 1 if the judge discharges all the debt and zero otherwise.[7] More precisely, the logistic regression equation is interpreted as predicting the natural log of the odds of a f ler benef ting from debt discharge. In model 1, shown in Table 2, we control the judgment with the debtor's reimbursement capacity, the sum of debts, the duration of the judicial proceedings, a dummy variable *Ind*(*Unemployed*) that equals 1 when the debtor is unemployed and zero otherwise (i.e., employed or retired), and a dummy variable *Ind*(*Real estate*) that equals 1 when the debtor owns real estate and zero otherwise. Model 2 includes the number of creditors, and model 3 takes f xed effects into account with a set of 19 dummy variables indicating the location of the bankruptcy court (that is, the number, minus 1, of regions in our data set; for each region, we set a dummy variable that equals 1 when the case is judged in the region, and zero otherwise). In each model, we simply evaluate the debtor's capacity to reimburse debt from (future) income by computing the difference between monthly resources and monthly expenses. It is worth mentioning that values can be either positive or negative, because for some debtors expenses are higher than resources. All these variables are then used to explore, f rst, whether a larger amount of debts may signif cantly inf uence the bankruptcy court's decision, and second, to what extent judges might consider the case of debtors who seem to have overborrowed in terms of the number (rather than the amounts) of their claims. We also compute the marginal effects and the effects of discrete changes in independent variables on the probability of debt discharge (see Table 3). More precisely, the marginal effect and quasielasticity are calculated for the continuous variables (*Reimbursement capacity*, *Sum of debts*, *Number of creditors*,

[7] See appendix A.2 for a logit multinomial regression analysis including the 35 debtors identifie as being mala f de. All key results are robust to this change in the data set. We have also run some regression analysis of the judge's decision where the mala or bona fide serves as an explanatory dummy variable of the judge's decision to discharge debts or not. Our results are also robust to this change. We do not report this result here.

Table 2

Explanation of Debt Discharge (number of debt discharges: 805; number of debt discharges denied: 279)

Explanation of the probability of debt discharge	Model 1: 1084 obs. Debt discharge versus No debt discharge		Model 2: 1084 obs. Debt discharge versus No debt discharge		Model 3: 1084 obs. Debt discharge versus No debt discharge	
Variables	Estimation	Prob. $> \chi^2$	Estimation	Prob. $> \chi^2$	Estimation	Prob. $> \chi^2$
Constant	0.3608	0.5753	0.6096	0.3549	−5.3549	0.2103
Reimbursement capacity	−0.00364***	< 0.0001	−0.00362***	< 0.0001	−0.00360***	< 0.0001
Sum of debts	2.75E−6	0.3010	4.51E−6*	0.0989	4.30E−6	0.1146
Number of creditors			−0.0508***	0.0011	−0.0532***	0.0011
Ind(Unemployed)	0.1203	0.4380	0.1486	0.3426	0.1153	0.4801
Ind(Real estate)	0.8648	0.1575	0.9528	0.1261	1.0831	0.1178
Duration of judicial proceedings	−0.0138***	0.0013	−0.0135***	0.0020	−0.0121**	0.0101
Fixed effects	no		no		yes	
Logit Regression	% concordant:	74.1	% concordant:	74.3	% concordant:	76.7
	Condition index:	5.63 < 30	Condition index:	6.80 < 30		
Estimation Method	Test		Test		Test	
Maximum likelihood	Likelihood	136.37 < 0.0001	Likelihood	146.80 < 0.0001	Likelihood	188.31 < 0.0001
	Score	122.63 < 0.0001	Score	132.99 < 0.0001	Score	168.01 < 0.0001
	Wald	101.32 < 0.0001	Wald	108.50 < 0.0001	Wald	130.77 < 0.0001

Note: We do not report the estimations for the fi ed-effect dummy variables (notice that seven variables are significan at the 5% level, and two variables are significan at the 10% level). Collinearity diagnostic: if condition index > 30 then there is strong collinearity. The sample is described in section 3.1. Coeff cients significan at the 1%, 5%, and 10% levels are indicated by ***, **, and *, respectively.

Table 3

Effects on Predicted Probability of Debt Discharge

Marginal Effects for Model 2 (without fixe effects)

Effect on predicted probability (for Ind(Unemployed) = 0)				Effect on predicted probability (for Ind(Unemployed) = 1)		
Variables	Discrete	Marginal effect	Quasielasticity	Variables	Marginal effect	Quasielasticity
Reimbursement capacity		< −0.001	−0.0897	Reimbursement capacity	< −0.001	−0.0879
Sum of debts		< 0.001	0.0235	Sum of debts	< 0.001	0.0230
Number of creditors		−0.0125	−0.0956	Number of creditors	−0.0123	−0.0937
Ind(Unemployed)	0.0364					
Duration of judicial proceedings		−0.0033	−0.1155	Duration of judicial proceedings	−0.0032	−0.1131

Marginal Effects for Model 3 (with fixe effects)

Effect on predicted probability (for Ind(Unemployed) = 0)				Effect on predicted probability (for Ind(Unemployed) = 1)		
Variables	Discrete	Marginal effect	Quasielasticity	Variables	Marginal effect	Quasielasticity
Reimbursement capacity		< −0.001	−0.0016	Reimbursement capacity	< −0.001	−0.0018
Sum of debts		< 0.001	< 0.001	Sum of debts	< 0.001	< 0.001
Number of creditors		< −0.001	−0.0018	Number of creditors	< −0.001	−0.0021
Ind(Unemployed)	< 0.001					
Duration of judicial proceedings		< −0.001	−0.0019	Duration of judicial proceedings	< −0.001	−0.0021

and *Duration of judicial proceedings*), and the binary variable *Ind*(*Unemployed*) is changed by one unit to observe the effect on probability. Due to the nonlinearity of the logit model, we have to take into account both the level of the analyzed variable and the level of all other explanatory variables. So the reimbursement capacity, the sum of debts, the number of creditors, and the duration of the judicial proceedings were set to their mean values. The dummy variable *Ind*(*Unemployed*) was assigned a value of either zero (see left side of Table 3) or 1 (see right side of Table 3). The dummy variable *Ind*(*Real estate*) was set to zero. Finally, we set each dummy variable indicating whether the debtor has f led for bankruptcy in a particular region (or not) to 1, and we multiplied each one by the proportion of cases judged in this region in our data set (that is, the number of cases judged in the region divided by the number of observations).

Obviously, reimbursement capacity is the most significa t factor to explain the probability of debt discharge. It suggests that those debtors who are least capable of repaying their debts (or those who are more f nancially distressed) have a greater probability of benef ting from a fresh start.[8] Further, our f ndings demonstrate that the effect of a one-percent change in the sum of debts on the predicted probability is much less: It increases the probability of debt discharge by as little as 0.023, or even less if we control for fi ed effects (see Table 3).

The main f nding in Table 2 is that the coeff cient associated with the number of creditors is negative and statistically signif cant (with a p-value of 0.0011). Moreover, if the number of creditors increases by one percent, then the predicted probability of debt discharge decreases by around 0.09 without f xed effects, or 0.0018 when controlling for f xed effects (i.e., a similar variation to the one resulting from an increase by one percent of the reimbursement capacity). This means that having a large number of creditors reveals more than does the debtors' level of f nancial distress. Judges might consider that having multiple creditors indicates that debtors have failed to balance their budget, or overborrowed. In other words, our data illuminate the extent to which specialized judges are inf uenced by their sense that debtors are responsible for their f nancial situation. For instance, as stated by Bizer and DeMarzo (1992), debtors would have an incentive to borrow from multiple lenders in order to borrow greater amounts (superior to those obtained from a unique lender). But, under such a sequential lending, the likelihood of default would then increase because borrowers may take actions that lower the probability of reimbursement of all the loans (i.e., a moral hazard problem). However, if we control for the structure of claims (meaning that we distinguish between fi ancial and ordinary creditors), we do not f nd whether it is the number of f nancial creditors rather than the number of ordinary creditors (or the reverse) that tends to disqualify debtors from debt discharge, because neither of these variables is significa t (we do not report the regression results in this paper). So, this result undermines the intuition

[8] We have controlled for the amount of resources, expenses, and debt separately, but we did not include these variables in the same regression model, because of the high degree of correlation between resources and expenses.

that the number of financial debts (mainly debt consumption or credit cards) may motivate bankruptcy courts to refuse debt discharge. Put simply, it is not very clear how the difference between financial and ordinary debts affects the judge's decision. Finally, recall that before filing for this bankruptcy procedure, debtors would have attempted to renegotiate their debts under the supervision of a CSUR, and have failed. Due to this sample selection, it is also plausible that debtors with a larger number of creditors are excluded from debt discharge, because the judge considers that they have failed to reach a compromise owing to the large number of parties involved. More precisely, there could be some debtors who are able to repay their debts under an appropriate repayment plan, but have failed to renegotiate such a plan (due to a large number of creditors) and thus are filing for the PRPs. In that case, the judge may order a repayment plan rather than a full debt discharge, but it is quite difficult empirically to resolve this case selection problem, due to the lack of data on the work of the CSUR.

In addition, Table 2 contains another interesting result that has to be interpreted with care. The duration of the judicial proceedings seems to play a role in the court's decision-making. Here, we consider that the duration of the judicial proceedings is closely correlated with the overwhelming number of personal bankruptcy filings with which the courts have to deal. Table 2 clearly shows that the longer the duration of the judicial proceedings, the smaller the likelihood of debt discharge. That means that debtors do not benefit from judges' leniency when judicial delays increase due to the rise in personal bankruptcy filings and the courts' congestion. However, some other factors may explain the duration of the judicial proceedings. First of all, it may be due to some strategic behavior by lenders or borrowers within the procedure, such as voluntary delays in furnishing information or documents, or claimants' requests to put off the judge's decision. Further, the duration of the judicial proceedings may also be linked to the complexity of the case. The most complex cases (or the longer-lived cases under this interpretation) would have a lower probability of debt discharge, because judges consider that the debtors should reimburse their debts, and that the negotiation on the repayment plan has failed earlier due to the complexity of the case. In other words, the negative relationship between the duration and the likelihood of debt discharge would arise from a case selection problem. Finally, as a robustness test, we exclude the duration of the legal process from the regression analysis, and we show in Table A2, in appendix A.2, that key results are similar.

3.3 Judges' Decisions and Macroeconomic or Social Context

To complete our analysis, we explore whether judges' decisions to discharge debts are influenced by the external environment of the bankruptcy case. More precisely, we test the sensitivity of the probability of debt discharge to a set of variables, representative of economic activity in the court's locality: unemployment rate, disposable income per head, growth rate of disposable income per head, gross domestic product per head, growth rate of gross domestic product per head, and real estate price (source: INSEE, French National Institute of Statistics and Economic Studies, www.insee.fr). To do so, we first compute for each variable the difference between its value in the court's locality (one of 20 French regions) and the mean national value of the variable questioned (except for the real estate price). More precisely, Δ*Unemployment rate* equals the difference between the unemployment rate in the court's locality and the mean unemployment rate in the country during the period studied in this paper. Δ*Disposable income per head* equals the difference between the disposable income per head in the court's locality and the mean disposable income per head in the country during this period. We calculate the same difference for the growth rate in the period 2002–2007 (see Δ*Growth rate of disposable income per head*). Δ*GDP per head* equals the difference between the growth in domestic product per head in the court's locality and the mean growth in domestic product per head in the country. We compute the same difference for the growth rate in the period 2002–2007 (see Δ*Growth rate of GDP per head*). To capture the effect of real estate prices on the judge's decision, we compute the dummy variable *Ind(Real estate price)* that equals one if the real estate price in the court's locality is larger than the mean real estate price in the country, and zero otherwise. At the next step, we report, for each debtor in the data set, these differences (one difference for each variable) according to the locality that settles the bankruptcy case. This new set of six variables then serves as explanatory variables for the judge's decision (see model 6 in Table 4). Second, for easier presentation, we focus on a single variable representative of the local economic context (unemployment rate) by setting an additional dummy variable (*Ind*(*Unemployment rate*)) that equals 1 if the unemployment rate in the court's locality is larger than the mean unemployment rate in the country during the period studied in the paper, and zero otherwise. In this way, we obtain a better look at the marginal effect on the probability of debt discharge of a change from a region where local unemployment is low to a region where it is high. Table 4 shows this new set of regressions, and the marginal effects on the predicted probability of debt discharge can be observed in Table 5 (see section 3.2 for a discussion of the methodology).

We first consider the regression results reported in Table 4: They confirm that judges tend to disqualify debtors with multiple debts from full debt discharge. Further, the regression estimates indicate that the higher the regional unemployment rate (in comparison with the mean rate in the country), the higher the likelihood of debt discharge. The reverse is true for the growth rate of disposable income per head. Regarding the marginal effects, if a debtor moves from a region where the unemployment rate is smaller than the mean national employment rate to a region where the reverse is true, then the predicted probability of debt discharge increases by 0.1976 (the strongest effect described in this paper). As a result, judges are more likely to enforce debt discharge in regions where there is a clear shortage of employment, or where debtors face increased risk of an adverse event (job loss) in the future (by controlling for the current debtor's employment status).

Our findings supplement the empirical results, thereby demonstrating that economic conditions do influence the decision of a judge. For instance, Ichino, Polo, and Rettore (2003) demonstrate that judges acting in Italian labor courts are more likely

Table 4

Explanation of Debt Discharge Related to External Environment (number of debt discharges: 805; number of debt discharges denied: 279)

Explanation of the probability of debt discharge	Model 4: 1084 obs. Debt discharge versus No debt discharge		Model 5: 1084 obs. Debt discharge versus No debt discharge		Model 6: 1084 obs. Debt discharge versus No debt discharge	
Variables	Estimation	Prob. > χ^2	Estimation	Prob. > χ^2	Estimation	Prob. > χ^2
Constant	0.5423	0.4025	0.1686	0.8015	0.5695	0.4022
Reimbursement capacity	−0.00356***	< 0.0001	−0.00367***	< 0.0001	−0.00358***	< 0.0001
Sum of debts	4.64E−6*	0.0894	5.42E−6*	0.0524	5.01E−6*	0.0745
Number of creditors	−0.0488***	0.0018	−0.0519***	0.0009	−0.0496***	0.0017
Ind(Unemployed)	0.1629	0.3030	0.1520	0.3370	0.1731	0.2828
Ind(Real estate)	0.9846	0.1072	1.0464	0.1054	1.0828*	0.0840
Duration of judicial proceedings	−0.0106**	0.0165	−0.0109**	0.0118	−0.0110**	0.0143
ΔUnemployment rate	25.9999***	0.0002			17.2344*	0.0553
Ind(Unemployment rate)			0.8175***	< 0.0001		
ΔDisposable income per head					5.1E−5	0.8203
ΔGrowth rate of disposable income per head					−78.16***	0.0034
ΔGDP per head					−1,00E−05	0.8307
ΔGrowth rate of GDP per head					46.3632	0.2357
Ind(Real estate price)					0.2292	0.2568
Logit Regression	% concordant:	74.9	% concordant:	75.2	% concordant:	75.2
	Condition index:	6.89 < 30	Condition index:	7.40 < 30	Condition index:	19.89 < 30
Estimation Method	Test		Test		Test	
Maximum likelihood	Likelihood	161.87 < 0.0001	Likelihood	170.42 < 0.0001	Likelihood	172.99 < 0.0001
	Score	147.35 < 0.0001	Score	153.96 < 0.0001	Score	157.36 < 0.0001
	Wald	120.11 < 0.0001	Wald	125.32 < 0.0001	Wald	126.95 < 0.0001

Notes: As we control for differences between courts' localities at several levels (unemployment rate, GDP, etc.), we do not introduce any dummy variables for f xed effects as in the previous regression analysis. Collinearity diagnostic: if condition index > 30 then there is strong collinearity. The sample is described in section 3.1. Coeff cients signif cant at the 1%, 5%, and 10% levels are indicated by ***, **, and *, respectively.

Table 5

Effects on Predicted Probability of Debt Discharge (related to regression model 5)

Marginal Effects for Model 5							
Effect on predicted probability (for Ind(Unemployed) = 0)				Effect on predicted probability (for Ind(Unemployed) = 1)			
Variables	Discrete	Marginal effect	Quasielasticity	Variables	Discrete	Marginal effect	Quasielasticity
Reimbursement capacity		< -0.001	-0.0917	Reimbursement capacity		< -0.001	-0.0920
Sum of debts		< 0.001	0.0284	Sum of debts		< 0.001	0.0286
Number of creditors		-0.0129	-0.0985	Number of creditors		-0.0129	-0.0988
Ind(Unemployed)	0.0379						
Duration of judicial proceedings		-0.0027	-0.0939	Duration of judicial proceedings		-0.0027	-0.094
Ind(Unemployment rate)	0.1976			Ind(Unemployment rate)	0.1925		

to decide in favor of workers when and where unemployment is higher. Further, Marinescu (2011) gives empirical evidence from British employment courts that judges are more pro-f rm when the unemployment rate is raised. All these results converge to support the idea that judges do not merely enforce debt or employment contracts: they also maximize the joint welfare of the trial parties. In our context, judges may not only be concerned with the monetary sanction when they force debtors to reimburse their debt. Judges may also take into consideration the debtor's welfare when they allow a fresh start for the most f nancially distressed debtors, especially when debtors are facing worse economic conditions (here, higher unemployment and lower growth rate in disposable income per head). However, there is another possible explanation for our last result. Judges may not directly consider the economic conditions (as explained above), but instead the case quality (here in an unobserved way), since debtors may be more likely to be bona f de (or may have better bankruptcy cases) when the unemployment rate is higher and, likewise, when their ability to repay is lower. But it seems to us unlikely that such heterogeneity in bankruptcy case quality is not reflecte in any way in our measured variables.

Finally, our results may also be linked to another strand of research. As noted in the introduction, some authors recently have observed significan differences in personal bankruptcy practices across localities or courts in the U.S. The closest paper to our analysis is perhaps the work of Norberg and Compo (2007). These authors analyzed how the debt discharge rate (and not directly the case decisions as in our paper) differs across U.S. districts, and across judges in the same district.[9] They also propose two additional explanations of the differences between localities in personal bankruptcy outcomes. Here, the identity of the attorney and whether the debtor is represented by an attorney or not may also affect the bankruptcy outcome, because the lawyers (attorney and judge) would locally drive the American debtor's choice between Chapter 7 and Chapter 13. Further, such differences in bankruptcy outcomes would result from a sociological aspect of bankruptcy law, namely, the local legal culture (not explored in the present paper). Again, the diff culty is to identify the factors that inf uence the lawyers' decision-making, especially their preferences concerning personal bankruptcy. According to these authors, lawyers would not only defend the fi ancial interest of their client. The same lawyers would also have some moral concerns about the bankruptcy process. For instance, the lawyers most involved in the fresh-start cases (or Chapter 7) would seek to maximize the fi ancial interest of debtors, whereas the lawyers most involved in reimbursement plans (or Chapter 13) would be more concerned with what is *morally* right (here, to reimburse the debts). In the same way, American bankruptcy judges might also infl ence the debtor's choice. They might set the amount of attorneys' fees contingent on the bankruptcy procedure (Chapter 7 versus Chapter 13) in order

[9] Notice that it is difficul to draw immediate comparisons between their empirical analysis and ours, simply because French and U.S. personal bankruptcy laws differ: for instance, American debtors have much more discretion in the choice between the reimbursement plan (Chapter 13) and the fresh start (Chapter 7) than French debtors.

to give attorneys a financia incentive to choose the “right” procedure according to their own view or preference.

4 Conclusion

As White (2007) noted in her cross-country analysis, it is generally agreed that French personal bankruptcy law tends to protect creditors’ interests rather than debtors’. But such a conclusion is debatable, because debt discharge ultimately depends on a judge’s ruling on a case. In this paper, we have explored how French judges decide whether or not debts are discharged and nonexempt assets are liquidated through PRPs. In particular, it appears that in the period 2004–2005, more than one-third of borrowers who f led for a fresh start or PRP were denied debt discharge. It was surprising that all these debtors were previously identif ed as f -nancially distressed by an administrative authority or CSUR (meaning that there was no chance of arranging a rescheduled debt payment, on the assumption that the CSUR makes no f ltering errors). Second, we f nd in our sample that a great majority of debtors who f led for a PRP had no or very few assets (real estate or other) to liquidate, mainly because these debtors were previously selected by the CSUR to get a PRP. Third, we also explore judicial criteria. We show that a debtor’s reimbursement capacity is the judge’s major consideration in the decision to discharge debts. More interestingly, we f nd that judges refuse debt discharge when debtors are indebted to multiple creditors. As a consequence, we suggest that judges may consider that some borrowers are responsible for their f nancial distress or overborrowing. In that case, the lower the probability of discharging the debt, the more the creditors (fi ancial or not) are protected from default. This could give f -nancial creditors some incentive to increase access to credit, at the risk of increasing the probability of overborrowing when an adverse event occurs. Finally, we show that it is necessary to control our estimate of the probability of debt discharge with some indicators of the macroeconomic context in which judges view the case. In particular, we f nd great statistical support for the hypothesis that French judges’ decision-making is inf uenced by the unemployment rate and the growth rate of disposable income per head in their locality. In summary, our observations on the French bankruptcy liquidation system (we do not explore the reorganization system) show that even if all countries appear to be acting similarly in their treatment of personal bankruptcy law, we need to conduct more realistic studies to better assess the work of courts in each country in order to f ll the gap between bankruptcy rules and judicial practice.

To conclude, there remains a large set of questions yet unanswered. To our knowledge, there is no cross-country analysis that relates the various national personal bankruptcy systems with specifi forms of bank debt contracts (size of loan, level and type of collateral, interest rate, duration). As in corporate bankruptcy law, the differences in lenders’ legal protection across these countries should correlate with signif cant differences in lenders’ strategies and outcomes. More simply, the design of personal bankruptcy law might help us to understand lenders' recovery rates or the success of informal renegotiations (preceding a bankruptcy f ling). We believe that sustained research in this direction can help f nd answers to these questions, thereby contributing to the development of the personal bankruptcy literature.

Appendix

A.1 Evolution of the Number of Bankruptcy Filings

In Table A1, we make the distinction between debtors who benef t from a reimbursement plan and debtors oriented toward a PRP, for the years 1993–2011 (source: Banque de France). More precisely, we report in the f rst column the annual number of f lers for a bankruptcy procedure (reschedule plan and PRP together). The second column indicates the percentage of fi ers who benefi from either a reimbursement plan or a PRP. The third column indicates the annual number of debtors who f le for a PRP (data are available only since 2003).

Table A1
Evolution of Bankruptcy Filings in the Period 1993–2011

Year	Number of bankrupts	Percentage of acceptance by CSUR	Number of debtors oriented to a PRP
1993	68 863	82.8	–
1994	68 608	86	–
1995	70 112	80.4	–
1996	86 999	82.3	–
1997	95 756	83.7	–
1998	117 854	80.1	–
1999	142 219	75.3	–
2000	148 373	84.6	–
2001	137 994	86	–
2002	145 348	81.6	–
2003	165 493	87.2	–
2004	188 176	81.4	22 034
2005	182 330	85.5	22 187
2006	184 866	85.4	27 504
2007	182 855	84.7	30 745
2008	188 485	84.9	34 919
2009	216 396	84.4	42 704
2010	217 325	82.9	44 361
2011	232 493	87.3	65 776

A.2 Robustness Analysis

In this appendix, we run two sets of robustness analyses. In the f rst, we exclude the duration of the judicial proceedings from the regression analysis (see Table A2). Key results are robust to this change in the set of explanatory variables. In the second, we include in the database 36 debtors who were identif ed as *mala fid* by the judge, and thus precluded from debt discharge. We also modify the bankruptcy outcome. Now, the bankruptcy outcome is either debt discharge for a bona f de debtor, no debt discharge for a bona fid debtor, or no debt discharge on account of the debtor's mala f des (see Table A3). Key results are robust to this change in the endogenous variable.

Table A2

Explanation of Debt Discharge Excluding the Duration of the Judicial Proceedings (number of debt discharges: 805, number of debt discharge denied: 279)

Explanation of the probability of debt discharge	Model 7: 1084 obs. Debt discharge versus No debt discharge			Model 8: 1084 obs. Debt discharge versus No debt discharge		
Variables	Estimation	Prob. > χ^2		Estimation	Prob. > χ^2	
Constant	−5.7561	0.1756		0.1390	0.8297	
Reimbursement capacity	−0.0036***	< 0.0001		−0.0036***	< 0.0001	
Sum of debts	4.11E−6	0.1221		4.87E−6*	0.0760	
Number of creditors	−0.0544***	0.0007		−0.0512***	0.0011	
Ind(Unemployed)	0.0843	0.6035		0.1462	0.3616	
Ind(Real estate)	1.1600*	0.0865		1.1161*	0.0682	
ΔUnemployment rate				19.9341**	0.0249	
ΔDisposable income per head				1.9E−5	0.9333	
ΔGrowth rate of disposable income per head				−74.0036***	0.0052	
ΔGDP per head				−1.59E−6	0.9799	
ΔGrowth rate of GDP per head				30.5186	0.4273	
Ind(Real estate price)				0.2261	0.2614	
Fixed effects	yes			no		
Logit Regression	% concordant:	76.3		% concordant:	74.8	
				Condition index:	19.38	< 30
Estimation Method	Test			Test		
Maximum likelihood	Likelihood	181.59	< 0.0001	Likelihood	119.91	< 0.0001
	Score	161.55	< 0.0001	Score	151.07	< 0.0001
	Wald	127.24	< 0.0001	Wald	123.48	< 0.0001

Note: See section 3.3 for a description of the variables. Collinearity diagnostic: if condition index > 30 then there is strong collinearity. The sample is described in section 3.1. Coeff cients significa t at the 1%, 5%, and 10% levels are indicated by ***, **, and *, respectively.

Table A3
Explanation of Debt Discharge Including the Mala Fide Debtors

Explanation of the probability of debt discharge	Model 9: 1120 obs. Output = debt discharge (ref. no debt discharge for bona fid debtors)		Output = no debt discharge (ref. no debt discharge for bona fid debtors)	
Variables	Estimation	Prob. > χ^2	Estimation	Prob. > χ^2
Constant	0.7653	0.2693	−16.5435	0.9734
Reimbursement capacity	−0.00351***	< 0.0001	−0.00245***	0.009
Sum of debts	4.80E−6*	0.0829	2.571E−6	0.6795
Number of creditors	−0.0458***	0.0028	0.0663**	0.012
Ind(Unemployed)	0.1971	0.2177	−0.0695	0.8592
Ind(Real estate)	1.0740*	0.0850	12.5354	0.9798
Duration of judicial proceedings	−0.0111**	0.0113	0.0135	0.1137
ΔUnemployment rate	16.1891*	0.0702	5.9944	0.7756
ΔDisposable income per head	3.9E−5	0.8590	−6.8E−4	0.2280
ΔGrowth rate of disposable income per head	−78.2872***	0.0029	161.6**	0.0244
ΔGDP per head	−9.04E−6	0.8863	0.0001	0.5138
ΔGrowth rate of GDP per head	43.1292	0.2615	−182.2*	0.0891
Ind(Real estate price)	−0.2401	0.2342	−0.6163	0.1972
Multinomial Independent Logit Regression	Number of debt discharge:			805
	Number of debt discharges denied by judge (for other reasons than mala f de):			279
	Number of debt discharges denied by judge due to debtor's mala f de:			36
Estimation Method	Test			
Maximum likelihood	Likelihood	211.06	< 0.0001	
	Score	202.02	< 0.0001	
	Wald	160.47	< 0.0001	

Note: See section 3.3 for a description of the variables. The sample is described in section 3.1. Coeff cients signif cant at the 1%, 5%, and 10% levels are indicated by ***, **, and *, respectively.

References

Banque de France (2008), *Enquête typologique 2007 sur le surendettement*, Secrétariat Général, DSRP, Paris.

Bizer, D. S., and P. M. DeMarzo (1992), "Sequential Banking," *Journal of Political Economy*, 100(1), 41–61.

Braucher, J. (1993), "Lawyers and Consumer Bankruptcy: One Code, Many Cultures," *The American Bankruptcy Law Journal*, 67(4), 501–583.

Dick, A. A., and A. Lehnert (2010), "Personal Bankruptcy and Credit Market Competition," *The Journal of Finance*, 65(2), 655–686.

Fay, S., E. Hurst, and M. J. White (2002), "The Household Bankruptcy Decision," *The American Economic Review*, 92(3), 706–718.

Filer, L., and J. D. Fisher (2007), "Do Liquidity Constraints Generate Excess Sensitivity in Consumption? New Evidence from a Sample of Post-Bankruptcy Households," *Journal of Macroeconomics*, 29(4), 790–805.

Filer II, L. H., and J. D. Fisher (2005), "The Consumption Effects Associated with Filing for Personal Bankruptcy," *Southern Economic Journal*, 71(4), 837–854.

Gan, L., and T. Sabarwal (2005), "A Simple Test of Adverse Events and Strategic Timing Theories of Consumer Bankruptcy," Working Paper No. 11763, National Bureau of Economic Research.

Gropp, R., J. K. Scholz, and M. J. White (1997), "Personal Bankruptcy and Credit Supply and Demand," *The Quarterly Journal of Economics*, 112(1), 217–252.

Ichino, A., M. Polo, and E. Rettore (2003), "Are Judges Biased by Labor Market Conditions?" *European Economic Review*, 47(5), 913–944.

Lusardi, A., and P. Tufano (2008), "Debt Literacy, Financial Experiences, and Overindebtedness," Working Paper No. 14808, National Bureau of Economic Research.

Marinescu, I. (2011), "Are Judges Sensitive to Economic Conditions? Evidence from the U.K. Employment Tribunals," *Industrial and Labour Relations Review*, 64(4), 673–698.

Musto, D. K. (2004), "What Happens when Information Leaves a Market? Evidence from Postbankruptcy Consumers," *The Journal of Business*, 77(4), 725–748.

Norberg, S. F., and N. S. Compo (2007), "Report on Empirical Study of District Variations, and the Role of Judges, Trustees and Debtors' Attorneys in Chapter 13 Bankruptcy Cases," *The American Bankruptcy Law Journal*, 81(4), 101–158.

Rachlinski, J. J., C. Guthrie, and A. J. Wistrich (2007), "Heuristics and Biases in Bankruptcy Judges," *Journal of Institutional and Theoretical Economics (JITE)*, 163(1), 167–186.

Sullivan, T. A., E. Warren, and J. L. Westbrook (1994), "The Persistence of Local Legal Culture: Twenty Years of Evidence from the Federal Bankruptcy Courts," *Harvard Journal of Law & Public Policy*, 17(3), 801–865.

White, M. J. (2007), "Bankruptcy Reform and Credit Cards," *The Journal of Economic Perspectives*, 21(4), 175–199.

Régis Blazy
EM Strasbourg Business School
and IEP Strasbourg
47 avenue de la Forêt Noire
67082 Strasbourg
France
regis.blazy@unistra.fr

Bertrand Chopard
Université Paris Ouest
Nanterre la Défense
200 avenue de la République
92001 Nanterre
France
Bertrand.Chopard@u-paris10.fr

Eric Langlais
Université Paris Ouest
Nanterre la Défense
200 avenue de la République
92001 Nanterre
France
elanglais@u-paris10.fr

Ydriss Ziane
Sorbonne Graduate Business School
21 rue Broca
75240 Paris
France
ydriss.ziane@univ-paris1.fr

Innovation, Tort Law, and Competition

by

Florian Baumann and Klaus Heine*

Received February 16, 2012; in revised form December 3, 2012;
accepted June 8, 2013

We examine the link between innovative activity on the part of f rms, the competitive pressure to introduce innovations, and optimal damages awards. While innovative activity brings forth valuable new products for consumers, competitive pressure in the ensuing innovation race induces fi ms to launch innovations too early, thereby raising the likelihood of severe product risks above the optimal failure rate. Introducing innovations too early may call for the application of punitive damages instead of mere compensation of harm caused, in order to decelerate such welfare-reducing innovation races. (JEL: K13, L13, O31)

1 *Innovation, Tort Law, and Competition: A 'ménage à trois'*

There is clearly a link between a f rm's innovative activity, competitive pressure, and the tort law regime: one might intuitively consider that a f rm's innovative activity is stimulated by competition, while the threat of tort liability dampens a f rm's appetite for innovation. For example, a pharmaceutical fir   may feel competitive pressure to be the f rst on the market with a new kind of drug. But at the same time the fir   may fear that the drug could cause adverse reactions, which may result in tort cases and subsequently high compensation payments, and even punitive damages. For instance, on 18 April 2012 jurors in federal court in New Haven ruled for a damages award payable by Pfize Inc. of at least $4 million to a woman who used the menopause drug Prempro distributed by Pf zer's Wyeth unit. The drug, which was assessed as an "unreasonable dangerous product," was assumed to have caused

* Düsseldorf Institute for Competition Economics, Heinrich Heine University, Düsseldorf, Germany, and Erasmus School of Law, Erasmus University, Rotterdam, The Netherlands. Corresponding author is Klaus Heine. We would like to thank Biung-Ghi Ju, two anonymous referees, Alexander Morell, Wendy Smith, the participants of the Annual Conference of the German Association of Law and Economics 2011, and the participants in the annual meeting of the International Society for New Institutional Economics 2012 for their highly appreciated comments.

Journal of Institutional and Theoretical Economics 169, 703–719 – ISSN 0932-4569
DOI: 10.1628/093245613X671201

the plaintiff's cancer.[1] Many firm may feel the tension between the prospective rewards of being the winner in a race for innovation and being cursed as a tortfeasor, if it turns out later that the innovation triggers unforeseen harm. The Ford Pinto case may illustrate this further. In 1968 the Ford Motor Company management felt the pressure to come up with a subcompact car, in order to regain market share in this market segment against Volkswagen and Japanese manufacturers. The Ford Pinto was developed on an accelerated schedule, whereby the management set the strategy "limits for 2000" in the initial production and testing phase (Leggett, 1999). The production cost was not to exceed $2000 and the weight not to exceed 2000 pounds. As a consequence a tank and fuel fille were installed that were according to the industry standard but that would break at a speed above 31 miles per hour. A safer technical solution would have cost $11 more. A cost–benef t analysis was undertaken that came to the conclusion that it was justifie to use the simpler and cheaper solution and to run the risk of more severe accidents. As it turned out, a lot of deadly fir accidents happened, due to the Pinto's fuel system. But only in 1977, after eight years, did Ford make some modifica ions to the fuel system. This means that early knowledge of a potential harm and the later realization of harm by no means guarantee that the selling or using of a defective product stops. The Pinto case raised a lot of fundamental questions about the correct design and application of tort law (Leggett, 1999). However, here it is important that Lee Iacocca, Ford's CEO, was fond of the statement "safety doesn't sell," which may be interpreted as saying that competitive pressure and the need of timely product innovations may outweigh safety issues from a firm s perspective.

Against the background of the given examples the research question of this article can be define as follows: How should the optimal damages amount be set in order to provide appropriate incentives regarding safety of newly invented products if there are competitors also striving for the winner's cup, and an early introduction of the innovation is associated with inf ated risks of product failure and consequent harm?

One may discuss why there is almost no literature on this important question. An explanation could be that in the economic analysis of tort law it is implicitly assumed that technological development is indirectly dependent on the inherent incentives provided by a liability rule to acquire knowledge about hidden risks (Shavell, 1992). However, the analysis of innovations and tort law remains diff cult. For example, under negligence liability in a tort case, it is often not possible for a court to verify whether a f rm has chosen the appropriate level of effort with regard to research and development, in order to bring potential harm down to an optimal level (Dari-Mattiacci and Franzoni, 2013). In addition, there are seemingly subtle trade-offs between the *activity level* of and the *due standard of care* for tortfeasors (firm) and victims (consumers), which cannot be overcome easily. In the end one may come

[1] "Pfi er Ordered to Pay $4 Million in Damages in Prempro Case"; see http://www.businessweek.com/news/2012-04-20/pfize -ordered-to-pay-4-million-in-damages-in-prempro-case.

to the conclusion that it is not possible to design a liability rule that gives optimal incentives to tortfeasors and victims at the same time (Shavell, 1980).

While there is not yet much literature on the effect of liability rules on firms' innovative activity in general (Parchomovsky and Stein, 2008, p. 286), there is some literature on the effect of punitive damages on innovation. This small literature mainly plays out against the legal background of the United States, where punitive damages are part of the tort law system.[2] From this literature one gains the impression that contemporary tort law regimes are seen as a hindrance to the socially beneficial amount of innovation, implying that there is too little innovation because of overdeterrence. The implied policy conclusion is, then, that there should be a tort law reform that removes those features from tort law that trigger overdeterrence (Huber, 1985; Parchomovsky and Stein, 2008; Shavell and Van Ypersele, 2001; Shavell, 2004).

However, overall no clear picture emerges; while some have called for punitive damages to be abolished, others have praised them, and others still have pled for a more balanced system of punitive damages. For example, Mahoney and Littlejohn (1989) claim that strict liability, in conjunction with huge jury awards and "uncontrolled" punitive damages in addition, creates immense legal uncertainty for innovators. As a result, innovators abstain from introducing new and safer products, in order to avoid punitive damages. The authors present as confirmative examples basic research, health care, and aviation, where the rate of innovations has dropped over the years and smaller companies in particular have had to close their businesses. In a more recent article, Epstein (2006) echoes this opinion, particularly with regard to pharmaceutical innovations.

It is argued in favor of punitive damages that they stimulate research on safer technologies, which is seen as directly socially beneficial for consumers. Moreover, relaxing punitive damages would mean that more explicit criminal sanctions ought to be applied, in order to deter managers effectively from introducing unsafe products (Rustad, 1992; Daniels and Martin, 1990).

Between these two camps there are some authors who hold an intermediate position. One of these is Viscusi (1998), who argues that disproportionately high punitive damages suppress innovation and may lead to the withdrawal of firms from the market, while more modest and tailor-made punitive damages may set incentives for firms to innovate and to search for safer technologies (see also Viscusi and Moore, 1993). However, the question of how high punitive damages should be in a particular case is left open. Rather, it is proposed that punitive damages should be abandoned completely and that a compensatory damages regime should be constructed that satisfies economic criteria and sets efficient incentives for potential tortfeasors (Viscusi, 1998). This conclusion is somewhat unsatisfactory, and in this contribution we will try to give an answer to the problem of which factors should determine optimal damages.

---

[2] This is in contrast to continental European law, where punitive damages are not seen as part of the tort law regime and do not even exist (see also section 5).

In doing so, we argue that taking into account the fierceness of competition between innovators plays a decisive role in achieving a better understanding of whether or not innovations are welfare-enhancing in a torts context. Our main finding is that competition forces innovators to introduce innovations too early, thereby raising the risk of harm for buyers. This finding suggests that inflated or punitive damages are an appropriate means of placing additional costs on innovations and slowing down the innovation race. That is, punitive damages prevent innovations being introduced prematurely, and will lead to a higher chance of defects of innovations being detected in time.

We derive our findings within a stylized model, in which we assume a *time–cost trade-off* for introducing innovations (see, e.g., Scherer and Ross, 1990), meaning that introducing innovations earlier results in higher costs. In addition, planning to introduce the product at an earlier date implies that the probability of harm arising from using the product will be higher, as less time will have been spent investigating the associated risks. With respect to competition, a *patent race* is assumed, in which the winner wins a monopoly, while the competitors receive nothing (see, e.g., Loury, 1979). Firms choose the speed at which they try to introduce a new innovation, which results in a hazard or success rate, which is modeled analogously to that used by Kamien and Schwartz (1972).

Even though the model presented in the next section is simple, it captures various important features of a firm's strategic decision-making process with regard to innovation, competition, and tort law. Moreover, the model is the first to introduce competitive pressure as an important variable in the relation between innovation and tort liability. We argue that the chosen time to market of an innovation depends not only on the incentives that are set by a torts regime, but also on those set by the interaction between a firm and its competitors. Fierce competition may induce a firm to develop an innovation as early as possible, fearing that another firm will come up with the innovation and take over the market. However, these innovation races come at a price. The innovation costs increase if the date of the innovation is set earlier, an effect that is well known as the time–cost trade-off of innovation (Scherer and Ross, 1990). An example of this is the Xerox 1045 copier, which was developed in the early 1980s on an accelerated schedule, and which should have been introduced to the market even more quickly in response to competition from Japanese firms. While the copier was already in pilot production, an important design problem was detected: a wire harness, holding about 40 wires connecting internal components, failed to meet the quality standards. Although the harness was swiftly redesigned, it could not then be built in using existing automated manufacturing equipment, which eventually had to be replaced at a cost of more than $1 million (Graves, 1989). This is an example of a typical trade-off faced in industrial R&D projects. Products have to be introduced quickly to be competitive, but this acceleration of development can increase costs.

In addition, another type of cost is involved if innovations are brought onto the market at an earlier date: the risk increases that there are more undetected failures or product defects, which may harm buyers. This cost category increases strongly when the innovation process is shortened, the reason being that it takes time to detect hidden failures of innovations. One may speak of an "information loss of acceleration" (Graves, 1989, p. 2). For example, adverse reactions to some drugs can be seen only after some time has passed. Another source of an increasing failure rate can be a f rm's managers, who may be tempted to deliberately ignore potential failures of innovations, in order to keep an early date for the introduction of an innovation, if there is f erce competition. Tort liability will, of course, ensure that the victims of an innovation introduced too early will be compensated, and it deters wrongdoing by forcing the potential wrongdoer to internalize the costs of harming others. However, the deterrence may be too low to ensure an optimal amount of innovative activity, if there is also competitive pressure to introduce the innovation.

In the following, we begin by outlining the model, in which we draw links between innovation, tort law, and competition (section 2). In the next step we analyze the f rms' behavior in our setting and compare it with the social optimum (section 3). This comparison reveals that, depending on the competitive pressure, a social planner is well advised to complement tort liability with punitive damages, in order to equilibrate innovative activity at an optimal level. The fourth section critically discusses the f ndings and proposes suggestions for future research. Finally, in the f fth section the policy implications of our model are discussed.

2 The Model

For the purpose of our investigation, we consider n firm competing for a market in which the representative consumer may buy one unit of the good. The consumer's valuation of the good is equal to v and exceeds the possible harm d that might occur when consuming the good.[3] Accordingly, demand in the market is equal to one for a price that is equal to or smaller than the difference between the consumer's evaluation v and the expected harm to be borne by the consumer (def ned below), and zero otherwise. The n firm have to invest in a research process associated with uncertainty about the time at which the product can be introduced into the market. Each fir i, $i = 1, ..., n$, chooses the hazard or success rate h_i at which it might succeed with its innovation, where the fir incurs costs $c(h_i) > 0$, in each time period until innovation occurs. Costs are strictly convex in the success rate h_i ($c'(h_i), c''(h_i) > 0$), and the quantity $1/h_i$ represents the expected time until the innovation can be realized. To ensure that there is a time–cost trade-off, we assume that the costs $c(h_i)$ are characterized by an elasticity with respect to the success rate h_i that is greater than one, i.e., that the average expected costs of an innovation,

[3] Huber (1985) claims that the actual praxis of tort law hinders the process of socially benef cial innovations, especially because, in his opinion, it is not accurately taken into account that new products are often safer than the incumbent products. However, in our setup any reduction in uncompensated risks (risks induced by nature) that would otherwise occur will be ref ected in the consumers valuation of the product, v.

$c(h_i)/h_i$, increase in the hazard rate:

$$\frac{d(c(h_i)/h_i)}{dh_i} = \frac{c'(h_i)h_i - c(h_i)}{h_i^2} = \frac{c(h_i)}{h_i^2}\left[\frac{c'(h_i)h_i}{c(h_i)} - 1\right] > 0\,.$$

The innovator retains the exclusive right to make use of the innovation; the f rm may, for example, obtain a patent. This f rm can sell the one unit of the innovative good, and we abstract from further production costs. The f rm charges a monopoly price, which relates to the perceived safety of the product. We assume that the probability of harm arising from the new product is inversely related to the expected time of its being marketed, or, equivalently, is positively related to the success rate. Following among others Daughety and Reinganum (1995), we assume that consumers cannot observe the chosen success rate directly, because fir behavior is private information. But they will form rational expectations about the safety of the product in equilibrium, since they comprehend the f rms' strategy of prof t maximization.[4] The success rate expected by consumers is denoted by $\bar{h}$. The actual probability of harm depends on the actually chosen level for the success rate h_i and is denoted by $x(h_i)$, where $x(h_i)$, $x'(h_i)$, $x''(h_i) > 0$. Given rational expectations, the success rate actually chosen will coincide with the expected success rate in equilibrium, but most importantly, any deviation in the actual choice of the success rate cannot inf uence expectations ex ante.

Given the informational structure described above, the prof t-maximizing strategy of the monopolist who has succeeded with the innovation is to charge a price that equals the consumer's maximal willingness to pay, which is given by

$$p = v - (1-\gamma)x(\bar{h})d\,, \tag{1}$$

where γ denotes the damages factor applied to the firm such that γd equals the f rm's liability in the event of an accident and $(1-\gamma)d$ is the share of harm not compensated to the consumer in the event of an accident.[5] Therefore, the willingness to pay must naturally depend on the expectations consumers have about the accident risk ($x(\bar{h})$). All actors within the economy discount future payments using the interest rate r and are assumed to be characterized by risk-neutral behavior.

3 *The Impact of Competition: An Equilibrium Analysis*

The description of the model so far allows us to determine the prof t equations for f rms and to solve for their optimally chosen success rate. After doing this, we will formulate the maximization problem faced by a social planner. As is standard in the law-and-economics literature, we assume that the social planner aims at maximizing a utilitarian social welfare function and chooses the socially optimal

[4] This is the standard approach in models pertaining to moral-hazard situations; see, e.g., the chapter on the liability of firm in Shavell (1987).

[5] The liability rule considered is the one of strict liability. Indeed, in many jurisdictions product liability often holds firm strictly liable for accidents resulting from defective products (see, e.g., Shavell, 2004; Cooter and Ulen, 2008; Geistfeld, 2009).

success rates. Policy implications regarding optimal damages can then be derived from the perspective of aligning privately and socially optimal choices.

The expected profits of a firm i after being the first to launch an innovation are given by the price the firm can charge as a monopolist (see equation (1)) minus the expected damages,

$$\pi^I(h_i) = p - \gamma x(h_i)d = v - (1-\gamma)x(\bar{h})d - \gamma x(h_i)d\,, \tag{2}$$

where it has to be noted that the consumer's willingness to pay depends on the ex ante expected success rate and that only the expected damages vary with the actual choice of the success rate made by firm i. In the process of innovating, costs $c(h_i)$ are incurred as long as no innovation takes place. The ex ante expected present value of profits for firm i can accordingly be written as

$$\begin{aligned} E\pi_i = \int_0^\infty e^{-rt}\Big[&e^{-\sum_{j=1}^n h_j t}(-c(h_i)) + e^{-\sum_{j\neq i} h_j t} h_i e^{-h_i t}\pi^I(h_i) \\ &+ \Big(1 - e^{-\sum_{j=1}^n h_j t}\Big)0\Big]dt\,, \end{aligned} \tag{3}$$

where r denotes the continuous interest rate. If no firm has succeeded with an innovation by time t, firm i incurs costs $c(h_i)$; given that the other firms have not succeeded in developing an innovation, the density function for an innovation released by firm i at time t is $h_i e^{-h_i t}$. If a firm has already generated the innovation, then the market is already served, implying zero future profits. Straightforward manipulation of equation (3) yields

$$E\pi_i = \int_0^\infty e^{-\left(r+\sum_{j=1}^n h_j\right)t}\big[h_i\pi^I(h_i) - c(h_i)\big]dt = \frac{h_i\pi^I(h_i) - c(h_i)}{r + \sum_{j=1}^n h_j}\,. \tag{4}$$

The expected present value of profits is given by the expected profits per period during the innovation contest, adjusted according to the appropriate discount rate. We assume that expected profits are positive in equilibrium.

Each firm maximizes ex ante expected profits by choosing its success rate h_i. This yields the first-order condition

$$\frac{\partial E\pi_i}{\partial h_i} = \frac{\Big[\pi^I(h_i) + h_i\frac{\partial \pi^I(h_i)}{\partial h_i} - c'(h_i)\Big]\Big(r + \sum_{j=1}^n h_j\Big) - h_i\pi^I(h_i) + c(h_i)}{\Big(r + \sum_{j=1}^n h_j\Big)^2} = 0\,,$$

which can be transformed into

$$\begin{aligned} A := \pi^I(h_i)\Big(r + \sum\nolimits_{j\neq i} h_j\Big) - h_i\gamma x'(h_i)d\Big(r + \sum\nolimits_{j=1}^n h_j\Big) \\ - c(h_i)\left[\frac{c'(h_i)}{c(h_i)}\Big(r + \sum\nolimits_{j=1}^n h_j\Big) - 1\right] = 0\,. \end{aligned} \tag{5}$$

The first-order conditions state that the profit from an earlier innovation, the first term in (5), equals the increase in costs due to higher expected damages and higher costs of developing innovations. The second-order condition for a maximum of ex ante expected profits, which requires $\partial A/\partial h_i < 0$, is fulfilled:

$$\frac{\partial A}{\partial h_i} = \big(-2\gamma x'(h_i)d - h_i\gamma x''(h_i)d - c''(h_i)\big)\Big(r + \sum\nolimits_{j=1}^n h_j\Big) < 0\,.$$

From the slope of the reaction curve $h_i(h_k)$, the conclusion can be drawn that the success rates are strategic complements; i.e., higher rates of innovation on the part of competitors induce f rms to increase their own innovation effort:

$$\frac{\partial h_i}{\partial h_k} = -\frac{\partial A/\partial h_k}{\partial A/\partial h_i} = \frac{h_i \pi^I(h_i) - c(h_i)}{-\big(r + \sum_{j=1}^n h_j\big)\partial A/\partial h_i} > 0\,,$$

where $k \neq i$, and both the f rst-order condition (5) and the assumption of positive expected prof ts in equilibrium have been applied.

Now, the market equilibrium (6) can be determined. With symmetric f rms and rational expectations formed by consumers, $h = \bar{h} = h_i$ applies in equilibrium, which yields

$$\begin{aligned} A^E &:= (v - x(h)d)(r + (n-1)h) - h\gamma x'(h)d(r + nh) \\ &\quad - c(h)\left[\frac{c'(h)}{c(h)}(r + nh) - 1\right] = 0\,. \end{aligned} \tag{6}$$

Stability of the equilibrium requires that $\partial A^E/\partial h < 0$, which we assume to be fulf lled.[6] Given this additional condition, it is easy to show that the equilibrium success rate h increases in the number of f rms taking part in the innovation race and decreases in the damages firm are liable for, i.e., the damages factor γ. More intense competition forces f rms to increase innovation costs and to reduce the expected time to market, because of the threat that another f rm might serve the market earlier. On the other hand, tougher liability rules dampen a f rm's incentive to launch innovations early, making it more advantageous to invest a greater amount of time in research and thereby reduce product risks.

In the next step, we will address the question of what level of research a social planner maximizing a utilitarian welfare function would choose. In other words, what is the socially optimal amount of time that a f rm should spend on research? In order to determine this optimum level of research, the social planner maximizes the expected present value of the sum of producer and consumer welfare. Due to the simplifying assumptions with respect to the market structure, the consumer surplus is equal to zero in equilibrium. Thus, it is suff cient to look at the f rms' expected present value of prof ts. Expected social welfare is then given by

$$\begin{aligned} ESW &= \int_0^\infty e^{-rt}\left[e^{-\sum_{j=1}^n h_j t}\Big(\sum\nolimits_{k=1}^n [h_k(v - x(h_k)d) - c(h_k)]\Big)\right]dt \\ &= \frac{\sum_{k=1}^n [h_k(v - x(h_k)d) - c(h_k)]}{r + \sum_{j=1}^n h_j}\,. \end{aligned}$$

Acknowledging the symmetric structure of the equilibrium, we can restate this as

$$ESW = \frac{nh(v - x(h)d) - nc(h)}{r + nh}\,.$$

[6] Indeed, in the second-best equilibrium, in which the policymaker chooses the optimal damages factor γ^*, the condition is fulfill d with certainty.

The first-best success rate h^* results from the first-order condition

$$\frac{\partial ESW}{\partial h} = \frac{n(v - x(h^*)d)r - nh^*x'(h^*)d(r + nh^*) - nc(h^*)\left[\frac{c'(h^*)}{c(h^*)}(r + nh^*) - n\right]}{(r + nh^*)^2} = 0\,,$$

implying that

$$(7)\quad B := (v - x(h^*)d)r - h^*x'(h^*)d(r + nh^*) - c(h^*)\left[\frac{c'(h^*)}{c(h^*)}(r + nh^*) - n\right] = 0\,.$$

A comparison between the market equilibrium, equation (6), and equation (7) reveals that the social planner values the monopoly profits $(v - x(h)d)$ to a lesser extent, since he/she notices that innovation on the part of one firm implies that no other firms will participate in the monopoly rent. In contrast, the social planner sees a benefit in the saving of total costs $nc(h)$ after an innovation has been introduced, in contrast to a single firm, which takes account only of its own cost savings.

Furthermore, the change in the accident probability is taken into account with a factor of one by the social planner, whereas firm apply the damages factor γ. With the second-order condition $\partial B/\partial h < 0$ for a welfare maximum being fulfilled, equation (7) indicates that the optimal success rate h^* is a decreasing function of the number n of market participants: $\partial h^*/\partial n < 0$. The intuition for this result is that a growing number of firms leads to increasing negative externalities between the firms' research efforts and the sum of the expected present value of profits.

Yet the question is how the policymaker can urge firms to adjust innovation efforts to the socially optimal level. For this goal the policymaker has the damages factor γ at his/her disposal, whereby γd equals the liability imposed on the firms. With respect to the optimal damages factor our main findings are summarized by:

PROPOSITION 1 *The optimal damages factor that equates the privately optimal research intensity with the socially optimal success rate exceeds one and requires damages exceeding harm as long as there is competition between firms Without competition in the innovative sector (n* $= 1$*), the optimal damages factor is equal to one and requires full compensation of accident victims.*

PROOF We solve for the optimal damages factor by equating the expressions in equations (6) and (7), while stipulating that $h = h^*$. This yields

$$(8)\qquad \gamma^* - 1 = \frac{n-1}{h^*x'(h^*)d}\,\frac{h^*(v - x(h^*)d) - c(h^*)}{r + nh^*} = \frac{n-1}{\varepsilon_{x,h}(h^*)}\,\frac{E\pi^*}{x(h^*)d} \geq 0\,,$$

where $\varepsilon_{x,h}(h^*) = h^*x'(h^*)/x(h^*)$. The term on the right-hand side is equal to zero for $n = 1$ and larger than zero for $n > 1$. *Q.E.D.*

PROPOSITION 2 *All else equal, the optimal damages multiplier depends (i) positively on the number of firms in the market, (ii) negatively on the elasticity of the accident probability with respect to the success rate in equilibrium, and (iii) positively on an innovating firm's value in relation to harm in equilibrium.*

PROOF Follows from equation (8).

As long as there is competition between f rms ($n > 1$), the optimal damages factor exceeds one and increases in expected profits This means that it is not enough that firm compensate actual harm in order to obtain the optimal level of innovative activity. There have to be extra f nes, exceeding the value of actual harm. This might be thought equivalent to the requirement of *punitive damages*. Only if competition is absent ($n = 1$) do we f nd an optimal damages factor equal to one and no punitive damages required.

The intuition for this result can be summarized as follows. A damages factor equal to one makes the f rm internalize all the effects of its decision regarding the innovation success rate on the value of the product as perceived by the consumers. This result conf rms the f ndings in the standard models of liability in market settings (see, for example, Shavell, 2004). Note that in our setup no consumer surplus exists, which implies that a monopoly is not necessarily welfare-reducing; hence, we obtain the f rst-best outcome for a monopoly and full liability. However, as outlined above, with competition, f rms' incentives with respect to the innovation hazard rate deviate from the social optimum. Firms do not take into account the lower expected profit of competitors when accelerating their research efforts, while, on the contrary a greater number of f rms in the market even increases the incentives for launching innovations early. However, by setting the damages factor above one, the policymaker tackles the exaggerated incentives for marketing innovations too early by punishing f rms whose products produce harm to consumers. Furthermore, a higher elasticity of the accident probability with respect to the success rate implies that higher damages will incentivize firm to limit their chosen success rate to a larger extent. Accordingly, the necessary deviation from the principle of full compensation is less pronounced in this case. Finally, the higher are the expected prof ts in comparison with expected harm, the more heavily the negative externality between f rms bears on firm ' incentives to increase their own success rate. This implies that a higher damages factor is necessary to align private prof t maximization with socially optimal choices.

So far, the total effect of an increase in the number of competitors on the optimal damages factor has not been established, since more intense competition (larger n) is associated with a lower equilibrium value of f rms ($E\pi^*$). Whereas the direct effect of an increase in the number of f rms calls for a higher damages factor, this is mitigated by the lower f rm value. To end the description of the model, we show that the damages factor is indeed a monotonic function of the number of competitors, i.e., the degree of competition in the market.

PROPOSITION 3 *The optimal damages factor increases monotonically in the number of f rms taking part in the innovation race.*

PROOF Differentiating equation (8) with respect to the number of f rms and noticing that the optimal success rate depends on the number of f rms, we obtain

$$\frac{d\gamma^*}{dn} = \frac{h^*(v - x(h^*)d) - c(h^*)}{(r + nh^*)^2} \frac{(r + h^*)}{h^* x'(h^*)d}$$
$$- \frac{h^*(v - x(h^*)d) - c(h^*)}{r + nh^*} \frac{(n-1)[x'(h^*)d + h^* x''(h^*)d]}{(h^* x'(h^*)d)^2} \frac{\partial h^*}{\partial n} > 0 .$$

Q.E.D.

The optimal damages factor increases due to two effects. First, it increases as the number of f rms increases, since the externalities that are due to competition become more prevalent. Second, it increases because the f rst-best innovation success rate decreases in the number of f rms, whereas private incentives for early innovation increase in the number of f rms. This means that f ercer competition makes firm more eager to launch innovations (too) early. This has then to be counteracted by making even more intensive use of inf ated damages.

4 Discussion

The general aim of this contribution is to provide an extended framework for the analysis of optimal tort law regimes. In our setup, we combine an innovation contest (with a variable number of f rms) with the idea of product risks and the subsequent liability of tortfeasors. This setup allows us to highlight the inf uence of the f erceness of competition on incentives for innovation and its subsequent results for product safety. We are also able to deduce the implications for the optimal damages factor in tort law. As a result of this, we can specify which factors should be taken into account when the responsible body determines the damages award. It emerges that the main factors in this calculation should be: (1) the degree of competition in the market, i.e., the number of f rms in the relevant sector; (2) the extent to which a lengthened development period reduces product risks, since this determines how tort law can affect a f rm's decision to launch an innovation; and (3) expected prof ts, which have been shown to indeed explain the level of punitive damages applied, at least to some extent (see, for example, Karpoff and Lott, 1999). Given these features, the model is rich in its predictive content, especially given its rather simple structure.

The simple structure naturally comes at a price. First, in our model we concentrate on a f rm's liability as the sole means for the policymaker to influenc research decisions. When considering innovations, another approach taken instead of the ex post instrument of liability may be ex ante regulation (for a comparison of different approaches see, for example, Shavell, 1987, chapter 12). This method of government intervention is prominent in the pharmaceutical industry, for instance, where the selling of a new drug most often necessitates approval by a state agency like the Food and Drug Administration in the U.S. In allowing for such an approach of government intervention in addition to f rms' liability, the relevance of a high damages factor applied to f rms is likely to be the less pronounced the more effective direct regulation is. However, given information asymmetry between the regulator and f rms, the sole use of the instrument of requiring off cial product approval might not result in a f rst-best allocation. We have extended our model to allow for imperfect ex ante regulation in this respect, in order to illustrate this line of argument. The formal calculations can be retraced in the Appendix. Indeed, high damages become less important if regulation is more effective. However, firm ' liability may still be part of the optimal policy package. In addition, relying on firms liability may also be associated with lower enforcement costs than direct regulation: the latter may require frequent state involvement, while the former only requires intervention in the event of an accident (see Shavell, 2012, who explicates this reasoning for a comparison of the negligence rule and regulation). This would increase the desirability of relying on the tort system.

Second, we neglect the fact that granting a monopoly in product markets may itself entail ineff ciencies. Given downward-sloping demand curves, firm will restrict output and charge higher prices. This may also affect decisions regarding product safety, as lower output may lead to reduced incentives to invest in avoiding harmful accidents (Baumann and Friehe, 2012). Third, a topic broadly discussed in the economic literature is that of positive research externalities associated with innovations, which may be the reason for alleged insuff cient innovative activity in the market. By contributing to the stock of knowledge in an economy, a generation of innovations may result in spillover effects by, for example, reducing the costs of further technological progress. Abstracting from the arguments brought forth here, whether or not incentives for innovation in markets with imperfect competition are too low or too high cannot be stated a priori (see, for example, Aghion and Howitt, 1992), and empirical studies are skeptical about the importance of such spillover effects (see, for example, Bottazzi and Peri, 2003). However, should such spillover effects be present, this may argue for lower damages, in line with the theory of second best (Lipsey and Lancaster, 1956).

A further issue concerns the difference between patenting and commercialization. Many patents are never commercialized, and for others it takes many years before a product is brought to market. Although this does not contradict our f ndings in general, it leads to a more nuanced discussion of the subject. While one cannot rule out the simple possibility that some patent holders do not commercialize a patent because of safety concerns, the more specifi question is, whether there are differences between industries. The *perceived* competitive pressure between industries may be different, as well as the propensity to commercialize a patent. Firms may regard themselves as belonging to a certain strategic group of f rms that shares a specif c understanding to take competitive action (Zucchini and Kretschmer, 2011). Quick commercialization of patents can be part of this shared understanding of how to compete within a group of f rms. As a result, in reality, our model applies the better the more the commercialization of patents is perceived as a crucial element of competition within a specif c industry. However, these more industry-specif c arguments need more empirical underpinning.

Finally, the hazard-rate model employed here (Kamien and Schwartz, 1972) might be replaced by a more comprehensive one, as for example proposed by Kamien and Schwartz (1980), in order to embrace the innovation process of f rms more thoroughly. Also, a more refine game-theoretical treatment might provide additional insights into the interaction of competing f rms facing the possibility of being sued in a tort case. The inclusion of more refi ed models would allow a firm s innovative behavior under certain circumstances to be investigated in more detail. However, we think that the basic relation between competition, innovation, and tort law, as outlined in this article, remains intact.

5 Policy Implications

With regard to the design of tort law, the policy implications of our f ndings are straightforward. With the exception of the case in which competition is absent, optimal investment behavior on the part of innovative f rms is induced only if there are damages awards exceeding actual harm, which might be likened to the application of punitive damages.

Our f ndings suggest that countries that do not yet have punitive damages at their legal disposal should consider that legal instrument. This holds in general for most European countries, where punitive damages are regarded critically. In most European countries it is argued that punitive damages are in conflic with the *ordre public*, which means that a clear distinction has to be made between the aim of tort law of compensating victims and the purpose of punishing a perpetrator of harm in a criminal court (see, for example, Licari, 2011). In a criminal court the procedural yardstick for determining liability is much higher than in a private law court (Cooter and Ulen, 2008, p. 489); also, in a criminal court the punitive award will not usually be awarded to the victim but to the public. However, from a consequentialist point of view, these legal arguments against punitive damages do not make so much sense when a legal system is aimed at providing incentives for agents to behave effic ently in the sense that society minimizes social costs.

Yet tort law would have to undergo a reform in the U.S. as well. The point is that today punitive damages are awarded more or less arbitrarily (see, for example, Viscusi, 1998), although in recent years in some U.S. states caps on punitive damages have been introduced or requirements have been established that more evidence be provided before punitive damages are awarded (Rubin and Shepherd, 2007). However, our finding suggest that in the case of innovations, punitive damages ought to be employed on a regular basis in a tort case along the parameters outlined above. As part of this, the option might be considered of awarding the punitive damages to the public instead of to the victim and the related lawyers, in order to prevent perverse incentives to sue.

However, we are not pleading for unrestricted use of punitive damages. First, our analysis is concerned with punitive damages related to innovative activity and competition; in the case that there is no innovative activity on the part of f rms, other arguments pro and contra punitive damages may play a role. For example, Daughety and Reinganum (1997) provide an analysis in which competition and punitive damages are both means of setting incentives for firm to reveal the true quality of a (noninnovative) product. Second, our f ndings suggest that the optimal amount of punitive damages depends on the intensity of competition within an industry to bring forth innovation. That is, the level of competitive pressure plays a decisive role in determining the optimal amount of punitive damages to be awarded. As a consequence, the level of punitive damages cannot be decided subjectively by a jury or a judge, but must be derived as a result of equilibrating different economic forces. Thus, our argument is very much in line with Viscusi's plea for a more economics-oriented tort law system (Viscusi, 1998), even though we do not share his opinion that punitive damages are superfluous

Another, more indirect policy implication is that punitive damages may avoid, in general, innovative products being introduced too early due to a patent race. Thus, while punitive damages cannot avoid the duplication of innovation costs in a patent race, they make the time–cost trade-off of innovative activity less severe by setting an incentive for f rms to slow down the pace of innovation. Society as a whole benefi s from this effect, since fewer resources are employed in order to accelerate innovation processes. In other words, the *common-pool* problem of patent races becomes mitigated by punitive damages. The positive effect on social welfare resulting from the mitigation of the common-pool problem will, of course, differ between countries, depending on the size of a country's R&D sector. However, the welfare gain may be considerable and may be understood as a second social dividend besides the social gain from reducing accidents.

Even though the model applied here is rather simple, it captures the basic features of the interaction between competition, innovation, and the tort law regime. Moreover, the model gives rise to the policy conclusion that infla ed damages awards can be an important means of calibrating the pace of innovation to an optimal level. At the very least, the conclusion may be drawn that without punitive damages the innovation process becomes less effic ent. However, we are aware that the simplicity and clarity of the model comes at a price. While the model may be suitable for deriving a general statement in favor of punitive damages, in a specifi tort case more factors may have to be taken into account before one can decide on the effic ency of punitive damages. Also, questions remain on the implementation level. In our model the awarded punitive damages depend on the level of competition. This implies that the degree of competition has to be measured and expressed as a number. Even though measuring the degree of competition in a market is not an easy task, one can refer to a growing literature that deals with that problem (see, for example, Aghion et al., 2005). Thus the model proposed here constitutes rather the beginning of a discussion than the end.

Appendix: Extending the Model by the Requirement of Offi ial Product Approval

In order to scrutinize the use of damages in the case that the policymaker has access to the instrument of directly regulating a f rm's market entry (e.g., a government agency has to approve the product before it can be sold in the market), we extend the model described in section 3 in the following way. To take into account that direct regulation is likely to be imperfect as well, we assume that the probability that a f rm i's product f nds approval by the agency is a function of the success rate h_i, where the probability is denoted by $q(h_i)$ and $q'(h_i) < 0 < q''(h_i)$ holds. That is, the regulator may have the idea that a certain success rate should be chosen. However, due to asymmetric information between the better-informed fir and the regulator, product approval might be granted also for higher values of the success rate or even be denied for lower values. Accordingly, we assume that the probability of the product being approved is an increasing function of the expected time until an innovation is made ($1/h_i$), as the product becomes more likely to meet given standards. Taking this additional aspect into account, we can state the expected prof ts of a f rm as

$$E\pi_i = \int_0^\infty e^{-\left(r+\sum_{j=1}^n q(h_j)h_j\right)t}\left[q(h_i)h_i\pi^I(h_i) - c(h_i)\right]dt = \frac{q(h_i)h_i\pi^I(h_i) - c(h_i)}{r + \sum_{j=1}^n h_j},$$

which might be compared to equation (4), while prof ts after successful approval of the product are still given by equation (2). After determining the f rst-order condition for a profi maximum and using the fact of a symmetric equilibrium, we can restate the condition determining the chosen success rate in the market equilibrium as

$$\begin{aligned} A^E := {} & q(h)\left[1 + \varepsilon_{qh}(h)\right](v - x(h)d)(r + (n-1)h) - h\gamma x'(h)d(r + nh) \\ & - c(h)\left[\frac{c'(h)}{c(h)}(r + nh) - q(h)[1 + \varepsilon_{qh}(h)]\right] = 0, \end{aligned} \tag{A1}$$

where $\varepsilon_{qh}(h) = q'(h)h/q(h)$ is the elasticity of the approval probability with respect to the chosen success rate. This elasticity is negative and should be absolutely higher for a more effective regulation system, that is, the easier it is for the regulation agency to determine the appropriateness of f rms' research efforts. Given the probability of approval $q(h)$, the expected social welfare reads

$$ESW = \frac{nq(h)h(v - x(h)d) - nc(h)}{r + nh},$$

and maximization of expected social welfare results in the optimal success rate being described by

$$\begin{aligned} B := {} & q(h^*)[1 + \varepsilon_{qh}(h^*)](v - x(h^*)d)r - h^*x'(h^*)d(r + nh^*) \\ & - c(h^*)\left[\frac{c'(h^*)}{c(h^*)}(r + nh^*) - nq(h^*)\left[1 + \varepsilon_{qh}(h^*)\right]\right] = 0. \end{aligned} \tag{A2}$$

Comparing equations (A1) and (A2) yields the result that the optimal damages factor γ^* can be found from

$$\begin{aligned} \gamma^* - 1 &= [1 + \varepsilon_{qh}(h^*)]\frac{n-1}{h^*x'(h^*)d}\frac{h^*(v - x(h^*)d) - c(h^*)}{r + nh^*} \\ &= [1 + \varepsilon_{qh}(h^*)]\frac{n-1}{\varepsilon_{x,h}(h^*)}\frac{E\pi^*}{x(h^*)d} \geq 0, \end{aligned}$$

which agrees with the corresponding condition in the main model, equation (8), except for the additional factor $[1 + \varepsilon_{qh}(h^*)]$. To summarize, in addition to the conditions outlined in the main text, the optimal damages factor now depends on the effectiveness of the ex ante regulation of product approval. Given this additional instrument, the optimal damages factor is lower than in the main analysis, which reflect the substitutability between the different means: regulation and firms liability. Punitive damages remain part of the optimal liability system as long as the approach of regulation is suff ciently ineffective, which translates into $\varepsilon_{qh}(h^*) > -1$. For a more effective system of regulation, optimal damages are lower. This might be likened to a change in the liability system in the direction of a negligence rule. For example, in the realm of product liability, liability is often incurred if the f rm's product is judged as defective, which incorporates an element of negligence into the otherwise applicable liability rule of strict liability.

References

Aghion, P., N. Bloom, R. Blundell, R. Griffith and P. Howitt (2005), "Competition and Innovation: An Inverted-U Relationship," *The Quarterly Journal of Economics*, 120(2), 701–728.

— and P. Howitt (1992), "A Model of Growth through Creative Destruction," *Econometrica*, 60(2), 323–351.

Baumann, F., and T. Friehe (2012), "Optimal Damages Multipliers in Oligopolistic Markets," Working Paper Series 2012-08, Department of Economics, University of Konstanz.

Bottazzi, L., and G. Peri (2003), "Innovation and Spillovers in Regions: Evidence from European Patent Data," *European Economic Review*, 47(4), 687–710.

Cooter, R., and T. Ulen (2008), *Law & Economics*, 5th edn., Prentice Hall, Upper Saddle River (NJ).

Daniels, S., and J. Martin (1990), "Myth and Reality in Punitive Damages," *Minnesota Law Review*, 75(1), 1–64.

Dari-Mattiacci, G., and L. A. Franzoni (2013), "Innovative Negligence Rules," Amsterdam Law School Legal Studies Research Paper No. 2013-14, University of Amsterdam.

Daughety, A. F., and J. F. Reinganum (1995), "Product Safety: Liability, R&D, and Signaling," *The American Economic Review*, 85(5), 1187–1206.

— and — (1997), "Everybody Out of the Pool: Products Liability, Punitive Damages, and Competition," *The Journal of Law, Economics, & Organization*, 13(2), 410–432.

Epstein, R. A. (2006), "The Pharmaceutical Industry at Risk: How Excessive Government Regulation Stifle Innovation," *Clinical Pharmacology & Therapeutics*, 82(2), 131–132.

Geistfeld, M. A. (2009), "Products Liability," in: M. Faure (ed.), *The Encyclopedia of Law and Economics, 2nd edn., Vol. 1: Tort Law and Economics*, Edward Elgar, Cheltenham, pp. 287–340.

Graves, S. B. (1989), "The Time–Cost Tradeoff in Research and Development: A Review," *Engineering Costs and Production Economics*, 16(1), 1–9.

Huber, P. (1985), "Safety and the Second Best: The Hazards of Public Risk Management in the Courts," *Columbia Law Review*, 85(2), 277–337.

Kamien, M. I., and N. L. Schwartz (1972), "Timing of Innovations under Rivalry," *Econometrica*, 40(1), 43–60.

— and — (1980), "A Generalized Hazard Rate," *Economics Letters*, 5(3), 245–249.

Karpoff, J. M., and J. R. Lott, Jr. (1999), "On the Determinants and Importance of Punitive Damage Awards," *The Journal of Law & Economics*, 42(S1), 527–573.

Leggett, C. (1999), “The Ford Pinto Case,” http://users.wfu.edu/palmitar/Law&Valuation/Papers/1999/Leggett-pinto.html, last access June 17, 2013.

Licari, F.-X. (2011), “La compatibilité de principe des punitive damages avec l’ordre public international: une décision en trompe-l’oeil de la Cour de cassation?” *Recueil Dalloz*, 187(6), 423–427.

Lipsey, R. G., and K. Lancaster (1956), “The General Theory of Second Best,” *The Review of Economic Studies*, 24(1), 11–32.

Loury, G. C. (1979), “Market Structure and Innovation,” *The Quarterly Journal of Economics*, 93(3), 395–410.

Mahoney, R. J., and S. E. Littlejohn (1989), “Innovation on Trial: Punitive Damages versus New Products,” *Science*, 246(4936), 1395–1399.

Parchomovsky, G., and A. Stein (2008), “Torts and Innovation,” *Michigan Law Review*, 107(2), 285–315.

Rubin, P. H., and J. M. Shepherd (2007), “Tort Reform and Accidental Deaths,” *The Journal of Law & Economics*, 50(2), 221–238.

Rustad, M. (1992), “In Defense of Punitive Damages in Products Liability: Testing Tort Anecdotes with Empirical Data,” *Iowa Law Review*, 78(1), 1–88.

Scherer, F. M., and D. Ross (1990), *Industrial Market Structure and Economic Performance*, 3rd edn. Houghton Miff in Co., Boston (MA).

Shavell, S. (1980), “Strict Liability versus Negligence,” *The Journal of Legal Studies*, 9(1), 1–25.

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

— (1992), “Liability and the Incentive to Obtain Information about Risk,” *The Journal of Legal Studies*, 21(2), 259–270.

— (2004), *Foundations of Economic Analysis of Law*, The Belknap Press of Harvard University Press, Boston (MA).

— (2012), “A Fundamental Enforcement Cost Advantage of the Negligence Rule over Regulation,” NBER Working Paper 18418, Cambridge (MA).

— and T. van Ypersele (2001), “Rewards versus Intellectual Property Rights,” *The Journal of Law & Economics*, 44(2), 525–548.

Viscusi, W. K. (1998), “The Social Costs of Punitive Damages against Corporations in Environmental and Safety Torts,” *The Georgetown Law Journal*, 87(2), 285–345.

— and M. J. Moore (1993), “Product Liability, Research and Development, and Innovation,” *Journal of Political Economy*, 101(1), 161–184.

Zucchini, L., and T. Kretschmer (2011), “Competitive Pressure: Competitive Dynamics as Reactions to Multiple Rivals,” Discussion Paper 2011-03, Munich School of Management, University of Munich.

Florian Baumann
Düsseldorf Institute for
Competition Economics
Heinrich Heine University Düsseldorf
Universitätsstraße 1
40225 Düsseldorf
Germany
baumann@dice.hhu.de

Klaus Heine
Rotterdam Institute of
Law and Economics
Erasmus School of Law
Erasmus University Rotterdam
Burgemeester Oudlaan 50
3000 DR Rotterdam
The Netherlands
heine@law.eur.nl

The Influenc of Lawyers and Fee Arrangements on Arbitration

by

Nathalie Chappe and Yannick Gabuthy*

Received September 14, 2012; in revised form July 25, 2013; accepted August 8, 2013

This paper proposes a theoretical analysis of fina -offer arbitration in which disputants may be represented by lawyers who can be paid by fla , contingent, or conditional fees. We derive the equilibrium lawyers' efforts to defend their clients and the equilibrium parties' proposals made to the arbitrator, and evaluate each payment mechanism's performance according to its ability to enhance effort and to promote convergence between the disputants' claims. Following these criteria, the contingent payment structure is shown to be the best regime, since it improves the client–lawyer relationship by enhancing the lawyer's incentives to provide effort, without altering the gap between the parties' positions in arbitration. (JEL: J52, K41)

1 Introduction

1.1 Motivation

When parties are embedded in a dispute, they may either turn to standard litigation or resort to an alternative dispute resolution mechanism, such as arbitration, in order to reach an agreement. Arbitration falls outside the judicial process and can be defi ed by the presence of a third party who is empowered to impose a binding settlement when parties fail to agree on one via bilateral negotiations. This mechanism is now widely used in developed countries for most types of disputes, and its popularity may be explained by the increasing caseload of public courts, a preference for confidentialit , and the desire to reach agreements with minimal delay and minimal cost.[1]

* CRESE, University of Franche-Comté, Besançon, France, Yannick Gabuthy being the corresponding author. We wish to thank Eve-Angéline Lambert and the anonymous referees for their useful comments.

[1] For example, arbitration has long been used to settle wage disputes in the U.S. public sector and is included in many contracts as an alternative to litigation (Deck, Farmer, and Zeng, 2007).

Journal of Institutional and Theoretical Economics 169, 720–738 – ISSN 0932-4569
DOI: 10.1628/093245613X13795082136101 – © 2013 Mohr Siebeck

This increasing reliance on arbitration in practice requires us to understand the strategies and incentives that this mechanism generates. As explained below, the theoretical literature on arbitration has focused attention on this issue by analyzing the strategic implications of different arbitration schemes for the disputants' bargaining behavior inside the procedure. The chief limitation of this literature is that it does not address the possibility for parties to be legally represented during the arbitration process, despite frequent use of legal counsel by disputants in practice: according to empirical studies, 70–80% of employers embedded in arbitration were represented by a lawyer, and the utilization rate by unions was 40–50% (Block and Stieber, 1987; Barnacle, 1991; Wagar, 1994). Ashenfelter and Dahl (2012) provide evidence that parties may derive benefi s from expert assistance, but show the occurrence of a prisoners'-dilemma situation: neither side benefi s from hiring a lawyer when both do so, and disputants would be better off if they could agree not to hire lawyers.

From a theoretical perspective, the aim of our paper is precisely to highlight the strategic implications of legal representation in arbitration by analyzing how the presence of lawyers may shape the disputants' bidding behavior and studying their incentives to get legal counsel. We investigate this issue by considering three different payment schemes for the client–attorney relationship. Under *fla fees*, the lawyer gets a f xed amount whatever the arbitration award, while both contingent and conditional fees pay for performance by compensating the lawyer with a higher fee if the case is won. Under *contingent fees*, the lawyer receives a share of the award if her client wins; under *conditional fees*, the lawyer gets a premium, not related to the adjudicated amount, in case of winning.[2] We argue that the payment mechanism in place may affect the lawyer's incentives to provide effort to defend her client and inf uence the arbitrator's opinion, which in turn may affect the parties' bidding behavior during the arbitration hearings.

1.2 Arbitration Literature

The potential role of legal representation in arbitration has been overlooked in the theoretical literature, which essentially evaluates the performance of different arbitration procedures. The evaluation criterion is the extent to which the procedure is able to promote convergence between the disputants' claims (viz., to reduce the gap between the plaintiff's demand and the defendant's offer). The arguments behind the use of this criterion are the following. First, enhancing convergence may lead parties findin themselves close enough to each other to conclude an agreement before the arbitrator imposes his/her decision. The second point refers to the acceptability of the procedure to the conf icting parties: a narrow gap between claims presumably makes it more likely that the arbitration award will be close to what each party considers reasonable, and therefore be more acceptable to both disputants. In other words, improving the convergence properties of the process would reduce the amount of

[2] The use of contingent fees is now widespread in the U.S.; the UK started introducing conditional fees in the nineties, followed by Belgium and the Netherlands.

arbitrariness in arbitration, which is an important argument given the consensual nature of the procedure.[3] Finally, it is argued that a narrow gap minimizes the risk faced by parties as well as any potential bias or error that could be introduced by the arbitrator's preferences. These arguments are widely developed in the literature (see, e.g., Chatterjee, 1981; Armstrong and Hurley, 2002; Zeng, 2003; Armstrong, 2004; Deck and Farmer, 2009).

Following this criterion, final-offer arbitration (FOA) is generally considered as superior to conventional arbitration (CA), as it induces disputants to make concessions and submit closer bids. CA and FOA are the two most prevalent forms of arbitration used in practice, and they perform as follows: CA mimics civil litigation in form, in that the arbitrator listens to the two sides' settlement proposals and is free to impose any award of his/her choice, while FOA requires that each disputant proposes a final offer and the arbitrator must choose one of the two (viz., the one that is closest to his/her own opinion of a fair settlement). The major criticism of CA in the literature is that arbitrators are inclined – or perceived – to compromise between the parties' final positions, which encourages them to exaggerate claims and avoid concessions (Farber, 1981). In view of this, the proponents of FOA argue that this mechanism induces disputants to stake out more reasonable positions, since no compromise is possible and the disputant with an extreme offer is likely to lose (Farber and Katz, 1979; Farber, 1980).[4] However, while FOA improves convergence and outperforms CA, several papers show that the design of the mechanism is not sufficient to harmonize the parties' positions (Chatterjee, 1981; Brams and Merrill III, 1983).

Aside from CA and FOA, researchers have proposed various alternative arbitration methods in a normative perspective – these methods are not used in real-world cases. These mechanisms include tri-offer arbitration, in which a neutral fact-finder makes a recommendation and the arbitrator must choose one of the three offers (Ashenfelter et al., 1992). Brams and Merrill III (1986) introduce combined arbitration, which is a mix of CA and FOA: both sides submit offers, and the arbitrator can pick either offer or a more extreme allocation. This mechanism induces perfect convergence, since it combines the possibility of extreme outcomes from CA with elimination of the middle-ground outcome as in FOA. Zeng (2003) proposes an amended final-offer arbitration mechanism whose logic is similar to that of a second-price auction: each disputant makes an offer, and the final outcome is based upon the arbitrator's preferred outcome and a penalty on the most extreme offer. This mechanism also induces convergence, since there is a unique equilibrium in which both sides make offers equal to the expected value of the arbitrator's preferred outcome.

---

[3] A key difference between the arbitration of disputes and their resolution in a court regards the process of selecting the fact-finder: both parties must agree to call in the arbitrator for the dispute to be settled by him/her.

[4] This result occurs if the parties are uncertain as to the arbitrator's preferred settlement. If both parties know with certainty the arbitrator's opinion, CA and FOA lead to the same outcome: the disputants' offers converge to the arbitrator's opinion.

1.3 Our Contribution

As mentioned above, this literature has ignored the potential influence of legal representation on the disputants' behavior in arbitration. We analyze this issue by building a model of FOA – which is widely used in practice – in which disputants may be represented by lawyers who can be paid by flat, contingent, or conditional fees.[5] Legal assistance is addressed by considering that the attorney's choice of effort to defend the case (e.g., presenting evidence, supporting documents, or counterarguments) may distort the arbitrator's opinion favorably for her client. In this framework, we derive the equilibrium lawyers' efforts and the equilibrium parties' final offers under the different fee contracts, and evaluate each payment mechanism's performance according to its ability to enhance effort and to promote convergence between the disputants' claims.

From this perspective, it is seen that the flat-fee mechanism displays poor incentive properties: the lawyers provide zero effort under this payment scheme, since defending the case is costly and they get a fixed payoff regardless of the arbitrated settlement. The disputants are hence not likely to be legally represented in this situation.

In contrast, it is highlighted that the contingent payment structure is a desirable one, since it improves the client–lawyer relationship by enhancing the lawyer's incentives to provide effort, without altering the gap between the parties' positions in arbitration (in comparison with the benchmark case where lawyers are absent). The range between the parties' proposals is not affected by legal representation, for the following reason: the use of a lawyer under contingent fees by the plaintiff makes him more demanding during arbitration (since his effort shifts the arbitrator's opinion favorably); however, this effect is perfectly balanced by the defendant, who increases his equilibrium offer by the same amount. The gap between the two disputants' bids is then constant. However, from a normative perspective, it is shown that the contingent percentage should be capped for legal representation to occur in equilibrium, because this share, which is paid to the lawyer, increases the arbitration cost for the client and reduces his residual returns from the procedure. This result provides some microeconomic foundations for the existing legislation, since the contingent percentage is usually capped in practice (see section 2.3 below).

Finally, the conditional fee arrangement provides similar effort incentives to lawyers to those under the contingent contract, since it pays for performance by compensating the attorney with a higher fee if the case is won. However, it induces a perverse effect on the parties' bidding behavior, which produces an increase in the range between their offers. Notice, interestingly, that the conditional payment has to be capped to be accepted by the lawyer: increasing the upscale fee improves her incentives to provide effort, which makes her client more demanding during the

---

5 For example, FOA has been used in major-league baseball to resolve wage conflict between players and team owners since 1979 (Burger and Walters, 2005).

arbitration hearings and could therefore reduce her probability of winning and being paid at the end of the process.

Overall, these results may have some policy implications, since several countries, such as Spain, France, Italy, and Portugal, are considering formally allowing contingent or conditional fees (which are currently forbidden).[6] Furthermore, from a research perspective, our insights shed new light on the convergence properties of FOA, which may be affected by the possibility for the parties to be legally represented during the arbitration hearings.

The remainder of the paper is organized as follows. Section 2 describes the FOA process and analyzes the players' equilibrium strategies, and section 3 concludes. Proofs are relegated to the appendix.

2 *The Final-Offer Arbitration Process*

2.1 Framework

Let us consider two risk-neutral parties, denoted A and B, entering arbitration. Disputant A would like the settlement to be as small as possible, while disputant B would like it to be as large as possible. The situation captured by our framework is the determination of an award for damages in a civil lawsuit, such that arbitration is considered as an alternative to a courtroom trial in which disputant A is the defendant and disputant B is the plaintiff. The FOA process may be characterized as the following four-stage game. At date 1, disputants A and B decide whether to hire lawyers, who choose, in turn, whether to accept the case. At date 2, the parties make proposals a and b to the arbitrator. At date 3, the lawyers – if any – choose their investment/effort levels to defend the case and thus to infl ence the arbitrator's decision. Finally, at date 4, the arbitrator selects one of the two parties' offers as the settlement without compromise. Regarding the timing of the game, we assume here that the lawyers' effort decisions are made after the disputants' proposals have been placed. As mentioned by Deck and Farmer (2009), this timing is consistent with the FOA process used in major-league baseball to resolve salary disputes: parties are required to submit bids several weeks prior to the actual hearings. Nevertheless, this approach may seem somewhat puzzling, since the alternative sequence of events is also common in practice: in other applications of FOA, the effort decisions are made f rst, in order to provide supporting documentation at the time bids are placed. In fact, we can show that both sequences lead to the same insights regarding the ranges between the parties' equilibrium offers and their incentives to hire lawyers. Our dynamic structure is thus chosen without loss of generality.

In order to emphasize the inf uence of different fee structures both on the lawyers' efforts and on the disputants' bidding behavior in arbitration, we consider three possible fee arrangements between disputant B and his attorney:

[6] Pactum de quota litis is not allowed by the ethical code of the European association of lawyers.

(i) A *fla* fee contract, which implies that the lawyer gets a f xed payoff f regardless of the arbitrated settlement.[7]

(ii) A *contingent* fee contract, in which the fee is given by a percentage α of the arbitrated amount (i.e., a or b). Under this fee regime, the lawyer gets a share of her client's recovery as a payment for her legal services.

(iii) A *conditional* fee contract, which is described by an upscale fee β, which is not related to the arbitrated amount, if the case is won. The lawyer gets nothing if disputant B loses (i.e., if proposal a is chosen by the arbitrator), and β if he wins (i.e., if proposal b is chosen by the arbitrator).

With this framework in place, we now proceed to solve our dynamic game using the backward induction procedure. We therefore begin by characterizing the arbitrator's expected decision at date 4, conditional on an arbitrary set of choices made at the previous stages.

2.2 Arbitrator's Decision

Following the literature on FOA, we assume that the arbitrator f rst decides what would be a reasonable award, and then selects whichever fi al offer is closest to it. We consider that the arbitrator's idea of a fair settlement, denoted by s, is a random variable drawn from a uniform distribution on $[0, 1]$, which is common knowledge to both disputants and lawyers.[8] The parameter s is exogenously determined, and it is through this parameter that the economic environment and arbitrator's reasoning affect the arbitrated outcome. Indeed, considering that $b > a$ in equilibrium, the arbitrator selects the offer of disputant A (disputant B) if $s \leq (a+b)/2$ (if $s > (a+b)/2$).

The parties' legal representation is addressed by assuming that the lawyers' efforts, denoted e_A and e_B respectively, shift the arbitrator's opinion favorably for their clients. In other words, by employing a lawyer, disputant A (B) can move the arbitrator's preferred award to the left (right) by the positive quantity e_A (e_B), which is endogenously determined. Typically, each party's claim may be interpreted as a story, and we consider that professional advocates may make this story more convincing by presenting evidence, counterarguments, and so on.

---

[7] Disputant A can only use a f at fee structure to remunerate his agent, since contingent/conditional fees are usually forbidden for defense lawyers (Gravelle and Waterson, 1993; Santore and Viard, 2001; Chen and Wang, 2007). Furthermore, for algebraic convenience, it is assumed that $f_A = f_B = f$.

[8] Assuming that the arbitrator's opinion is uniformly distributed is a simplifying modeling restriction that is commonly used in literature on arbitration (Hanany, Kilgour, and Gerchak, 2007; Deck and Farmer, 2007, 2009). However, Brams and Merrill III (1983) and Brams, Kilgour, and Merrill III (1991) show that the parties' bidding behavior is quite sensitive to the specifi distribution assumed over the arbitrator's preferences.

Given the uniform distribution and the potential influence of lawyers, the probability that the arbitrator chooses disputant A's offer is given by

$$\text{(1)} \qquad \operatorname{Prob}\left\{s - \lambda_A e_A + \lambda_B e_B \leq \frac{a+b}{2}\right\} = \frac{a+b}{2} + \lambda_A e_A - \lambda_B e_B =: \phi\,,$$

where λ_i is a dummy variable equal to one if party i hires a lawyer ($i = A, B$).

This way to model the lawyers' influence on the arbitrator's opinion is related to the framework by Deck and Farmer (2009). The authors examine the disputants' bidding and investment strategies in FOA, depending on the sequence and observability of choices. The investment decision is a continuous variable, which may be interpreted as hiring an attorney, spending time developing and presenting a case, and so on. In our paper, we focus the analysis on a particular – binary – investment decision (viz., to hire or not to hire a lawyer) and extend the framework by considering different fee schedules, whereas Deck and Farmer (2009) consider only a linear investment cost function (which captures a flat fee structure). Furthermore, following Ashenfelter and Dahl (2012), we choose here to view lawyers as agents whose aim is to influence the arbitrator. However, some other reasons might explain legal representation. For instance, a lawyer might be considered as an expert with better information about the merits (i.e., the expected return) of the case than her client. The disputant might hence rely upon his lawyer's recommendation on the proposal to make or on whether to pursue the case in arbitration. The important question of the screening role of lawyers has been discussed in the literature on litigation (see, e.g., Dana and Spier, 1993), and could be addressed in an extension of our analysis.

Having explained the arbitrator's decision rule, we now investigate the attorneys' choices of effort at date 3.

2.3 Lawyer's Investments

Each lawyer exerts effort to maximize her expected payoff, taking her opponent's effort as given.

Given that disputant A uses a flat fee contract, his attorney gets the fixed payoff f regardless of the arbitration outcome, which implies the following payoff function:

$$\text{(2)} \qquad \Pi_A^j = f - \frac{e_A^2}{2}\,,$$

where $e_A^2/2$ is lawyer A's quadratic effort cost, and j characterizes the fee contract used by disputant B (i.e., $j = F$ under flat fees, $j = C$ under contingent fees, and $j = K$ under conditional fees).[9]

The payoff of attorney B depends on the fee regime in place:

[9] For simplicity, we assume that both lawyers share the same effort cost function. Furthermore, to alleviate notation, party i's lawyer will be called lawyer i in the rest of the paper ($i = A, B$).

(i) Under a *fla* fee contract, following (2), her payoff function is given by

$$\Pi_B^F = f - \frac{e_B^2}{2}\,. \tag{3}$$

(ii) Under a *contingent* fee contract, lawyer B gets a fraction α of the arbitrated amount. Her expected payoff may hence be written as

$$\Pi_B^C = \alpha[a\phi + b(1-\phi)] - \frac{e_B^2}{2}, \quad \text{where} \quad 0 < \alpha < 1\,. \tag{4}$$

(iii) Under a *conditional* fee contract, she gets the upscale fee β in case of winning, independently of the arbitrated amount. Her expected payoff is then

$$\Pi_B^K = \beta(1-\phi) - \frac{e_B^2}{2}, \quad \text{where} \quad 0 < \beta < b\,. \tag{5}$$

Given the values of α and β, which are exogenously determined by the public authorities, each lawyer chooses her level of effort in order to maximize her payoff, for arbitrary proposals a and b made by the parties. An alternative way to analyze the client–lawyer relationship would be to consider a principal–agent setup where the lawyer's remuneration (i.e., the fi ed fee, the contingency percentage, or the upscale bonus) is determined endogenously by contract. While such a more normative approach would be interesting, the issue of lawyer's control and related agency problems has been widely analyzed in the literature (see, e.g., Miller, 1987; Emons and Garoupa, 2006; Emons, 2007). Furthermore, even if the value of the fees is indeed generally based on a contractual arrangement between the attorney and her client, it is also limited by local rules for "reasonableness" in practice. For example, in most jurisdictions in the U.S., the contingent percentage is capped at 45% and equal to 33% in most cases. Such "norms of conduct" apply also for conditional fees and thus induce some exogeneity into the determination of the attorneys' remunerations.

The following proposition shows how the fee regime affects lawyers' incentives to invest in the case and to distort the arbitrator's opinion.

PROPOSITION 1 *Lawyer A is never induced to provide effort, while the equilibrium effort made by lawyer B depends on the fee regime in place:*

$$e_A^j = 0 \ (\forall j \in \{F, C, K\}); \quad e_B^F = 0; \quad e_B^C = \alpha(b-a); \quad e_B^K = \beta\,, \tag{6}$$

where j stands for the equilibrium value of the lawyers' strategies in each fee regime.

These results may be interpreted as follows. Overall, in each fee regime, the marginal productivity of effort is equated with its marginal cost. In case of f at fees, both attorneys provide zero effort, since investing to defend the case is costly and they get a f xed payment that does not depend on the arbitration outcome. Under the contingent fee contract, lawyer B gets a fraction α of the arbitrated amount, and increasing e_B decreases the probability that a is chosen by the arbitrator (i.e., ϕ) and enhances the likelihood that b is selected (i.e., $1-\phi$). Therefore, given the marginal effects of e_B on these probabilities, as highlighted in (1), it follows immediately that the marginal benefi from an increase in e_B is given by αb, while the marginal cost equals $\alpha a + e_B$, so that $e_B = \alpha(b - a)$ in equilibrium. Under the conditional fee contract, an increase in e_B provides a marginal gain of β, since lawyer B receives this payment in case of winning.

Having characterized the outcome of date 3, we now determine the equilibrium f nal offers chosen by the parties at stage 2.

2.4 Parties' Final Offers

In choosing a fina offer, each disputant trades off the benefi of a larger (or smaller) award against the probability that his offer will be selected by the arbitrator. This trade-off must be affected by legal representation, since the lawyer increases the probability of winning an arbitration case (ceteris paribus), depending on the legal fee structure in place.

Given that disputant A uses a f at fee contract, he has to pay f regardless of the arbitrated settlement, and his expected utility may be written as

(7) $$U_A^j = -[a\phi + b(1 - \phi) + \lambda_A f]\,,$$

where $j \in \{F, C, K\}$.

The expected utility of disputant B depends on the fee contract in place:
(i) Under a *fla* fee contract, disputant B has to pay f in any case. His expected payoff is then

(8) $$U_B^F = a\phi + b(1 - \phi) - \lambda_B f\,.$$

(ii) Under a *contingent* fee contract, disputant B gets the fraction $1 - \alpha$ of the arbitrated amount. His expected utility may hence be written as

(9) $$U_B^C = (1 - \lambda_B \alpha)[a\phi + b(1 - \phi)], \quad \text{where} \quad 0 < \alpha < 1\,.$$

(iii) Under a *conditional* fee contract, the lawyer gets β in case of winning, independently of the arbitrated amount. The disputant B's expected utility is then

(10) $$U_B^K = a\phi + (b - \lambda_B \beta)(1 - \phi), \quad \text{where} \quad 0 < \beta < b\,.$$

Each disputant chooses his settlement offer in order to maximize his expected utility. The following proposition highlights the effect of legal representation and fee structure on the range R between the disputants' equilibrium offers (i.e., $R \equiv b - a$).

PROPOSITION 2 *Legal representation affects the range between equilibrium offers only when disputant B hires a lawyer under a conditional fee contract.*

This result may be interpreted as follows. Legal representation does not affect the parties' behavior under f at fees, since they anticipate that lawyers will not defend the case under this payment mechanism. In this context, the parties' offers and the range between them are equal to those that would arise without legal representation:

(11) $$a^F = 0 \quad \text{and} \quad b^F = 1 \quad \Rightarrow \quad R^F = 1\,.$$

Under contingent fees, we get similar results because the parties integrate perfectly the inf uence of the lawyer B's effort on the arbitrator's decision when formulating their f nal offers:

$$(12)\qquad a^C = 2\alpha \quad \text{and} \quad b^C = 1 + 2\alpha \quad \Rightarrow \quad R^C = 1 \quad (\text{when } \lambda_B = 1)\,.$$

In other words, even if the use of a lawyer by disputant B infl ences his bargaining behavior in arbitration, this effect is perfectly by disputant A, who corrects his fi al offer: given the inf uence of lawyer B on the arbitrator's opinion, her client increases his demand by 2α, and the equilibrium reaction from the opponent is to increase his proposal by the same amount.[10] In contrast, the conditional fee arrangement affects this basic strategy trade-off by introducing a distortion in disputant B's expected payoff: under conditional fees disputant B gives a bonus to his lawyer only in case of winning, while under contingent fees he pays a f xed share of the expected arbitrated amount. In this context, we obtain

$$(13)\qquad a^K = \beta \quad \text{and} \quad b^K = 1 + \frac{3\beta}{2} \quad \Rightarrow \quad R^K = 1 + \frac{\beta}{2} \quad (\text{when } \lambda_B = 1)\,.$$

Furthermore, comparing (11), (12), and (13), we can remark that the range between equilibrium offers is enhanced when the conditional payment mechanism is implemented:

$$R^K = 1 + \frac{\beta}{2} > R^F \equiv R^C \equiv 1 \quad \Leftrightarrow \quad \beta > 0\,.$$

Indeed, in comparison with the f at-fee situation, the conditional payment regime increases the arbitration cost in case of victory, which induces disputant B to increase his demand in order to maximize his residual return from arbitration, while disputant A's behavior is less affected by the conditional fee contract:

$$b^K - b^F = \frac{3\beta}{2} \equiv e_B^K - e_B^F + \frac{\beta}{2} \quad \text{and} \quad a^K - a^F = \beta \equiv e_B^K - e_B^F\,.$$

Given that $e_B^K > e_B^F$, the increase in disputant A's proposal (i.e., $a^K - a^F$) is typically more than offset by the rise of disputant B's claim (i.e., $b^K - b^F$), because disputant B shifts half of the upscale fee β into his claim in order to be reimbursed from this cost in case of winning. The gap between the disputants' equilibrium offers is then enlarged under conditional fees.[11] Similar arguments may be used to compare the disputants' behavior under conditional and contingent fees.

In the next subsection, we emphasize the parties' incentives to hire lawyers and the lawyers' incentives to accept the case, given the strategic implications of legal representation that are highlighted above.

---

[10] As shown in appendix A.2, attorney B's effort evaluated at the equilibrium parties' offers is $e_B^C = \alpha(b^C - a^C) = \alpha$.

[11] Notice that disputant B is not induced to shift the full payment β in his proposal, because by boosting his claim he gets a greater award in case of victory, but at the same time increases the risk of not being chosen by the arbitrator.

2.5 Hiring Stage

When deciding whether to hire a lawyer, each disputant will trade off the cost of legal representation with the benefit (which is notably captured by the effort made by the advocate to influence the arbitrated settlement favorably for her client). In turn, each lawyer agrees to defend the case if her participation constraint is satisfied. The conditions under which legal representation occurs in equilibrium, depending on the fee contract in place, are highlighted in the following proposition.

PROPOSITION 3 *Disputant B is legally represented under contingent (conditional) fees if* $\alpha < 1/2$ *(if* $\beta \leq 2/3$*); legal representation never occurs under flat fees.*

The intuition behind this proposition is the following. First, the poor incentive properties of flat fees obviously explain why legal representation does not occur in equilibrium with this remuneration scheme in place. Both parties anticipate that their lawyers will not invest to influence the arbitrator under this payment mechanism. Therefore, given that legal representation is costly and the arbitration outcome would be the same without lawyers, both disputants are better off by choosing not to hire advocates. Furthermore, the results show that disputant A's incentives not to be legally represented are not altered by the payment contract chosen by his adversary.

Second, as shown in appendix A.3, lawyer B's participation constraint is always satisfied under contingent fees, and the cap at 50% on α is thus induced by her client's incentives:

$$\Pi_B^C > 0 \ (\forall \alpha > 0), \quad \text{while} \quad U_B^C(\lambda_B = 1) > U_B^C(\lambda_B = 0) \text{ iff } \alpha < \frac{1}{2}\,.$$

Indeed, even if increasing the contingent percentage improves the lawyer's effort to defend the case (since $e_B^C = \alpha$ in equilibrium), this percentage should be capped – if it is to be used – because it increases the arbitration cost for the client and reduces his residual benefits from the procedure. However, it is easy to show that, when $\alpha < 25\%$, increasing α improves the client's incentives to be legally represented, precisely because his lawyer is more prone to make effort and to distort the arbitrator's opinion favorably, while her services are not too costly:

$$\frac{\partial U_B^C(\lambda_B = 1)}{\partial \alpha} > 0 \quad \Leftrightarrow \quad \alpha < \frac{1}{4}\,.$$

To the contrary, the results highlight that the cap at $2/3$ on β is due to the lawyer's participation constraint:

$$U_B^K(\lambda_B = 1) > U_B^K(\lambda_B = 0) \ (\forall \beta > 0), \quad \text{while} \quad \Pi_B^K \geq 0 \text{ iff } \beta \leq \frac{2}{3}\,.$$

In other words, the attorney agrees to defend the case only if her conditional payment is lower than a given threshold. This paradoxical result is due to the perverse influences that an increase in the conditional fee may have on the disputants' behavior, via its effect on lawyer B's effort, and hence on the arbitration outcome (it being known that the advocate is paid only in case of winning). In fact, an increase in the conditional fee makes the lawyer more prone to invest in the case (since

$e_B^K = \beta$ in equilibrium), which increases her effort cost and has two opposite effects on the probability $(1 - \phi)$ of winning in arbitration and, hence, of being paid. First, this probability is positively affected by e_B – ceteris paribus – since the arbitrator's opinion is moved favorably by this quantity. However, the lawyer's effort has also a negative – strategic – effect on the probability of being chosen by the arbitrator, via its inf uence on the parties' equilibrium f nal offers: following (13), increasing e_B (or β) induces both parties to boost their claims (i.e., a^K and b^K), which in turn enhances the likelihood of losing for disputant B and raises the probability of his lawyers's being unpaid. Overall, the value of β should be capped to prevent its positive effect on the lawyer's payoff from being more than offset by its effect on the lawyer's effort cost and by its perverse effect on the parties' strategies and hence on the arbitrator's decision.

3 Conclusion

In this paper, we highlight the strategic implications of legal representation in FOA by analyzing how the presence of lawyers may shape the disputants' bidding behavior and studying their incentives to get legal counsel. This issue is investigated by considering three different remuneration contracts behind the client–attorney relationship (viz., f at, contingent, and conditional fees). Following our main f ndings, the contingent payment mechanism has interesting properties in that it enhances the lawyer's incentives to provide effort in defending her client (in comparison with the case without legal representation), without altering the gap between the parties' claims in arbitration. However, the contingent share of the arbitrated settlement given to the lawyer must be capped for legal representation to occur in equilibrium. Indeed, this share increases the arbitration cost for the client and reduces his residual benef ts from the procedure.

This paper may be considered as a f rst step in the theoretical investigation of the role of legal representation in arbitration, and several extensions should be made to improve our approach to this issue. Firstly, we consider that the disputants alone decide which f nal offers to make to the arbitrator. This view corresponds to the *instrumentalist* approach to the decision-making allocation, according to which the lawyer merely supplies the technical knowledge and skills necessary to implement the decision, while the client has the ultimate authority over the objectives. It would be interesting to construct a *paternalist* model, in which lawyers get the exclusive decision-making authority (Maute, 1984; Choi, 2003) and choose the proposals that maximize their own payoffs, depending on the fee schedule in place.

Furthermore, in accordance with most of the literature, we measure conflic by the range between the parties' proposals, and such an approach may be discussed in light of the following arguments. First, it is not clear that less extreme bids do in fact lead to more settlements in the pre-arbitration phase, because more reasonable offers imply less risk in arbitration, which lowers the uncertainty costs of the procedure and may induce more disagreements. Herein lies a potential explanation for the empirical finding of higher dispute rates with FOA than with a comparable CA mechanism (see, e.g., Ashenfelter et al., 1992; Dickinson, 2004). The original idea behind the suggestion that FOA would increase voluntary settlements over CA is that FOA eliminates the middle of the arbitrator award distribution (by removing the ability of the arbitrator to compromise). However, as underlined by Farber and Bazerman (1989), FOA may also decrease uncertainty by ruling out the tails of the distribution and allowing the parties to mitigate the risk by manipulating their proposals so as to affect the choice probabilities. In other words, FOA is not more risky than CA in a mean-preserving spread or stochastic dominance sense. These considerations suggest that FOA may produce higher dispute rates than CA precisely because it induces convergence between the parties' claims. Second, as mentioned in footnote 8, the parties' bidding behavior is quite sensitive to the arbitrator's settlement distribution, and the uniform case that is considered in the paper makes bids diverge extremely under FOA. Although each individual increases the probability that his proposal will be chosen by moving toward the midpoint of the distribution, this gain in probability is more than offset by the loss in payoff he suffers by moderating his offer. The uniform distribution implies that if this is true at one point, it is true at all points in the distribution. It is then a dominant strategy to make an extreme proposal. For these two reasons, it would be relevant to relax the uniform assumption and to explicitly include the negotiation stage that would precede the arbitration procedure. Some papers in the literature contemplate arbitration and bargaining as alternatives that parties choose through a negotiation process (Manzini and Mariotti, 2001; Compte and Jehiel, 2004; Mercedes Adamuz and Ponsatí, 2009), and such a design would be useful for understanding how the linkages between bargaining and arbitration actually work.

Moreover, we design the arbitration process as a four-stage game where the parameters α and β are common knowledge. In practice, it is impossible for the defendant to see the exact values of these parameters, since the contracts for α and β are a kind of privileged communication or document between the plaintiff and his lawyer (Baik and Kim, 2007). To reflect this situation, it would be interesting to consider a framework in which the defendant is constrained by asymmetric information.

Finally, we consider a setting where the arbitrator is presumed to know more about the case than the disputants themselves: the arbitrator is assumed to construct the *appropriate* award (i.e., the "right" decision s) as a function of the merits of the case, while the disputants' offers are based on their beliefs about this appropriate settlement. This setting is consistent with most of the articles on arbitration, but two notable exceptions are Gibbons (1988) and Samuelson (1991).[12] Following these papers, we could consider that the parties themselves, and not the arbitrator, have substantial information on the dispute (e.g., the true amount of damages suffered by the plaintiff). In this setup, the arbitrator would learn from the parties' final offers and base his/her assessment of the right decision on those offers, knowing that

[12] See also Emons and Fluet (2009) in the context of litigation.

each disputant might be incited to misrepresent his private information and distort the truth by boosting his claim and hiring a lawyer to provide evidence. A relevant welfare criterion to evaluate each lawyer's payment mechanism could be to minimize both the arbitrator's error (i.e., the discrepancy between the arbitrated settlement and the truth) and the inf uence costs (i.e., the costs incurred by the disputants to infl ence the arbitrator's decision, via legal representation for example). This analysis would be really interesting, but it would create some technical diff culties in arriving at theoretical predictions. In particular, the players' decisions would be part of a perfect Bayesian equilibrium, and, in view of the seminal paper by Crawford and Sobel (1982), this kind of communication game is often plagued by a multiplicity of equilibria.

Overall, a framework based on some of these extensions would certainly provide a more complete and robust analysis of the role of lawyers in arbitration. Our aim was to develop a theoretical basis to understand this role under idealized conditions, as a prerequisite to analyze it in a more integrative process.

Appendix

A.1 Proof of Proposition 1

We consider that $\lambda_i = 1$, $\forall i \in \{A, B\}$, since each lawyer decides whether to exert effort only if she has been hired at stage 1 of the game.

Following (2), the f rst-order condition for attorney A implies immediately that $e_A^j = 0$, $\forall j \in \{F, C, K\}$. Concerning lawyer B, using (3)–(5), we get the following f rst-order conditions:

$$\frac{\partial \Pi_B^F}{\partial e_B} = 0 \;\Leftrightarrow\; e_B = 0,$$

$$\frac{\partial \Pi_B^C}{\partial e_B} = 0 \;\Leftrightarrow\; -e_B + \alpha(b-a) = 0,$$

$$\frac{\partial \Pi_B^K}{\partial e_B} = 0 \;\Leftrightarrow\; -e_B + \beta = 0.$$

The results stated in (6) follow immediately. *Q.E.D.*

A.2 Proof of Proposition 2

Given that disputant A uses fla fees, three cases have to be considered, depending on the fee contract chosen by disputant B.

(1) Flat Fees. Following (7) and (8), the f rst-order conditions for disputants A and B are respectively

$$\frac{\partial U_A^F}{\partial a} = -\phi - \frac{1}{2}(a-b) = 0, \qquad \frac{\partial U_B^F}{\partial b} = 1 - \phi + \frac{1}{2}(a-b) = 0\,,$$

with $e_A^F = e_B^F = 0$, implying that $\phi = (a+b)/2$.

Solving these for optimal values a^F and b^F yields

(A1) $$a^F = 0 \quad \text{and} \quad b^F = 1\,.$$

The range R^F between these offers is then given by

$$R^F \equiv b^F - a^F = 1\,.$$

The parties' offers would be the same without legal representation, since both lawyers provide zero efforts under f at fees.

(2) Contingent Fees. Following (7) and (9), the f rst-order conditions for disputants A and B are respectively

$$\frac{\partial U_A^C}{\partial a} = -\phi - \frac{1}{2}(a-b) + \lambda_B\alpha(b-a) = 0,$$
$$\frac{\partial U_B^C}{\partial b} = (1-\lambda_B\alpha)\left[1 - \phi + \left(\frac{1}{2} - \alpha\lambda_B\right)(a-b)\right] = 0\,,$$

with $e_A^C = 0$ and $e_B^C = \alpha(b-a)$, implying that $\phi = (a+b)/2 - \lambda_B\alpha(b-a)$.

Since $1 - \lambda_B\alpha > 0$, solving these for optimal values a^C and b^C yields

(A2) $$a^C = 2\lambda_B\alpha \quad \text{and} \quad b^C = 1 + 2\lambda_B\alpha\,.$$

The range R^C between these offers is then given by

$$R^C = 1 \equiv R^F\,.$$

Lawyer B's effort evaluated at the equilibrium parties' offers is then $e_B^C = \alpha$.

(3) Conditional Fees. Following (7) and (10), the f rst-order conditions for disputants A and B are respectively

$$\frac{\partial U_A^K}{\partial a} = -\phi - \frac{1}{2}(a-b) = 0, \qquad \frac{\partial U_B^K}{\partial b} = 1 - \phi + \frac{1}{2}(a - b + \lambda_B\beta) = 0,$$

with $e_A^K = 0$ and $e_B^K = \beta$, implying that $\phi = (a+b)/2 - \lambda_B\beta$.

Solving these for optimal values a^K and b^K yields

(A3) $$a^K = \lambda_B\beta \quad \text{and} \quad b^K = 1 + \frac{3\lambda_B\beta}{2}\,.$$

The range R^K between these offers is then given by

$$R^K = 1 + \frac{\beta}{2} > R^F \equiv R^C \quad \Leftrightarrow \quad \beta > 0\,.$$

Q.E.D.

A.3 Proof of Proposition 3

For simplicity and without loss of generality, we normalize to zero the lawyer's reservation payoff.

(1) Flat Fees. Using (2), (3), and (6), the equilibrium payoff to lawyer i is

(A4) $$\Pi_i^F = f > 0 \quad \forall i \in \{A, B\}\,.$$

Using (6)–(8), and (A1), the equilibrium utilities to disputants A and B are, respectively

$$U_A^F = -\frac{1}{2} - \lambda_A f \quad \text{and} \quad U_B^F = \frac{1}{2} - \lambda_B f\,. \tag{A5}$$

Hiring a lawyer is a strictly dominant strategy for disputants A and B if

$$U_A^F(\lambda_A = 1) > U_A^F(\lambda_A = 0) \quad \text{and} \quad U_B^F(\lambda_B = 1) > U_B^F(\lambda_B = 0)\,. \tag{A6}$$

Using (A5), we can state that

$$\begin{aligned} U_A^F(\lambda_A = 1) &= -\frac{1}{2} - f, \quad U_A^F(\lambda_A = 0) = -\frac{1}{2}, \\ U_B^F(\lambda_B = 1) &= \frac{1}{2} - f, \quad U_B^F(\lambda_B = 0) = \frac{1}{2}\,. \end{aligned} \tag{A7}$$

Using (A4) and (A6)–(A7), we can state that each lawyer's participation constraint is satisf ed, but no disputant is induced to be legally represented under f at fees.

(2) Contingent Fees. Using (2), (4), and (6), the equilibrium payoffs to lawyers A and B are, respectively,

$$\Pi_A^C = f > 0 \quad \text{and} \quad \Pi_B^C = \frac{\alpha}{2}(1+\alpha) > 0, \text{ since } \alpha > 0\,. \tag{A8}$$

Using (7), (9), and (A2), the equilibrium utilities to disputants A and B are, respectively,

$$U_A^C = -\frac{1}{2} - \lambda_A f - \lambda_B \alpha \quad \text{and} \quad U_B^C = (1 - \lambda_B \alpha)\left(\frac{1}{2} + \lambda_B \alpha\right). \tag{A9}$$

Hiring a lawyer is a strictly dominant strategy for disputants A and B if

$$U_A^C(\lambda_A = 1) > U_A^C(\lambda_A = 0) \quad \text{and} \quad U_B^C(\lambda_B = 1) > U_B^C(\lambda_B = 0)\,. \tag{A10}$$

Using (A9), we can state that

$$\begin{aligned} U_A^C(\lambda_A = 1) &= -\frac{1}{2} - f - \lambda_B \alpha, \qquad U_A^C(\lambda_A = 0) = -\frac{1}{2} - \lambda_B \alpha, \\ U_B^C(\lambda_B = 1) &= (1-\alpha)\left(\frac{1}{2} + \alpha\right), \quad U_B^C(\lambda_B = 0) = \frac{1}{2}\,. \end{aligned} \tag{A11}$$

Using (A8) and (A10)–(A11), we can state that disputant A does not hire a lawyer, while disputant B is legally represented if $\alpha < 1/2$.

(3) Conditional Fees. Using (2), (5), and (6), the equilibrium payoffs to lawyers A and B are, respectively,

$$\Pi_A^K = f > 0 \quad \text{and} \quad \Pi_B^K = \frac{\beta}{2}\left(1 - \frac{3\beta}{2}\right) \geq 0 \Leftrightarrow \beta \leq \frac{2}{3}\,. \tag{A12}$$

Using (7), (10), and (A3), the equilibrium utilities to disputants A and B are, respectively,

$$U_A^K = -\left[\frac{1}{2} + \lambda_A f + \lambda_B \beta - \frac{(\lambda_B \beta)^2}{8}\right] \quad \text{and} \quad U_B^K = \frac{1}{2} + \frac{\lambda_B \beta}{2} + \frac{(\lambda_B \beta)^2}{8}\,. \tag{A13}$$

Hiring a lawyer is a strictly dominant strategy for disputants A and B if

$$\text{(A14)} \qquad U_A^K(\lambda_A = 1) > U_A^K(\lambda_A = 0) \quad \text{and} \quad U_B^K(\lambda_B = 1) > U_B^K(\lambda_B = 0)\,.$$

Using (A13) , we can state that

(A15)

$$U_A^K(\lambda_A = 1) = -\frac{1}{2} - f - \lambda_B\beta + \frac{(\lambda_B\beta)^2}{8}, \quad U_A^K(\lambda_A = 0) = -\frac{1}{2} - \lambda_B\beta + \frac{(\lambda_B\beta)^2}{8},$$
$$U_B^K(\lambda_B = 1) = \frac{1}{2} + \frac{\beta}{2}\left(1 + \frac{\beta}{4}\right), \quad U_B^K(\lambda_B = 0) = \frac{1}{2}\,.$$

Using (A12) and (A14)–(A15), we can state that disputant A does not hire a lawyer, while disputant B is legally represented if $\beta \leq 2/3$. *Q.E.D.*

References

Armstrong, M. J. (2004), "A Comparison of Arbitration Procedures for Risk-Averse Disputants," *Decision Sciences*, 35(4), 639–664.

— and W. J. Hurley (2002), "Arbitration Using the Closest Offer Principle of Arbitrator Behavior," *Mathematical Social Sciences*, 43(1), 19–26.

Ashenfelter, O., J. Currie, H. S. Farber, and M. Spiegel (1992), "An Experimental Comparison of Dispute Rates in Alternative Arbitration Systems," *Econometrica*, 60(6), 1407–1433.

— and G. B. Dahl (2012), "Bargaining and the Role of Expert Agents: An Empirical Study of Final-Offer Arbitration," *The Review of Economics and Statistics*, 94(1), 116–132.

Baik, K. H., and I.-G. Kim (2007), "Contingent Fees versus Legal Expenses Insurance," *International Review of Law and Economics*, 27(3), 351–361.

Barnacle, P. J. (1991), *Arbitration of Discharge Grievances in Ontario: Outcomes and Reinstatement Experiences*, Industrial Relations Centre, Queen's University, Kingston (ON).

Block, R. N., and J. Stieber (1987), "The Impact of Attorneys and Arbitrators on Arbitration Awards," *Industrial & Labor Relations Review*, 40(4), 543–555.

Brams, S. J., D. M. Kilgour, and S. Merrill III (1991), "Arbitration Procedures," in: H. P. Young (ed.), *Negotiation Analysis*, The University of Michigan Press, Ann Arbor, pp. 47–65.

— and S. Merrill III (1983), "Equilibrium Strategies for Final-Offer Arbitration: There is No Median Convergence," *Management Science*, 29(8), 927–941.

— and — (1986), "Binding versus Final-Offer Arbitration: A Combination Is Best," *Management Science*, 32(10), 1346–1355.

Burger, J. D., and S. J. K. Walters (2005), "Arbitrator Bias and Self-Interest: Lessons from the Baseball Labor Market," *Journal of Labor Research*, 26(2), 267–280.

Chatterjee, K. (1981), "Comparison of Arbitration Procedures: Models with Complete and Incomplete Information," *IEEE Transactions on Systems, Man, and Cybernetics*, 11(2), 101–109.

Chen, K.-P., and J.-S. Wang (2007), "Fee-Shifting Rules in Litigation with Contingency Fees," *The Journal of Law, Economics, & Organization*, 23(3), 519–546.

Choi, A. (2003), "Allocating Settlement Authority under a Contingent-Fee Arrangement," *The Journal of Legal Studies*, 32(2), 585–610.

Compte, O., and P. Jehiel (2004), "Gradualism in Bargaining and Contribution Games," *The Review of Economic Studies*, 71(4), 975–1000.

Crawford, V. P., and J. Sobel (1982), "Strategic Information Transmission," *Econometrica*, 50(6), 1431–1451.

Dana, Jr., J. D., and K. E. Spier (1993), "Expertise and Contingent Fees: The Role of Asymmetric Information in Attorney Compensation," *The Journal of Law, Economics, & Organization*, 9(2), 349–367.

Deck, C. A., and A. Farmer (2007), "Bargaining over an Uncertain Value: Arbitration Mechanisms Compared," *The Journal of Law, Economics, & Organization*, 23(3), 547–579.

— and — (2009), "Strategic Bidding and Investments in Final Offer Arbitration: Theory and Experimental Evidence," *Journal of Economic Behavior & Organization*, 70(1–2), 361–373.

Deck, C., A. Farmer, and D.-Z. Zeng (2007), "Amended Final-Offer Arbitration over an Uncertain Value: A Comparison with CA and FOA," *Experimental Economics*, 10(4), 439–454.

Dickinson, D. L. (2004), "A Comparison of Conventional, Final-Offer, and 'Combined' Arbitration for Dispute Resolution," *Industrial & Labor Relations Review*, 57(2), 288–301.

Emons, W. (2007), "Conditional versus Contingent Fees," *Oxford Economic Papers*, 59(1), 89–101.

— and C. Fluet (2009), "Accuracy versus Falsificatio Costs: The Optimal Amount of Evidence under Different Procedures," *The Journal of Law, Economics, & Organization*, 25(1), 134–156.

— and N. Garoupa (2006), "US-Style Contingent Fees and UK-Style Conditional Fees: Agency Problems and the Supply of Legal Services," *Managerial and Decision Economics*, 27(5), 379–385.

Farber, H. S. (1980), "An Analysis of Final-Offer Arbitration," *The Journal of Conflic Resolution*, 24(4), 683–705.

— (1981), "Splitting-the-Difference in Interest Arbitration," *Industrial & Labor Relations Review*, 35(1), 70–77.

— and M. H. Bazerman (1989), "Divergent Expectations as a Cause of Disagreement in Bargaining: Evidence from a Comparison of Arbitration Schemes," *The Quarterly Journal of Economics*, 104(1), 99–120.

— and H. C. Katz (1979), "Interest Arbitration, Outcomes, and the Incentive to Bargain," *Industrial & Labor Relations Review*, 33(1), 55–63.

Gibbons, R. (1988), "Learning in Equilibrium Models of Arbitration," *The American Economic Review*, 78(5), 896–912.

Gravelle, H., and M. Waterson (1993), "No Win, No Fee: Some Economics of Contingent Legal Fees," *The Economic Journal*, 103(420), 1205–1220.

Hanany, E., D. M. Kilgour, and Y. Gerchak (2007), "Final-Offer Arbitration and Risk Aversion in Bargaining," *Management Science*, 53(11), 1785–1792.

Manzini, P., and M. Mariotti (2001), "Perfect Equilibria in a Model of Bargaining with Arbitration," *Games and Economic Behavior*, 37(1), 170–195.

Maute, J. L. (1984), "Allocation of Decisionmaking Authority under the Model Rules of Professional Conduct," *U.C. Davis Law Review*, 17(4), 1049–1115.

Mercedes Adamuz, M. de las, and C. Ponsatí (2009), "Arbitration Systems and Negotiations," *Review of Economic Design*, 13(3), 279–303.

Miller, G. P. (1987), "Some Agency Problems in Settlement," *The Journal of Legal Studies*, 16(1), 189–215.

Samuelson, W. F. (1991), "Final-Offer Arbitration under Incomplete Information," *Management Science*, 37(10), 1234–1247.

Santore, R., and A. D. Viard (2001), "Legal Fee Restrictions, Moral Hazard, and Attorney Rents," *The Journal of Law & Economics*, 44(2), 549–572.

Wagar, T. H. (1994), "The Effect of Lawyers on Non-Discipline/Discharge Arbitration Decisions," *Journal of Labor Research*, 15(3), 283–293.

Zeng, D.-Z. (2003), “An Amendment to Final-Offer Arbitration,” *Mathematical Social Sciences*, 46(1), 9–19.

Nathalie Chappe
CRESE
University of Franche-Comté
45, avenue de l’observatoire
25030 Besançon
France
nathalie.chappe@univ-fcomte.fr

Yannick Gabuthy
CRESE
University of Franche-Comté
45, avenue de l’observatoire
25030 Besançon
France
yannick.gabuthy@univ-fcomte.fr